Optimierung biometrischer Hash-Algorithmen für die dynamische Handschrift

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von Dipl.-Inform. Tobias Scheidat

geb. am 27.03.1972 in Magdeburg

Gutachterinnen/Gutachter

Prof. Dr. Jana Dittmann, Otto-von-Guericke-Universität Magdeburg
Prof. Dr. Claus Vielhauer, Fachhochschule Brandenburg
Prof. Dr. Andreas Uhl, Universität Salzburg
Prof. Dr. Stefan Katzenbeisser, Technische Universität Darmstadt

Magdeburg, den 21.01.2015

Danksagung

Mein besonderer Dank gilt Frau Prof. Dr. Jana Dittmann für die interessante Aufgabenstellung und die Betreuung der Arbeit. Zusammen mit Herrn Prof. Dr. Claus Vielhauer hat sie mich mit Ideen, Hinweisen und Ratschlägen über den Verlauf der Entstehung dieser Arbeit umfangreich unterstützt und mir in Diskussionen immer wieder neue Blickwinkel auf die einzelnen Bereiche der Thematik ermöglicht. Weiterhin danke ich beiden sowie Prof. Dr. Andreas Uhl (Universität Salzburg) und Prof. Dr. Stefan Katzenbeisser (Technische Universität Darmstadt) für die Begutachtung dieser Arbeit.

Ich bedanke mich bei allen, die direkt oder indirekt an der Entstehung dieser Arbeit beteiligt waren. Ohne die vielen freiwilligen Handschriftspender wären die Tests, die in dieser Arbeit durchgeführt wurden, nicht möglich gewesen. Den aktuellen und ehemaligen Kollegen der Arbeitsgruppen *Advanced Multimedia and Security* an der Otto-von-Guericke-Universität in Magdeburg und *Angewandte Informatik/Medieninformatik, insbesondere Datensicherheit* an der Fachhochschule Brandenburg danke ich für die gute Zusammenarbeit sowie die Vielzahl fruchtbarer und interessanter Diskussionen. Weiter bin ich sehr dankbar für die Möglichkeiten, die mir durch die Mitarbeit in verschiedenen nationalen und internationalen Projekten gegeben wurden. In der projektbezogenen Arbeit und dem Austausch mit Kollegen aus aller Welt konnte ich viel an Erfahrungen gewinnen, vor allem in den Bereichen IT-Sicherheit, Biometrie, Benchmarking und digitale Forensik.

Meiner Frau Steffi gilt mein besonderer Dank, da sie mich während meiner Arbeit an der Dissertation häufig moralisch unterstützt, auf viel gemeinsame Zeit verzichtet und mir oft aufbauend zur Seite gestanden hat. Ihr und Claudia Ihnen bin ich für die große Hilfe bei der grammatikalischen und rechtschreiblichen Korrektur der Arbeit zu Dank verpflichtet.

Zusammenfassung

In dieser wissenschaftlichen Arbeit untersuchen wir verschiedene Möglichkeit zur Verbesserung der Performanz biometrischer Algorithmen. Dabei adressieren wir unter dem Begriff Performanz sowohl die Erkennungsgenauigkeit der Authentifikation, als auch die Reproduzierbarkeit und Kollisionsresistenz biometrischer Hashes. Wir fokussieren in der dieser Arbeit zugrunde liegenden Forschung die biometrische Modalität der Handschrift, wobei die Herangehensweise und die gewählten Ansätze auch auf andere biometrische Modalitäten übertragbar sind.

Die im Rahmen dieser Arbeit identifizierten und umgesetzten Optimierungsansätze werden mit anonymen Handschriftendaten einer eigenen Datenbank experimentell evaluiert. Zu deren Erstellung wurden mittels eines TabletPC von 53 Personen jeweils zehn Handschriftproben mit fünf unterschiedlichen Inhalten (Semantiken) erfasst. Dieses Vorgehen wurde nach Ablauf von einem beziehungsweise zwei Monaten wiederholt. Dadurch ergibt sich eine Gesamtzahl von 7950 Schriftproben. Bei der experimentellen Evaluierung konnten die Fehlerraten sowohl für die Authentifikation als auch die Hash-Generierung zum Teil signifikant verringert werden. So kann beispielsweise für die Semantik *Symbol* die Equal Error Rate durch Parameteranpassung und Merkmalsselektion von einem Ausgangswert von 0,16158 auf 0,03286 verbessert werden. Aus dem Blickwinkel der Hash-Generierung kann im besten Fall eine sehr gute Collision Reproduction Rate von 0,02848 ebenfalls durch Optimierung von Parametern und Selektion von Merkmalen erzielt werden. Dieser Wert ergibt sich für das *Symbol* aus einer Reproduction Rate von 0,98491 und einer Collision Rate von 0,04187. Das bedeutet, dass ein Nutzer seinen individuellen biometrischen Hash mit einer Wahrscheinlichkeit von circa 98,5% reproduzieren kann, während eine Kollision durch Hashes anderer Personen nur zu circa 4,2% wahrscheinlich ist.

Zusätzlich konnten wir basierend auf unserer Datenbank und Evaluierungsmethodologie zeigen, dass sich mit zunehmendem zeitlichen Abstand zwischen den Referenz- und Testdaten die Authentifikationsperformanz verschlechtert. Dieses Phänomen ist bereits für geringe Zeiträume von einem beziehungsweise zwei Monaten sowohl für beide Algorithmen, als auch jeweils für die fünf Semantiken zu beobachten. So konnte festgestellt werden, dass sich die Equal Error Rate bis um das siebenfache verschlechtert, wenn zwischen der Aufnahme der Referenz- und Testdaten zwei Monate liegen.

Mit dieser Arbeit wird ein Beitrag in den Forschungsbereichen der biometrischen Benutzererkennung und der biometrischen Hash-Generierung im Kontext der dynamischen Handschrift geleistet. Sie bereitet Grundlagen für weitere Untersuchungen in Bezug auf die Optimierung biometrischer Verfahren auf Basis von zusätzlichen Merkmalen, der Analyse und Selektion von Merkmalen und der single-modalen Fusion, auch in Bezug auf weitere biometrische Modalitäten beziehungsweise Forschungs- und Anwendungsbereiche.

Inhaltsverzeichnis

1 Einleitung und Motivation

Die Verwendung und Verbreitung von Technik für die Speicherung, den Transfer und die Verarbeitung von Informationen und zur Kommunikation hat im Laufe der letzten Jahre immer mehr zugenommen, wobei ein Ende der Entwicklung nicht abzusehen ist. Aus diesem Grund gewinnt verständlicherweise die Forderung nach Sicherheit der zugrunde liegenden IT-Systeme immer mehr an Bedeutung, da zunehmend Interessen Dritter an der Erlangung und Verwendung von Informationen bestehen. Diese Daten sind häufig von sensibler Natur, insbesondere wenn sie Informationen über bestimmte Personen oder Personenkreise preisgeben oder mit monetären Aspekten verbunden sind. Beispielhaft können hier die Adresse, persönliche Korrespondenz, auf das Online-Banking bezogene Informationen, aber auch Vorlieben bei der Nutzung von Online-Angeboten wie das Einkaufs- oder die Suchgewohnheiten von Nutzern genannt werden. Natürlich sind sicherheitsrelevante Informationen auch in anderen Bereichen, wie der kommunalen Verwaltung oder in industriellen beziehungsweise dienstleistenden Unternehmen zu finden. Für jedes der genannten Beispiele existiert eine Vielzahl von Interessenten, die einen Nutzen aus der Gewinnung und Verwertung entsprechender Informationen ziehen können und oft auch wollen. Dabei können potenzielle Angreifer unterschiedliche Ziele verfolgen: Einige nutzen Sicherheitslücken dazu, um Schwachstellen aufzuzeigen und deren Behebung zu motivieren, andere haben das Abfangen, die Löschung oder die Veränderung von Daten zum Ziel, während wiederum andere Angreifer Daten sammeln, um sie für kommerzielle Zwecke zu verwenden. Ein wichtiges Ziel des Forschungs- und Anwendungsbereiches der IT-Sicherheit ist es, IT-Systeme und damit verbundene Informationen zu sichern beziehungsweise zu schützen. Rechtlich ist der Umgang mit personenbezogenen Daten so geregelt, dass der Betroffene nicht in seinem Persönlichkeitsrecht beeinträchtigt wird. Grundlage dafür bilden das Bundesdatenschutzgesetz (BDSG[1]) und die Datenschutzgesetze der Länder, wie beispielsweise das Datenschutzgesetz des Landes Sachsen-Anhalt (DSG LSA[2]). Die Datenschutzgesetze geben organisatorische und technische Regeln vor, die dazu beitragen sollen, den Einzelnen vor dem Missbrauch seiner persönlichen Daten zu schützen. Dem DSG LSA zufolge sind personenbezogene Daten *„Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person“*. Weiter handelt es sich um personenbezogene Daten besonderer Art bei *„Angaben über die rassische oder ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschafts-*

[1] Bundesdatenschutzgesetz in der Fassung vom 20. Dezember 1990 (BGBl. I S. 2954), neugefasst durch Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), zuletzt geändert durch Gesetz vom 29.07.2009 (BGBl. I, S. 2254), durch Artikel 5 des Gesetzes vom 29.07.2009 (BGBl. I, S. 2355 [2384] und durch Gesetz vom 14.08.2009 (BGBl. I, S. 2814)

[2] Gesetz zum Schutz personenbezogener Daten der Bürger (Datenschutzgesetz Sachsen-Anhalt - DSG LSA) vom 12. März 1992 (GVBl. S. 152) in der Fassung der Bekanntmachung vom 18. Februar 2002 (GVBl. LSA S. 54), zuletzt geändert durch Artikel 1 des Gesetzes vom 27. September 2011 (GVBl. LSA S. 648)

zugehörigkeit, Gesundheit oder Sexualleben". Zu biometrischen Daten werden keine speziellen Regelungen getroffen, allerdings sind eine Vielzahl durch biometrische Sensoren aufgenommene Daten einer der beiden Kategorien zuzuordnen. Entsprechend sind Vorgaben zum Umgang mit personenbezogenen Daten auch bei der Erhebung, Speicherung, Verarbeitung und Löschung von biometrischen Daten zu beachten.

Der Ansatz, mit dem wir uns im Rahmen dieser Arbeit befassen, bietet eine Möglichkeit, Aspekte des Datenschutzes beim Umgang mit biometrischen Daten zu berücksichtigen. Wir fokussieren dazu biometrische Hash-Funktionen, welche biometrische Eingangsdaten in andere Wertebereich transferieren. Die wesentlichen Ziele sind dabei sowohl der Schutz biometrischer Informationen vor dem Missbrauch, damit verbunden die Durchführung von Authentifikationen im Hash-Bereich als auch die potentielle Generierung biometrischer Hashes zu kryptographischen Zwecken. Inspiriert wird die biometrische Hash-Generierung durch kryptographische Hash-Funktionen. Dabei werden die Grundeigenschaften kryptographischer Hash-Funktionen (Unumkehrbarkeit, Reproduzierbarkeit und Kollisionsresistenz, siehe zum Beispiel [DK07], [Eck05]) gemäß der Gegegebenheiten biometrischer Daten adaptiert (weitere Informationen dazu folgen in den Abschnitten 2.1.4 *Biometrisches Hashing* und 4 *Verwendete Algorithmen*). Die genannten drei Anwendungsbereiche, Datenschutz, Authentifikation und Hash-Generierung, werden wir im Rahmen dieser Arbeit basierend auf den beiden biometrischen Handschriftverfahren *Biometric Hash* [Vie06] und *Secure Sketch* [MSV12] betrachten. Im Mittelpunkt steht die jeweilige Optimierung der Algorithmen in Hinblick auf die Authentifikation und die Generierung von stabilen individuellen biometrischen Hashes. Mit einer nicht umkehrbaren Transformation der statistischen Merkmale, welche aus biometrischen Daten generiert wurden, in einen Hash-Raum wird auch der Datenschutz adressiert. Die Nutzung der biometrischen Hashes als Referenzdaten für ein biometrisches System hat den Vorteil, dass im Fall einer Kompromittierung, beispielsweise dem Diebstahl der Daten, aus diesen keine Rückschlüsse auf den Urheber gezogen werden können. Als dynamische biometrische Eigenschaft besitzt gerade die Handschrift zusätzlich die Möglichkeit, den Inhalt des geschriebenen zu verändern. Das kann beispielsweise als Reaktion auf die Kompromittierung erfolgen, aber auch, um die Handschrift für unterschiedliche Authentifikationszwecke einsetzen zu können.

1.1 Wissenschaftliche Herausforderung

Die sichere Authentifikation von Personen ist in der heutigen technisierten Welt nicht in jedem Fall auf direktem Wege möglich. Kommunikation zwischen mehreren Partnern findet immer häufiger über weite Entfernungen statt, ermöglicht durch lokale oder globale Telekommunikations- und Computer-Netzwerke. Aus diesem Grund kann die Identität der teilnehmenden Parteien nicht mehr von Angesicht zu Angesicht festgestellt werden. Daher übernehmen automatische Systeme die Aufgabe der Ermittlung beziehungsweise Sicherstellung der Identität. Heutzutage wird die automatische Authentifikation, neben den traditionellen Methoden *geheimes Wissen* und *persönlicher Besitz*, vermehrt mittels *Biometrie* durchgeführt.

Die Bedeutung biometrischer Systeme im Bereich der Benutzerauthentifikation hat in den

letzten Jahrzehnten immer mehr zugenommen. Biometrische Verfahren nutzen physiologische beziehungsweise auf Verhalten basierende Eigenschaften einer Person. Diese Vorgehensweise bietet Vorteile, die die herkömmlichen Authentifikationsmethoden nicht aufweisen. Beispielhaft sei hier das Präsentieren eines Authentifikationsobjektes genannt, welches sich direkt auf die jeweilige Person bezieht und dadurch nicht verloren gehen oder von anderen Personen verwendet werden kann. Auf der anderen Seite existieren auch in der biometrischen Benutzererkennung spezifische Nachteile. Neben natürlichen Schwankungen der Daten eines Benutzers beziehungsweise Ähnlichkeiten zwischen Daten unterschiedlicher Personen und den daraus resultierenden Nicht- beziehungsweise Falscherkennungen entstehen auch ganz neue Anforderungen an den Datenschutz. Da in der Biometrie mit personenbezogenen beziehungsweise personenbeziehbaren Daten gearbeitet wird, müssen diese vor Missbrauch entsprechend geschützt werden. Folglich bestehen zwei wesentliche Ziele der Forschung in der Biometrie darin, die Fehleranfälligkeit bei der Authentifikation zu verringern und die Sicherheit der biometrischen Informationen gegenüber Unbefugten zu gewährleisten. Im weiteren Verlauf verwenden wir - sofern nicht anders angegeben - den Begriff *Performanz* synonym zur Fehlerunanfälligkeit eines biometrischen Algorithmus oder Systems bezogen auf die Erkennung beziehungsweise Unterscheidung von Personen. Das bedeutet beispielsweise, dass eine Verbesserung der Performanz eines biometrischen Algorithmus in diesem Kontext durch eine Verringerung der Fehleranfälligkeit verursacht wird und umgekehrt.

Um die Authentifikationsperformanz biometrischer Systeme zu verbessern, werden nicht nur immer mehr Algorithmen entwickelt, sondern vorhandene Verfahren werden in vielen Fällen optimiert, zum Beispiel durch Hinzufügen zusätzlicher statistischer Merkmale. Auf der einen Seite können dadurch oft bessere Ergebnisse im Bereich der Benutzerauthentifikation erzielt werden. Für die Verwendung biometrischer Algorithmen in alternativen Anwendungsbereichen der Biometrie ist andererseits das stete unkritische Hinzufügen von Merkmalen nicht immer zielführend. Beispielhaft ist hier der Biometric Hash-Algorithmus für die dynamische Handschrift [Vie06] zu nennen, bei dem die Merkmalswerte einzeln auf separate Hash-Werte abgebildet werden. Verfolgt man hier durch zusätzliche Merkmale nur die einseitige Optimierung eines Ziels (beispielsweise die Kollisionsresistenz) geht dies zu Lasten anderer Eigenschaften des Systems (hier zum Beispiel der Reproduzierbarkeit). Zusätzlich besteht die Möglichkeit, dass die Menge der verwendeten Merkmale auch solche Elemente enthält, die der Unterscheidung verschiedener Personen beziehungsweise der Erkennung einzelner Individuen nicht dienlich sind. Vor allem in solchen Fällen kann es sinnvoll sein, vorhandene, aber auch neue Merkmale auf ihre Eignung für den spezifischen Verwendungszweck (wie zum Beispiel Authentifikation oder Hash-Generierung) zu untersuchen und entsprechend auszuwählen. Die Analyse und die darauf aufbauende Selektion von Merkmalen wird im Bereich der Biometrie häufig vernachlässigt, wenn das Hinzufügen neuer Merkmale allein schon zu einer nennenswerten Verbesserung führt. Zudem kann das Aussortieren von Merkmalen auch zu einer Verringerung der Laufzeit beziehungsweise der benötigten Rechenleistung und Speicherkapazität führen. Dies kann vor allem im Kontext heutzutage immer häufiger eingesetzter mobiler beziehungsweise eingebetteter Systeme von Vorteil sein, wenn eine bestimmte Anzahl von Merkmalen nicht berechnet werden braucht. Zusätzlich entfällt deren

Transfer in den Hash-Raum, wodurch gegebenenfalls Rechenzeit und -kapazität eingespart werden kann.

Da in biometrischen Systemen das Authentifikationsobjekt auf einer biologischen Eigenschaft oder auf einer bestimmten Aktion einer Person beruht, ist dieses abhängig von der körperlichen und mentalen Verfassung der Person. Diese Abhängigkeit kann sich beispielsweise einerseits durch schwächere Ausprägung des Merkmals oder auf der anderen Seite durch Schwierigkeiten bei der Bedienung des biometrischen Systems auswirken. Durch den aktuell anhaltenden demografischen Wandel, der vor allem in den Industrieländern zu verzeichnen ist, steigt der Anteil der älteren Menschen innerhalb relativ kurzer Zeit beachtlich. Verbunden mit der Technisierung und dem wachsendem Bedarf an Authentifikationionssystemen stellt sich zum einen die Frage, wie sich die Performanz biometrischer Systeme in Bezug auf die Authentifikation aber auch die Generierung biometrischer Hashes verändert, wenn der zeitliche Abstand zwischen der Erfassung der Referenzdaten und der Authentifikationsdaten steigt. Damit ist zu rechnen, wenn Personen über einen langen Zeitraum ein biometrisches System verwenden. Ein Hauptgrund, ältere Menschen und Auswirkungen der Alterung in der biometrischen Forschung zu berücksichtigen, ist, dass die menschliche Lebensspanne zusammen mit der Vitalität im Alter immer mehr zunimmt. Basierend auf der Betrachtung der europäischen Entwicklung der Demografie der letzten 64 Jahre und deren Voraussage für die kommenden 36 Jahre (siehe Tabelle 1.1, [Uni10]), kann man einen demografischen Wandel beobachten, der eine Alterung der Bevölkerung zur Folge hat. Entsprechend Tabelle 1.1 betrug beispielsweise 1950 der Anteil der ab 65-jährigen nur circa 8,22% der europäischen Gesamtbevölkerung. Im Verlauf der folgenden 50 Jahre steigerte sich dieser Betrag im Jahr 2000 auf 14,76%.

Tabelle 1.1: Anteil verschiedener Altersgruppen bezogen auf die Gesamtpopulation im Verlauf der Dekaden von 1950 bis 2050 für Europa ([Uni10])

Jahr	Population in Mio.	Altersgruppe 0-19	Altersgruppe 20-64	Altersgruppe 65-79	Altersgruppe 80+	Altersgruppe 65+
1950	547,287	34,59%	57,20%	7,11%	1,10%	8,22%
1960	603,852	33,71%	57,43%	7,53%	1,34%	8,86%
1970	655,879	33,59%	55,94%	8,86%	1,62%	10,48%
1980	692,872	30,28%	57,33%	10,37%	2,03%	12,39%
1990	720,498	27,68%	59,59%	9,93%	2,80%	12,73%
2000	726,778	24,53%	60,72%	11,80%	2,96%	14,76%
2010	738,197	21,21%	62,62%	11,99%	4,19%	16,18%
2020	744,177	21,11%	60,04%	13,69%	5,17%	18,85%
2030	741,232	20,88%	56,72%	16,41%	5,98%	22,40%
2040	731,826	20,37%	54,65%	17,18%	7,80%	24,98%
2050	719,256	21,00%	52,08%	17,59%	9,33%	26,92%

Eine weitere Zunahme der Menschen mit einem Alter über 64 Jahren wird für die folgenden 50 Jahre prognostiziert: Für das Jahr 2050 sagt die Studie einen Anteil von ca. 26,92% vorher. Basierend auf diesen Werten ist also von einer Verdreifachung des Anteils der älteren Menschen an der Gesamtbevölkerung innerhalb der untersuchten 100 Jahre auszugehen. Die Verschiebung der Altersstruktur ist ein weiteres Anzeichen für die zunehmende Alterung der Bevölkerung. Das zeigt sich neben der Zunahme des Durchschnittsalters und des Anteils an der Bevölkerung

auch in einer anteiligen Abnahme jüngerer Menschen. Zusätzlich zu dieser Verschiebung werden die Menschen deutlich älter. Wie in Tabelle 1.1 zu sehen ist, sind im Jahr 2000 circa 14,76% der 726,778 Millionen Einwohner 65 Jahre oder älter während der Anteil der unter 20-jährigen circa 24,53% beträgt. 50 Jahre später betragen die jeweiligen Anteile bei einer geschätzten Gesamtbevölkerungszahl von 719,256 Millionen 26,92% (65+) beziehungsweise 21,00% (0-19). Das bedeutet, dass sich im genannten Zeitraum die Rate der älteren Menschen nahezu verdoppeln wird und gleichzeitig die Anzahl der Menschen unter 20 um circa 14,5% zurück geht.

Die Vorhersagen der World Population Prospects ([Uni10]) zeigen ähnliche Veränderungen der Bevölkerungsstruktur für nahezu jedes Industrieland voraus. Basierend auf diesen Prognosen müssen ältere Menschen bei Design, Entwicklung, Evaluierung und Verwendung von biometrischen Systemen berücksichtigt werden. Da die biometrischen Systeme auf physischen Eigenschaften beziehungsweise typischem Verhalten von Personen basieren, sind sie abhängig von körperlichen und mentalen Veränderungen, die durch das biologische Altern hervor gerufen werden.

1.2 Aufgabenstellung

In dieser Arbeit entwickeln und untersuchen wir unterschiedliche neue Möglichkeiten, biometrische Systeme zu optimieren. Dabei werden wir die beiden Anwendungsfälle *biometrische Authentifikation* und *biometrische Hash-Generierung* zugrunde legen. Zum einen wird in diesem Zusammenhang der Biometric Hash-Algorithmus für die dynamisch Handschrift [Vie06] und zum anderen ein von uns entwickeltes Verfahren [MSV12], welches das Prinzip des Secure Sketch nutzt (siehe [DRS04]), erforscht. Im Folgenden wird auf die einzelnen Aufgaben, denen wir uns in dieser Arbeit stellen eingegangen.

A.1 Hinzufügen zusätzlicher Merkmale. Beide Systeme basieren auf demselben Set statistischer Merkmale, welches zu Beginn dieser Arbeit aus 69 Elementen bestand. Im Rahmen dieser Aufgabe sollen neue Merkmale entwickelt, implementiert und ihre Wirkung auf die Anwendungsszenarien für beide Algorithmen evaluiert werden. Das Hinzufügen neuer Merkmale passiert vorerst unkritisch ohne Überprüfung der Eignung der einzelnen Merkmale und ihrer möglichen Kombinationen für den jeweiligen Anwendungsfall.

A.2 Optimierung von Parametern. Als zweiter Ansatz zur Optimierung soll überprüft werden, welche Parameter die Algorithmen zur Verfügung stellen, die in beiden Anwendungsszenarien zu Verbesserungen führen können. Werden entsprechende Parameter identifiziert, wird für beide Verfahren je einer gewählt, jeweils entsprechend der zur Verfügung stehenden Daten und eingesetzten Methodologie ein optimaler Wert bestimmt und die anwendungsspezifische Eignung experimentell überprüft.

A.3 Analyse und Selektion von Merkmalen. Die Überprüfung der Eignung der einzelnen Merkmale für die Anwendungsmodi Authentifikation und Hash-Generierung der beiden Algorithmen ist Inhalt dieser Teilaufgabe. Ob ein einzelnes Merkmal oder Kombinationen aus mehreren zur Verbesserung der Ergebnisse beiträgt oder nicht soll mittels Verfahren zur

Merkmalsanalyse untersucht werden. Darauf aufbauend ist eine Selektion durchzuführen, welche diejenigen Merkmale ausschließt, die das Ergebnis negativ beziehungsweise gar nicht beeinflussen, abhängig vom jeweils verwendeten Algorithmus und Anwendungsszenario. Der Erfolg der Merkmalsanalyse und -selektion wird experimentell evaluiert.

A.4 Biometrische Fusion. Als weiterer Ansatz zur Optimierung ist in dieser Aufgabe die biometrische Fusion zu untersuchen. Ziel dabei ist, festzustellen, ob die Kombination einzelner biometrischer Komponenten im Kontext der Anwendungsszenarien und der genutzten Daten im Vergleich zur Performanz der einzelnen Komponenten zu besseren Ergebnissen führt. Dafür sind geeignete Verfahren einzusetzen und zu evaluieren.

Die vier oben genannten Optimierungsschritte werden nacheinander ausgeführt und die entsprechenden Resultate je nach Möglichkeit für die nachfolgenden Schritte übernommen. Das bedeutet im einzelnen, dass die in *A.1* umgesetzten und hinzugefügten Merkmale auch in den nachfolgenden Schritten untersucht werden. Ebenso fließen die in *A.2* ermittelten Parameter für die beiden Algorithmen sowohl in die Merkmalsanalyse und -selektion als auch in die Fusion mit ein. Eine Ausnahme bilden die Ergebnisse aus *A.3* die für die Fusion in *A.4* nicht berücksichtigt werden. Der Grund für diese Entscheidung liegt darin, dass die Analyse der Merkmale und ihre darauf aufbauende Selektion teilweise auf Verfahren beruhen, welche zur Optimierung den jeweiligen Algorithmus verwenden. Daher wären weitere Merkmalsanalysen basierend auf den unterschiedlichen Fusionsstrategien notwendig gewesen, ebenso die Durchführung der darauf aufbauenden experimentellen Evaluierungen.

B Alterung biometrischer Daten. Aufbauend auf den zur Verfügung stehenden Daten soll im Kontext von Authentifikation und Hash-Generierung überprüft werden, ob ein kurzer zeitlicher Abstand zwischen der Erfassung der Referenz- und Testdaten (hier circa ein Monat) nachweisbare Auswirkungen zur Folge hat. Diese Untersuchung ist experimentell durchzuführen.

Aufgrund der unterschiedlichen Konzipierung der Testsets für die Optimierung und die Alterung werden die Resultate der Untersuchungen zu *A.1* bis *A.4* nicht für die Untersuchung der Auswirkungen der Kurzzeitalterung herangezogen.

Die Eignung der Optimierungsansätze zur Verbesserung der Performanz beider Algorithmen in den zwei Modi Authentifikation und Hash-Generierung wird nach jedem Schritt experimentell evaluiert. Dazu werden zum einen die in der Biometrie gebräuchliche Fehlerraten Equal Error Rate, False Non Match Rate und False Match Rate (Authentifikation) herangezogen. Auf der anderen Seite haben wir zur Überprüfung der Auswirkung der Optimierungsschritte auf die Hash-Generierung mit Reproduction Rate, Collision Rate und Collision Reproduction Rate geeignete Maße entwickelt.

1.3 Wissenschaftlicher Beitrag

Der Fokus des wissenschaftlichen Beitrags liegt in der Optimierung von zwei biometrischen Hash-Algorithmen in Kontext der Betriebsmodi Hash-Generierung und Authentifikation. Dieser Abschnitt vermittelt die Struktur des Inhaltes der vorliegenden Arbeit im Zusammenhang mit jeweils einer kurzen Angabe des entsprechenden wissenschaftlichen Beitrages.

Kapitel 2 *Ausgewählte Grundlagen und Stand der Technik*

Im 2 Kapitel werden die wichtigsten Grundlagen und Begriffe aufgearbeitet, mit dem Ziel, das Verständnis des nachfolgenden Inhalts zu erhöhen. Es wird ebenfalls auf den Stand der Technik in den betreffenden Bereichen der Forschung eingegangen, um potentielle Forschungslücken aufzuzeigen und den Inhalt dieser Arbeit von denen Dritter abzugrenzen.

Kapitel 3 *Herausforderungen, Anforderungen und Lösungsansätze*

Das Kapitel 3 grenzt die zuvor identifizierten Forschungslücken ein, um die daraus resultierenden Heraus- beziehungsweise Anforderungen abzuleiten, die den wissenschaftlichen Rahmen für diese Arbeit bilden.

Kapitel 4 *Verwendete Algorithmen*

Für die vergleichende Untersuchung der Wirksamkeit der von uns unternommenen Optimierungschritte basierend auf der biometrischen Modalität der Handschrift verwenden wir zwei Algorithmen deren Grundlage dieselben Merkmale bilden. Eine detaillierte Beschreibung sowohl des in [Vie06] vorgestellten *Biometric Hash* als auch des von uns umgesetzten *Secure Sketch* für die biometrische Handschrift [MSV12] wird im Kapitel 4 gegeben.

Kapitel 5 *Ansätze zur Optimierung*

Wir beschäftigen uns in Kapitel 5 mit der Optimierung der im vorherigen Kapitel vorgestellten Algorithmen bezüglich der Anwendungsszenarien *biometrische Verifikation* und *biometrische Hash-Generierung*. Im Rahmen dieser Arbeit werden verschiedene Ansätze entwickelt, implementiert und evaluiert, mit dem Ziel, die Erkennungsgenauigkeit jeweils für beide Anwendungsszenarien zu verbessern. Unsere Ansätze zur Optimierung der Hash-Verfahren unterteilen wir in diesem Kapitel, wie in Abschnitt 1.2 eingeführt.

A.1 Hinzufügen zusätzlicher Merkmale. Das Hinzufügen neuer Merkmale erfolgt zunächst unkritisch, das heißt, ohne Betrachtung der potentiellen Wirkung eines Merkmals beziehungsweise seiner Kombination mit anderen auf den jeweiligen Anwendungsprozess. Im Abschnitt 5.1 beschreiben wir die 62 neuen Merkmale formal, grafisch beziehungsweise durch Pseudocode.

A.2 Optimierung von Parametern. Beide Algorithmen bieten Werte zur Parametrisierung der Quantisierung, die dem Transfer der einzelnen Merkmale in den Hash-Raum zugrunde

liegt. In Abschnitt 5.2 wählen wir jeweils einen Quantisierungsparameter pro Algorithmus aus und definieren und erläutern die einzelnen Strategien zur Bestimmung des jeweils für die hier untersuchten Anwendungsszenarien optimalen Wertes.

A.3 Analyse und Selektion von Merkmalen. Die Optimierung basierend auf den 131 Merkmalen unterteilen wir in Abschnitt 5.3 in die beiden Teilbereiche Analyse der Merkmale mittels ausgewählter Verfahren (Wrapper und Filter) abhängig von den einzelnen Anwendungsszenarien und der entsprechenden Parametrisierung, gefolgt von der darauf aufbauenden Selektion von Merkmalen.

A.4 Biometrische Fusion. Für den Optimierungsansatz der biometrischen Fusion den wir in Abschnitt 5.4 untersuchen wird eine Auswahl solcher Strategien betrachtet, die lediglich auf der Kombination von Komponenten einer einzelnen biometrischen Modalität beruhen (*single-modal Fusion*). Wir beschränken uns in dieser Arbeit auf die biometrische Modalität der dynamischen Handschrift, die vorgeschlagenen Strategien lassen sich jedoch auch auf andere Modalitäten anwenden.

B Alterung biometrischer Daten. Neben der Optimierung betrachten wir auch den Einfluss der Alterung auf die Performanz der beiden Algorithmen. Dabei kann unter dem Term *Alterung* im biometrischen Sinne sowohl die physische als auch die mentale Veränderung des Menschen über längere Zeiträume, als auch der Zeitabstand zwischen einzelnen Datenaufnahmen (zum Beispiel Referenz- und Verifikationsdaten) verstanden werden. Wir adressieren an dieser Stelle letztere Interpretation des Begriffes Alterung im Kontext der Biometrie.

Kapitel 6 *Aufbau und Methodologie der experimentellen Evaluierung*

Die Vorstellung und Beschreibung der den Tests zugrunde liegenden Methodologie als auch des Aufbaus der verwendeten Datenbasen sowie der experimentellen Evaluierung stehen im Mittelpunkt von Kapitel 6.

Für die Einschätzung und den Vergleich der Algorithmen untereinander, sowie zur Verdeutlichung der Veränderungen durch die von uns entwickelten und adaptierten Optimierungsansätze, verwenden wir biometrische Fehlermaße.

Zum einen setzen wir die in der Biometrie gebräuchlichen Maße *False Acceptance Rate* und *False Rejection Rate* ein. Aus diesen beiden Raten kann die *Equal Error Rate* bestimmt werden, welche als normalisierter Vergleichswert zur Messung der Performanz der Verifikation genutzt werden kann.

Auf der anderen Seite kommen bei der Einschätzung und zum Vergleich der Performanz der Hash-Generierung die von uns zu diesem Zweck entwickelten Maße *Reproduction Rate*, *Collision Rate* [SVD08a] und *Collision Reproduction Rate* [MSV11a] zum Einsatz. Diese ermöglichen es, die Algorithmen nicht nur bezüglich der Verifikation, sondern auch in Hinblick auf die Hash-Generierung zu betrachten, zu optimieren und zu evaluieren.

Kapitel 7 *Präsentation und Diskussion der experimentellen Ergebnisse*

In Kapitel 7 überprüfen wir in welchem Umfang die Verfahren zur Optimierung beider Algorithmen für beide Anwendungsszenarien erfolgreich waren. Grundlagen dafür liefern experimentelle Evaluierungen entsprechend der in Kapitel 6 *Aufbau und Methodologie der experimentellen Evaluierung* vorgestellten Maße, Testsets und Methodologien. Unterteilt ist das Kapitel entsprechend der zu untersuchenden Aufgabenstellungen.

A.1 Hinzufügen zusätzlicher Merkmale. Das Hinzufügen zusätzlicher Merkmale ist ein häufig eingesetztes Mittel zur Optimierung biometrischer Verfahren. Wir zeigen durch die experimentelle Evaluierung, dass sich in unserem Testaufbau für die Verifikation in den meisten Fällen Verbesserungen erzielen lassen. So kann zum Beispiel für den Biometric Hash-Algorithmus im besten Fall die Equal Error Rate von 0,16311 (69 Merkmale) auf 0,12512 (131 Merkmale) um circa 23% gesenkt werden. Die beste Verringerung der Equal Error Rate basierend auf dem Secure Sketch-Algorithmus beträgt etwa 38% für die Senkung von 0,19680 mit 69 Merkmalen auf 0,12141 mit 89 Merkmalen.

A.2 Optimierung von Parametern. Beide Algorithmen bieten Werte zur Parametrisierung der zugrunde liegenden Quantisierung, welche die Hash-Generierung beeinflussen, deren Auswirkungen auf beide Modi mit dem Ziel der jeweiligen Optimierung untersucht werden sollen. Durch die Anpassung der Parameter kann in Abhängigkeit vom Algorithmus für die Verifikation eine Verbesserung von bis zu 41% (Biometric Hash) beziehungsweise 55% (Secure Sketch) erreicht werden. Auch im Hash-Generierungmodus werden signifikante Verbesserungen beider Algorithmen erreicht. Diese liegen bei ca. 82% (Biometric Hash) beziehungsweise 65% (Secure Sketch) basierend auf den besten Ergebnissen für die Collision Reproduction Rate.

A.3 Analyse und Selektion von Merkmalen. Zusätzlich zum Hinzufügen neuer Merkmale wird eine anschließende Analyse der gesamten Merkmalsmenge mittels ausgewählter Wrapper und Filter und eine darauf aufbauende Selektion von Merkmalen in Abhängigkeit von den beiden Betriebsmodi und deren Parametrisierung durchgeführt und experimentell evaluiert. Ein interessantes Ergebnis der experimentellen Evaluierung zur Analyse und Selektion der Merkmale ist, dass es zum Teil möglich ist, auch mit wenigen Merkmalen gute Ergebnisse zu erhalten. So ermittelt der *simple* Wrapper 22 Merkmale für den Biometric Hash-Algorithmus ohne Parametrisierung, bei deren Verwendung die Equal Error Rate um circa 29% gesenkt werden kann. Dabei sinkt die EER von 0,12512 auf 0,08850. Zusätzlich zur experimentellen Untersuchung der Auswirkungen der Wrapper- und filterbasierten Merkmalsanalyse und -selektion führen wir Betrachtungen durch, wie häufig beziehungsweise in welcher Reihenfolge Merkmale durch die einzelnen Verfahren selektiert werden. Ebenso untersuchen wir, wie häufig das beste Evaluierungsergebnis auf den einzelnen Verfahren basiert, jeweils für beide Algorithmen, verschiedene Schreibinhalte und unterschiedliche Parametrisierungen. Eine wichtige Erkenntnis dieses Abschnittes liefert die Betrachtung der Häufigkeiten, mit denen Wrapper und Filter das jeweils beste Ergebnis ermitteln. Dabei

zeigt sich sowohl für beide Algorithmen, unterschiedliche Schreibinhalte als auch für jeweils fünf verschiedene Parameter-Sets, dass die Wrapper deutlich erfolgreicher sind als die Filter.

A.4 Biometrische Fusion. Bei der Fusion wird eine Auswahl solcher Strategien betrachtet, die lediglich auf der Kombination von Komponenten der biometrischen Modalität der Handschrift beruhen (*single-modal Fusion*). Die entsprechenden Gewichte werden für die zu untersuchenden Strategien berechnet. Die Fusionsergebnisse werden experimentell ermittelt und diskutiert. Basierend auf dem Secure Sketch-Algorithmus und der Kombination von zwei unterschiedlichen Schreibinhalten kann beispielsweise die EER signifikant verringert werden. Diese beträgt für die multi-instance Fusion 0,03002, wobei die Einzelergebnisse für die zugrunde liegenden geschriebenen Inhalte *Symbol* und *Woher* bei 0,07259 beziehungsweise 0,06817 liegen.

B Alterung biometrischer Daten. Die Ergebnisse der Evaluierung im Kontext der Kurzzeitalterung von biometrischen Referenzdaten zeigen basierend auf unseren Testdaten und der genutzten Methodologie einen deutlichen Effekt. Es ist festzustellen, dass sich bereits nach kurzer Zeit (ein beziehungsweise zwei Monaten) die Performanz der Verifikation verschlechtert. Dabei erhöht sich beispielsweise die Equal Error Rate im ungünstigsten Fall auf mehr als das fünffache (nach einem Monat) beziehungsweise das bis zu siebenfache (nach zwei Monaten).

Kapitel 8 *Fazit und Ausblick*

In diesem Kapitel fassen wir die wichtigsten Erkenntnisse und Ergebnisse, die im Rahmen dieser Arbeit entstanden sind, zusammen. Weiter diskutieren wir deren Auswirkung auf andere Bereiche der Informatik. Anschließend geben wir einen Überblick über Arbeiten, die sich aus den hier beschriebenen Forschungen ergeben beziehungsweise durch diese ermöglicht werden.

In den Anhängen sind Informationen zu den verwendeten statistischen Merkmalen der beiden untersuchten Hash-Algorithmen (Anhang A *Statistische Merkmale*), sowie zur Hardware und Software, die für die Evaluierung verwendetet wurde (Anhang B *Verwendete Software und Hardware*) und tabellarische beziehungsweise grafische Darstellungen der relevanten Testergebnisse (Anhang C *Evaluierungsergebnisse*) zu finden.

2 Ausgewählte Grundlagen und Stand der Technik

Der Inhalt dieser Arbeit basiert hauptsächlich auf den zwei bekannten Mechanismen der IT-Sicherheit *Authentifikation* und *Kryptographie*. Die Biometrie als Teilbereich der Benutzerauthentifikation fokussiert die sichere Feststellung (Identifikation) beziehungsweise Überprüfung (Verifikation) der Identität von Personen und dient damit der Sicherung der Authentizität. Durch die Nutzung von biometrischen Daten, die in vielen Fällen personenbezogene beziehungsweise -beziehbare Informationen enthalten, entsteht der Bedarf, diese Daten vor dem Zugriff Unbefugter zu schützen. Für Dritte interessante Informationen können beispielsweise der gesundheitliche Zustand (gegebenenfalls ablesbar aus einem Bild des Gesichtes oder der Hand) oder spezifische Angaben zur Person wie die eigenhändige Unterschrift sein. Zum Schutz biometrischer Daten eignen sich unter anderem weit verbreitete Mechanismen der Kryptographie, wie beispielsweise die Verschlüsselung der kritischen Informationen. Ebenso ist die Transformation der biometrischen Daten mittels einer Hash-Funktion in einem anderen Raum möglich. Die zur Authentifikation durchzuführenden Vergleiche finden dann auf Basis von Hash-Werten statt, ähnlich, wie dies schon seit einiger Zeit bei verschiedenen Betriebssystemen üblich ist. Ziel dieses Vorgehens ist der Schutz der Referenzdaten vor Missbrauch (engl. Template Protection) und dient somit der Wahrung der Vertraulichkeit. Denkbar ist in diesem Kontext ebenfalls die Kombination von Biometrie und Kryptographie.

Ziele dieser Arbeit sind zum einen das Optimieren zweier biometrischer Hash-Algorithmen unter Berücksichtigung der Anwendungsbereiche Authentifikation und Hash-Generierung und andererseits die Betrachtung der Auswirkung der Alterung biometrischer Daten auf eben diese. In diesem Kapitel werden ausgewählte Grundlagen in den Bereichen IT-Sicherheit, insbesondere der biometrischen Benutzerauthentifikation und des biometrsichen Hashing, erläutert, die zum Verständnis dieser Arbeit erforderlich sind (Abschnitt 2.1). Um eine Abgrenzung zwischen den von Dritten publizierten Erkenntnissen und eigenen Resultaten zu ermöglichen, wird im Anschnitt 2.2 auszugsweise der für diese Arbeit wesentliche Stand der Technik beleuchtet.

2.1 Ausgewählte Grundlagen

Mit dem Begriff Sicherheit wird im deutschen Sprachgebrauch der Schutz vor Gefahr beziehungsweise deren Abwesenheit verbunden. Gefahren können beispielsweise nicht geplante Ereignisse oder ungewollte, aber auch beabsichtigte Handlungen von Personen darstellen. Unterschieden werden können hierbei zwei Arten zur Erzeugung von Gefahren: Einerseits gibt es Gefahren,

die unbeabsichtigt und zufällig entstehen. Dazu zählen natürliche oder durch Fahrlässigkeit ausgelöste Ereignisse. Auf der anderen Seite existieren Gefahren, die beabsichtigt herbeigeführt werden, um gezielt Schaden anzurichten. Die vor Angriffen zu schützenden Ziele können in zwei Klassen eingeteilt werden, materielle und immaterielle Schutzgüter. Zu den materiellen Schutzgütern zählen zum Beispiel das eigene Leben, die IT-Technik, Gebäude oder die Natur. Daten, Informationen oder auch das persönliche Ansehen können den immateriellen Schutzgütern zugeordnet werden.

In der englischen Sprache werden in diesem Zusammenhang zwei verschiede Begriffe verwendet, um Art und Ziel eines Angriffes besser unterscheiden zu können: *Security* und *Safety*. Dabei bezieht sich einerseits der Begriff Safety auf die Zuverlässigkeit eines Systems, beispielsweise bezüglich dessen Ausfallsicherheit. Auf der anderen Seite wird der Begriff Security verwendet, wenn es um den Schutz eines Systems vor beabsichtigten Angriffen geht. Zum Beispiel kann ein Sicherheitsgurt (engl. safety belt) oder ein Sicherheitsnetz (engl. safety net) dem Begriff Safety zugeordnet werden. In den Bereich Security können Wachpersonal (engl. security guard) oder Software zur Vermeidung und Beseitigung von Schadsoftware (engl. security suite) eingeordnet werden. Allerdings können beide Begriffe nicht immer vollkommen unabhängig voneinander betrachtet werden, da sie sich teilweise auch gegenseitig bedingen können. Beim Studium von Fachliteratur muss darauf geachtet werden, dass einige Autoren diese Unterscheidung gar nicht oder eventuell anders vornehmen.

Zur Abschätzung potentieller Angriffe wurden unterschiedliche Schutzziele entwickelt, welche in der Fachliteratur als Sicherheitsaspekte bezeichnet werden (siehe zum Beispiel [Dit00]): Integrität, Authentizität, Verfügbarkeit, Verbindlichkeit (auch Nachweisbarkeit oder Nicht-Abstreitbarkeit), Vertraulichkeit und Privatsphäre. Maßnahmen zur Erlangung beziehungsweise Erhaltung eines bestimmten Sicherheitsniveaus erfüllen einen oder mehrere dieser Sicherheitsaspekte. Eine wichtige Rolle im Rahmen dieser Arbeit spielt die automatische Sicherstellung der Authentizität von Personen, auf die im nachfolgenden Abschnitt genauer eingegangen wird.

2.1.1 Benutzerauthentifikation

In der heutigen Zeit spielt die Authentifikation von Personen und Informationen eine bedeutende Rolle. Ausschlaggebend dafür ist die Tatsache, dass durch die zunehmende Technisierung Kommunikationspartner oft sehr weit voneinander entfernt sind. Dadurch besteht der Bedarf, sicher zu stellen, dass der jeweilige Partner derjenige ist, der er vorgibt zu sein. Dies kann durch verschiedene Mechanismen erreicht werden, die jeweils Vor- und Nachteile aufweisen.

Bishop beschreibt die Authentifikation als die Bindung einer Identität an ein Subjekt [Bish2005]. Weiter gibt er die folgenden fünf Komponenten für ein System zur Benutzerauthentifikation an:

1. Eine Menge von aktuellen Authentifizierungsdaten dient als Grundlage für spezielle Informationen, mit der eine Person ihre Identität nachweisen will.

2. Eine Menge von Referenzdaten ist im System gespeichert und wird zur Überprüfung der

Identität genutzt.

3. Eine Menge von Funktionen, die die Authentifizierungsdaten in das gleiche Format wie die Referenzdaten transferiert, um einen Vergleich zu ermöglichen.

4. Eine Menge von Authentifizierungsfunktionen, die die Identität überprüfen.

5. Eine Menge von Funktionen, die es einer Person erlauben, Authentifizierungs- beziehungsweise Referenzdaten zu erzeugen oder zu ändern.

Abbildung 2.1 stellt das allgemeine Schema eines Authentifikationsprozesses dar. Dieser besteht aus vier Prozessmodulen: Datenerfassung, Merkmalsextraktion, Vergleich und Entscheidung. Zusätzlich verfügt das System über eine Datenbank zur Speicherung der Referenzdaten. Ein Versuch zur Authentifikation läuft dabei folgendermaßen ab: Zum Zeitpunkt der Datenaufnahme wird das Authentifikationsobjekt einem entsprechenden Sensor des Systems präsentiert (zum Beispiel Passwort über Tastatur, Schriftprobe über Grafik-Tablet). Die dabei aufgezeichneten Sensordaten werden dann an die Merkmalsextraktion übergeben. Hier werden aus den Daten Merkmale berechnet, die die Grundlage für den Vergleich darstellen. Im Vergleich werden die zuvor ermittelten Merkmale des aktuell präsentierten Authentifikationsobjektes mit den im System hinterlegten Daten (Referenzdaten, Template) aus der Referenzdatenbank verglichen. Basierend auf dem Ergebnis dieses Vergleichs wird eine Entscheidung bezüglich der Authentizität getroffen. Die derzeit gebräuchlichsten Authentifikationsverfahren basieren auf geheimem Wissen (*„Was weiß ich?“*), persönlichem Besitz (*„Was habe ich?“*) und Biometrie (*„Was bin ich?“* oder *„Was kann ich?“*). In den folgenden Abschnitten werden diese drei unterschiedlichen Möglichkeiten der Benutzerauthentifikation beschrieben.

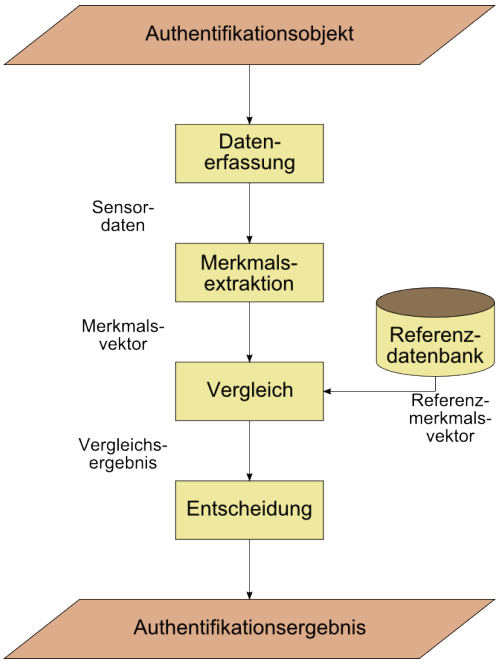


Abbildung 2.1: Allgemeines Schema eines Authentifikationsprozesses

Authentifikation durch geheimes Wissen

Authentifikation durch geheimes Wissen (auch wissensbasierte Authentifikation) nutzt Informationen, die nur der autorisierten Person (beziehungsweise einem eingeschränkten Personenkreis) bekannt sind. Gebräuchliche Beispiele aus der Praxis ist die Nutzung von persönlichen Identifikationsnummern (PINs) oder Passwörtern. Diese geheimen Informationen muss sich der Nutzer merken und bei Aufforderung an das System übergeben (zum Beispiel eine PIN über die Tastatur). Das Geheimhalten des Wissens ist dabei zur Wahrung der Sicherheit des entsprechenden Systems unabdingbar. Um eine Überwindung des auf geheimem Wissen basierenden Systems zusätzlich zu erschweren, müssen entsprechende Maßnahmen ergriffen werden. Bei der Wahl des geheimen Wissens zählen dazu beispielsweise Vorgehensweisen, die das Ermitteln der Information sowohl durch zufälliges als auch systematisches Probieren oder auch durch Social Engineering vermeiden. Dies kann zum Beispiel durch folgende Eigenschaften gewährleistet werden: eine möglichst große Auswahl an Zeichen, die Sonderzeichen und Zahlen enthält, nicht in einem Wörterbuch zu finden und unabhängig vom Umfeld des Besitzers ist. Außerdem sollte das zur Authentifikation genutzte geheime Wissen in regelmäßigen Abständen erneuert werden. Diese Anforderungen führen jedoch dazu, dass sich der Nutzer unter Umständen mehrere lange Zeichenkombinationen für diverse Anwendungen merken muss, die dazu noch häufig wechseln sollten. Die Nutzung des geheimen Wissens ist jedoch für die automatische Authentifikation die technisch am einfachsten umsetzbare Variante. Nicht zuletzt dadurch findet sie eine große Verbreitung, zum Beispiel bei der Anmeldung bei Betriebssystemen oder bei Online Shops. Dabei darf jedoch ein gravierender Nachteil nicht vernachlässigt werden: Geheimes Wissen kann durch Vergessen verloren gehen oder durch Verraten weitergegeben werden, auch ein Ausspähen durch Beobachten oder durch heimliches Protokollieren ist möglich. Ein wesentlicher Vorteil der Authentifikation durch geheimes Wissen besteht darin, dass bei beabsichtigter/unbeabsichtigter Offenlegung durch die Sperrung der kompromittierten und Erstellung einer neuen geheimen Information die Sicherheit des betroffenen Systems mit verhältnismäßig geringem Aufwand wiederhergestellt werden kann.

Authentifikation durch persönlichen Besitz

Der persönliche Besitz eines Authentifikationsobjektes ermöglicht die Authentifizierung einer Person über einen physischen Gegenstand, wie beispielsweise Schlüssel oder SmartCard, um den Zugang zu gesicherten Bereichen, wie nicht öffentliche Räume oder Computersysteme, zu gewähren. Der Nutzer muss durch das Präsentieren des physischen Gegenstandes seine Berechtigung nachweisen. Eine Gefährdung des zu sichernden Objektes ist durch den Verlust (zum Beispiel durch Diebstahl oder Verlieren) des Gegenstandes möglich. Ähnlich wie beim geheimen Wissen kann die Gültigkeit weitergegebener, verlorener oder gestohlener physischer Gegenstände aufgehoben werden. Der betroffenen Person kann dann ein neuer Gegenstand zur Authentifizierung zur Verfügung gestellt werden. Beim Notieren von geheimem Wissen ist zu beachten, dass es sich dabei nicht mehr um geheime Informationen handelt, sondern vielmehr um persönlichen Besitz, der entsprechend zu behandeln ist.

Ein großer Nachteil bei der Nutzung von geheimem Wissen als auch von persönlichem Besitz zur Benutzerauthentifikation ist, dass nicht überprüft werden kann, ob sich gerade eine bestimmte Person gegenüber einem System authentifiziert. Es wird vielmehr nur festgestellt, ob das präsentierte Authentifikationsobjekt (geheimes Wissen beziehungsweise persönlicher Gegenstand) den Überbringer zum Zugang zum gesicherten Bereich berechtigt oder nicht. Das hat zur Folge, dass die Identität der Person nicht überprüft werden kann, die Zugang verlangt.

Biometrische Benutzerauthentifikation

Anders verhält es sich bei der Verwendung von biometrischen Eigenschaften zur Benutzerauthentifikation. Diese Charakteristiken, auch biometrische Modalitäten genannt, sind entweder fest mit dem Körper (statisch, offline, physisch) oder dem Verhalten (dynamisch, online, verhaltensbasiert) eines Menschen verbunden. Bekannte Beispiele sind hier die Nutzung des Fingerabdruckes, der Iris (beide statisch), der Sprache (dynamisch) oder der Handschrift (sowohl statisch als auch dynamisch). Die biometrischen Verfahren basieren also auf der Tatsache des Vorhandenseins der berechtigten Person, wogegen bei den zwei oben genannten Verfahren nur die Präsenz eines bestimmten Wissens oder Gegenstandes überprüft werden kann.

Die in den vorangegangenen Abschnitten angeführten Nachteile von Authentifikationsobjekten, die geheimes Wissen beziehungsweise persönlichen Besitz nutzen, zeigen deutlich, dass eine Gefährdung der Sicherheit hauptsächlich auf Vergessen, Verlieren, Diebstahl beziehungsweise Weitergabe basieren. Dieses Risiko soll durch die Verwendung biometrischer Authentifikationsobjekte vermieden werden.

Im Folgenden werden zum besseren Verständnis wichtige Grundbegriffe der Biometrie definiert, die in dieser Arbeit häufig verwendet werden.

Modalität (engl. Modality). Als biometrische Modalität bezeichnet man eine physiologische Eigenschaft oder ein typisches Verhalten einer Person, welches zur biometrischen Authentifikation herangezogen wird (zum Beispiel Fingerabdruck, Handschrift).

Merkmal, Merkmalsvektor (engl. Feature, Feature Vector). Der Term Merkmal bezeichnet in der Biometrie eine einzelne numerische Größe, die aus einer biometrischen Modalität durch die Merkmalsextraktion bestimmt wird (zum Beispiel Schreibdauer einer Schriftprobe, Anzahl oder Art der Minutien eines Fingerabdrucks). Ein Merkmalsvektor entsteht durch die Zusammenfassung mehrerer Merkmale einer biometrischen Modalität.

Enrollment. Der Operationsmodus eines biometrischen Systems, innerhalb dessen eine Person diesem ihre biometrischen Informationen als Referenzdaten zur Verfügung stellt, wird als Enrollment bezeichnet. Die Hinterlegung von Referenzen ist bei einer Vielzahl von biometrischen Systemen notwendig, da darauf basierend die erforderlichen Vergleiche während der Authentifikation durchgeführt werden. Aus Gründen des Datenschutzes sollten Referenzdaten so gespeichert werden, dass aus ihnen allein kein Rückschluss auf die Person gezogen werden kann. In Abbildung 2.2 ist der Enrollment-Prozess zur Generierung der

Referenzdaten schematisch dargestellt. Die Daten der präsentierten biometrischen Modalität werden mittels Sensor erfasst. Das Modul zur Merkmalsextraktion berechnet daraus einen Merkmalsvektor, welcher, verknüpft mit der Identität der Person, in der Referenzdatenbank gespeichert wird.

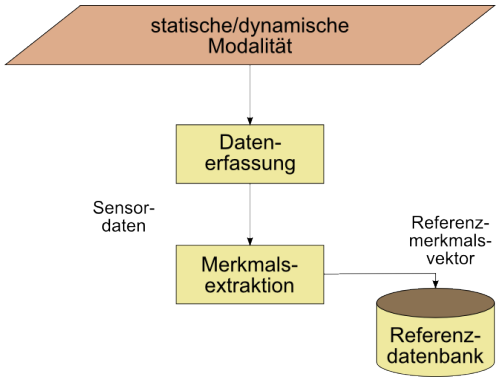


Abbildung 2.2: Schematische Darstellung eines Enrollment-Prozesses in einem biometrischen System

Authentifikation (engl. Authentification). Die Authentifikation stellt den zweiten Operationsmodus eines biometrischen Systems dar und dient der Überprüfung beziehungsweise Ermittlung der Identität einer Person, die zu einem gesicherten Bereich oder Computer-System Zugang verlangt. Die Überprüfung der Identität wird als *Verifikation* bezeichnet. Dabei wird sichergestellt, dass die Person diejenige ist, für die sie sich ausgibt. Dazu werden die aktuell aufgewiesenen Daten mit den in der Datenbank hinterlegten Referenzdaten verglichen, die sich auf den Identifikator der behaupteten Identität (zum Beispiel Benutzername oder ID) beziehen. Stimmen diese in einem ausreichendem Maße überein, wird die Identität der Person verifiziert. Im anderen Fall wird sie abgewiesen. Soll im Verlauf einer Authentifikation die Identität einer Person nicht bestätigt sondern ermittelt werden, spricht man von einer *Identifikation* . Hierbei werden die Eingabedaten der zunächst unbekannten Person mit den Referenzdaten aller dem System bekannten Personen verglichen. Im Falle einer innerhalb vorgegebener Schwellwerte ausreichenden Übereinstimmung mit einem vorhandenen Referenzdatensatz gilt die Person als identifiziert. Für beide Authentifikationsmodi wird, wie in Abbildung 2.3 dargestellt, die entsprechende biometrische Modalität während der Datenerfassung aufgenommen und aus den Sensordaten von der Merkmalsextraktion ein Merkmalsvektor generiert. Dieser wird beim Vergleich mit dem entsprechenden Merkmalsvektor aus der Referenzdatenbank abgeglichen. Das Ergebnis dieses Vorgangs ist ein einzelner Wert, welcher die Ähnlichkeit/Unähnlichkeit beider Vektoren zueinander ausdrückt und die Eingabe für die Entscheidung darstellt. Dieser gibt entweder an, ob die Verifikation erfolgreich war oder nicht (wahr/falsch) oder ob beziehungsweise als wer eine Person identifiziert wurde.

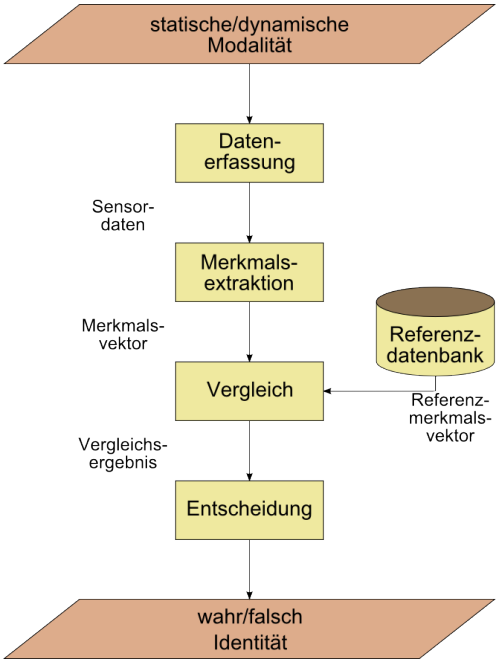


Abbildung 2.3: Schematische Darstellung eines Authentifikationsprozesses in einem biometrischen System

Um als biometrische Modalität beziehungsweise als biometrisches Merkmal für den Zweck der Authentifikation geeignet zu sein, müssen nach Jain et al. [JBP99] eine Reihe von Eigenschaften erfüllt sein. Diese Eigenschaften werden im folgenden aufgelistet und kurz erläutert.

Universalität (engl. Universality). Die Eigenschaft, dass eine biometrische Charakteristik bei möglichst vielen Menschen vorkommt, wird als Universalität bezeichnet. Demnach sollte sichergestellt werden, dass dieses Merkmal zumindest bei den meisten potentiellen Nutzern des geplanten biometrischen Systems vorhanden ist.

Einzigartigkeit (engl. Uniqueness). Trotz der geforderten Universalität ist es aber auch notwendig, dass eben dieses Merkmal bei jedem potentiellen Nutzer typische individuelle Ausprägungen annimmt, um eine ausreichende Trennschärfe zwischen einzelnen Personen zu erreichen. Diese Eigenschaft wird als Einzigartigkeit bezeichnet.

Konstanz (engl. Permanence). Die Eigenschaft der Konstanz verlangt, dass die der biometrischen Modalität zugrunde liegende Eigenschaft über einen langen Zeitraum unverändert bleibt beziehungsweise nur einer geringen Varianz unterliegt.

Messbarkeit (engl. Collectability). Einen wichtigen Aspekt für die Nutzung einer biometrischen Modalität stellt deren Messbarkeit dar. Dabei muss sichergestellt werden, dass die jeweilige biometrische Modalität bei jeder teilnehmenden Person in ausreichender Qualität und mit möglichst einfachen technischen Mitteln erfasst werden kann. Dies schließt natürlich vor allem die individuellen Eigenschaften mit ein, welche zur Unterscheidung zwischen den einzelnen Personen herangezogen werden.

Performanz (engl. Performance). Die Aufnahme der Daten und deren Weiterverarbeitung bis hin zur Erkennung einer Person sollte in einem zeitlich annehmbaren Rahmen und mit akzeptabler Hardware-Anforderung erfolgen.

Akzeptanz (engl. Acceptability). Die Akzeptanz eines biometrischen Systems umschreibt den Grad der Annahme durch die möglichen Nutzer. In den meisten Fällen wird die Eigenschaft durch die Höhe der Sicherheit und die Bedienbarkeit beeinflusst. Der Begriff der Sicherheit kann hier sowohl als diejenige Sicherheit interpretiert werden, welche durch das biometrische System erreicht wird, als auch als jene, die sich auf den Umgang mit den gespeicherten personenbezogenen beziehungsweise-beziehbaren Daten bezieht.

Sicherheit vor Angriffen (engl. Circumvention). Ein weiterer wichtiger zu beachtender Punkt ist, wie sicher eine Modalität oder ein Merkmal in Bezug auf potentielle Angriffe wie beispielsweise Fälschung oder Nachahmung ist.

Während die nicht biometrischen Verfahren zur Benutzerauthentifikation auf einer eindeutigen Übereinstimmung des benutzten Authentifikationsobjektes basieren, ist dies in der Biometrie nicht möglich. Der Grund dafür ist, dass eine biometrische Charakteristik einer Person von Aufnahme zu Aufnahme variiert. Das kann beispielsweise durch unterschiedliche Sensor Hardware, Verletzungen oder auch durch das aktuelle Befinden der Person hervorgerufen werden. Diese Art der Schwankung wird in der Biometrie als Intra-Klassen-Variabilität bezeichnet. Einen weiteren Grund stellt die so genannte Inter-Klassen-Ähnlichkeit dar, welche den Umstand beschreibt, dass sich biometrische Daten einer Modalität von zwei Personen sehr ähneln können. Beispielhaft kann hier die Ähnlichkeit bei sehr engen Verwandten wie Eltern und Kindern oder (Zwillings-) Geschwistern genannt werden.

Biometrische Modalitäten

Eine biometrische Authentifikation kann entweder auf physischen Eigenschaften, auf bestimmten Verhaltensweisen oder einer Kombination aus beidem basieren. Physische Eigenschaften die häufig für biometrische Zwecke eingesetzt werden sind Strukturen (zum Beispiel Haut, Iriden) oder bestimmte topografische Eigenschaften (zum Beispiel Gesicht, Ohr). Verhaltensweisen in der Biometrie sind zumeist Handlungen, die von einer Person bewusst (zum Beispiel Schrift, Sprache) oder zum Großteil unbewusst (zum Beispiel Gang, Mimik) durchgeführt werden. In der Biometrie wird eine einem Verfahren zugrunde liegende Charakteristik als biometrische Modalität bezeichnet.

Bei den *statischen biometrischen Modalitäten* (auch als passiv oder offline bezeichnet) werden physiologische Eigenschaften des Körpers einer Person (zum Beispiel Fingerabdruck, Gesicht) zur Authentifikation herangezogen. Die Erfassung einer statischen Modalität stellt üblicherweise eine Momentaufnahme dar, die zeitunabhängig durchgeführt werden kann. Die biometrische Information kann beispielsweise aus Foto- oder Videoaufnahmen extrahiert werden. Dieser Umstand ermöglicht die Erfassung der Daten in den meisten Fällen auch ohne bewusste Teilnahme des Inhabers, zum Beispiel zur Gesichtserkennung basierend auf Überwachungsvideos.

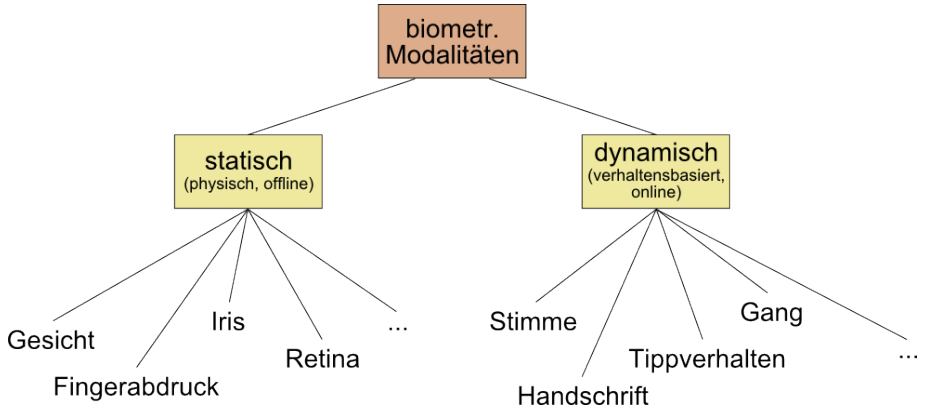


Abbildung 2.4: Klassifizierung biometrischer Modalitäten und einige Beispiele

Dynamische biometrische Modalitäten (auch als aktiv oder online bezeichnet) verwenden Informationen, die aus typischem Verhalten einer Person abgeleitet werden können (zum Beispiel Handschrift). Die Erfassung einer dynamischen Modalität ist demzufolge eine Beobachtung des Verhaltens über einen bestimmten Zeitraum hinweg. Damit bieten die meisten dynamischen biometrischen Modalitäten den Vorteil, dass die aktive und willentliche Mitwirkung der Person erforderlich ist. Abbildung 2.4 verdeutlicht diese Klassifizierung anhand einiger Beispiele.

Den im Rahmen dieser Arbeit durchgeführten Untersuchungen und Evaluierungen liegt ausschließlich die biometrische Modalität der dynamischen Handschrift zugrunde. Aus diesem Grund werden wir im folgenden nicht weiter auf andere Modalitäten eingehen und uns im anschließenden Abschnitt den Grundlagen der dynamischen Handschrift widmen.

2.1.2 Handschrift als biometrische Modalität

Haarmann umschreibt in [Haa07] die Schrift als *"Technologie, die sich der Mensch geschaffen hat, um Informationen für den Wiedergebrauch zu konservieren"*. Dieses Ziel haben die Menschen schon sehr früh verfolgt. Bereits um 5.000 v. Chr. verzierte die Donauzivilisation, eine vorindoeuropäische kupfersteinzeitliche Kultur des Balkans, Skulpturen und Kultgegenstände durch das Einritzen von Symbolen mit asymmetrischem Aussehen. Diese dienten nach Haarmann hauptsächlich der Verwendung zu religiösen Zwecken. Der Zweck der Informationsweitergabe und -speicherung spiegelt sich auch in Tontafeln mit Warenlisten und Aufrechnungen aus Mesopotamien um 3.200 v. Chr. wider. Um etwa dieselbe Zeit wurden in Ägypten Siegel mit Schriftzeichen verwendet, um Vorratsbehälter zu verschließen. Diese können beispielsweise als Möglichkeiten zur Authentifikation (des Inhaltes oder des Besitzers/Verschießenden) beziehungsweise zum Schutz der Integrität des Inhaltes interpretiert werden.

Im Laufe der Zeit hat sich die Schrift ihre Funktion zur Bewahrung und Weitergabe von Informationen erhalten. Ebenso ist die handgeschriebene Unterschrift auch heute noch wesentlicher Bestandteil geschäftlicher als auch privater Vorgänge. Dabei bestehen die Hauptfunktionen der Unterschrift zum einen in der Sicherstellung der Authentizität und andererseits dem Nachweis des willentlichen Einverständnisses des beziehungsweise der Unterzeichnenden.

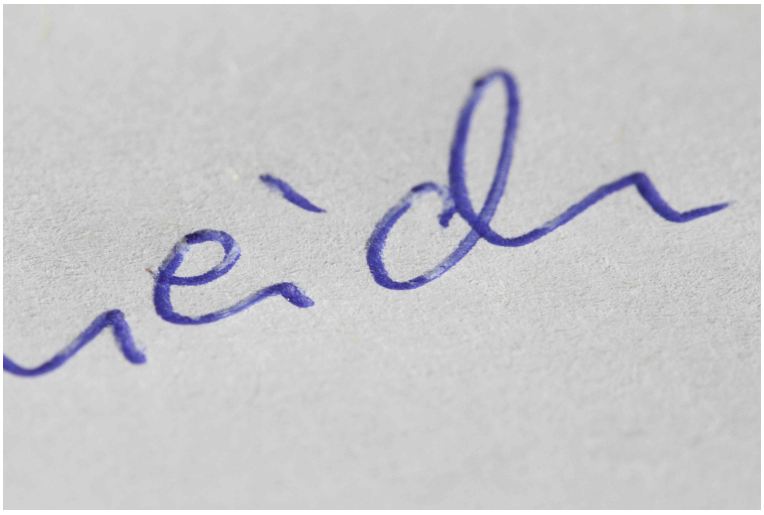


Abbildung 2.5: Ausschnitt einer Handschriftenprobe (Fotografie: Claudia Ihnen)

Die Handschrift einer Person weist eine Vielzahl von individuellen Eigenheiten auf, anhand derer der Urheber von anderen Menschen erkannt werden kann. Diese Individualität entsteht unter anderem durch physische und mentale Eigenschaften der jeweiligen Person aber auch durch persönliche Vorlieben bei der Gestaltung der eigenen Schrift. Typische Merkmale bilden neben der Verwendung von Druck- oder Schreibschrift die so genannte Oberlänge (beispielsweise beim kleinen H) beziehungsweise Unterlänge (beispielsweise beim kleinen G), die Neigung der Schrift oder der Schreibdruck. Die in Abbildung 2.5 dargestellte Handschrift verwendet zum Beispiel vorwiegend Schreibschrift und weist eine rechtslastige Neigung auf. Weiterhin sind die Schwankungen des aufgewendeten Drucks durch verschiedene Vertiefungen im Papier deutlich an mehreren Stellen sichtbar.

Merkmale für die biometrische Verwendung der Handschrift können sowohl aus dem Ergebnis des Schreibvorgangs (statische Repräsentation) als auch aus der Beschreibung des Schreibprozesses (dynamische Repräsentation) gewonnen werden. Als statische Repräsentation kann das Ergebnis nach der Vollendung des Schriftzuges angesehen werden. Für eine biometrische Untersuchung können darauf aufbauend unterschiedliche Merkmale bestimmt werden, zum Beispiel der durchschnittliche Neigungswinkel der Schrift, die Flächeninhalte bestimmter Bereiche oder die Relationen von unterschiedlichen Schnittpunkten zueinander. Während statische Methoden auf dem Ergebnis des Schreibprozesses, dem Schriftbild, basieren, betrachten dynamische Methoden den Schreibprozess selbst.

Bei der dynamischen Erfassung der Schrift werden bestimmte physische Merkmale der Schrift in Abhängigkeit von der Zeit t erfasst (siehe auch Abbildung 2.6). Bei aktuell verfügbarer Hardware sind dies vorwiegend:

- x(t): das horizontale Positionssignal des Stiftes,
- y(t): das vertikale Positionssignal des Stiftes und
- p(t): das Signal des auf das Tablet oder die Spitze des Stiftes ausgeübten Druckes.

Darüber hinaus ist spezielle Sensor-Hardware auch in der Lage, auf den Stift bezogene Winkel zu messen, zum Beispiel erfassen einige Grafik-Tablets zusätzlich:

- $\phi(t)$: Höhenwinkel des Stiftes über dem Tablet und/oder
- $\theta(t)$: Seitenwinkel des Stiftes über dem Tablet.

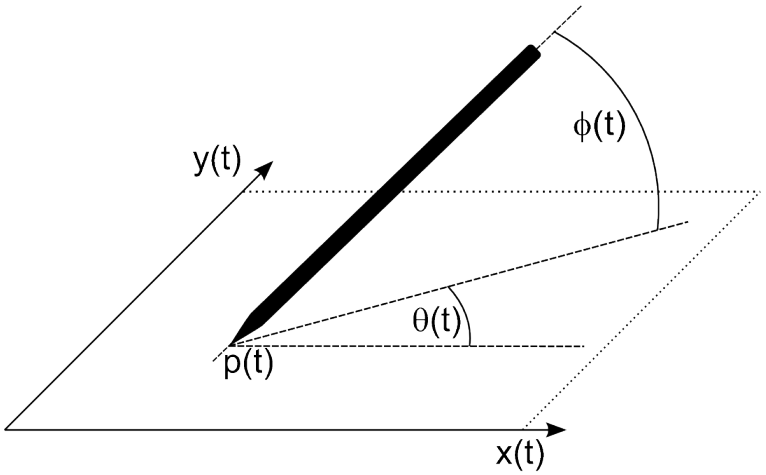


Abbildung 2.6: Signale eines aktuellen Handschrift-Sensors

Die automatisierte Weiterverarbeitung des an sich analogen Vorgangs des Schreibens erfordert das Sampling des Signals. Dabei wird dieses kontinuierliche Signal von der Hardware in eine Sequenz zeitdiskreter Werte umgewandelt. Während des Schreibprozesses werden zu definierten Zeitpunkten jeweils die physischen Information erfasst und gespeichert. In den meisten Fällen ist der zeitliche Abstand zwischen den einzelnen Erfassungspunkten identisch. Die auf die beschriebene Weise erfasste Sequenz von zeitdiskreten Stiftinformationen einer Schriftprobe wird in der Biometrie auch als *Sample* bezeichnet. Als das Ergebnis eines beispielhaften Sampling-Prozesses sind, zusammen mit dem Schriftzug, die Signale der Koordinaten X und Y und des ausgeübten Druckes in Abbildung 2.7 anhand einer beispielhaften Schriftprobe dargestellt. Deutlich zu erkennen ist der Anstieg von x während des Schreibvorganges (siehe Abbildung 2.7b), ein Indiz dafür, dass die Schreibrichtung von Links nach Rechts verläuft. Kleinere Ausschläge nach unten weisen auf kurze Bewegungen in die Gegenrichtung hin, beispielsweise bei den oberen Enden des großen *M* am Wortanfang oder beim Unterschwung des g am Ende des Wortes. Der Verlauf des y-Signals (siehe Abbildung 2.7c) weist eine gewisse Ähnlichkeit zum Schriftzug selbst auf, da dieser die Auf- und Abbewegungen des Stiftes mit steigendem t festhält. In Abbildung 2.7d sind im Druckverlauf gut die Absetzpunkte des Stiftes zu erkennen. Hier sinkt das Drucksignal stark ab, erreicht aber aufgrund einer gewissen Trägheit des Sensors nicht den Wert 0, bevor der Stift wieder aufgesetzt wird.

Tabelle 2.1 zeigt beispielhaft die Sequenz der ersten 27 Erfassungspunkte der Schriftprobe aus Abbildung 2.7. Basierend auf den Daten eines Samples können statische oder dynamische Werte bestimmt werden, die als Basis zur Unterscheidung von unterschiedlichen Nutzern beziehungsweise

zur Erkennung eines Nutzers dienen können. Typische statistische Werte für die dynamische Handschrift sind zum Beispiel die Schreibgeschwindigkeit, die Dauer des Schreibprozesses oder der Verlauf des Drucks. Im Folgenden wird auf unterschiedliche Hardware-Klassen zur Erfassung dynamischer Handschriftendaten eingegangen.

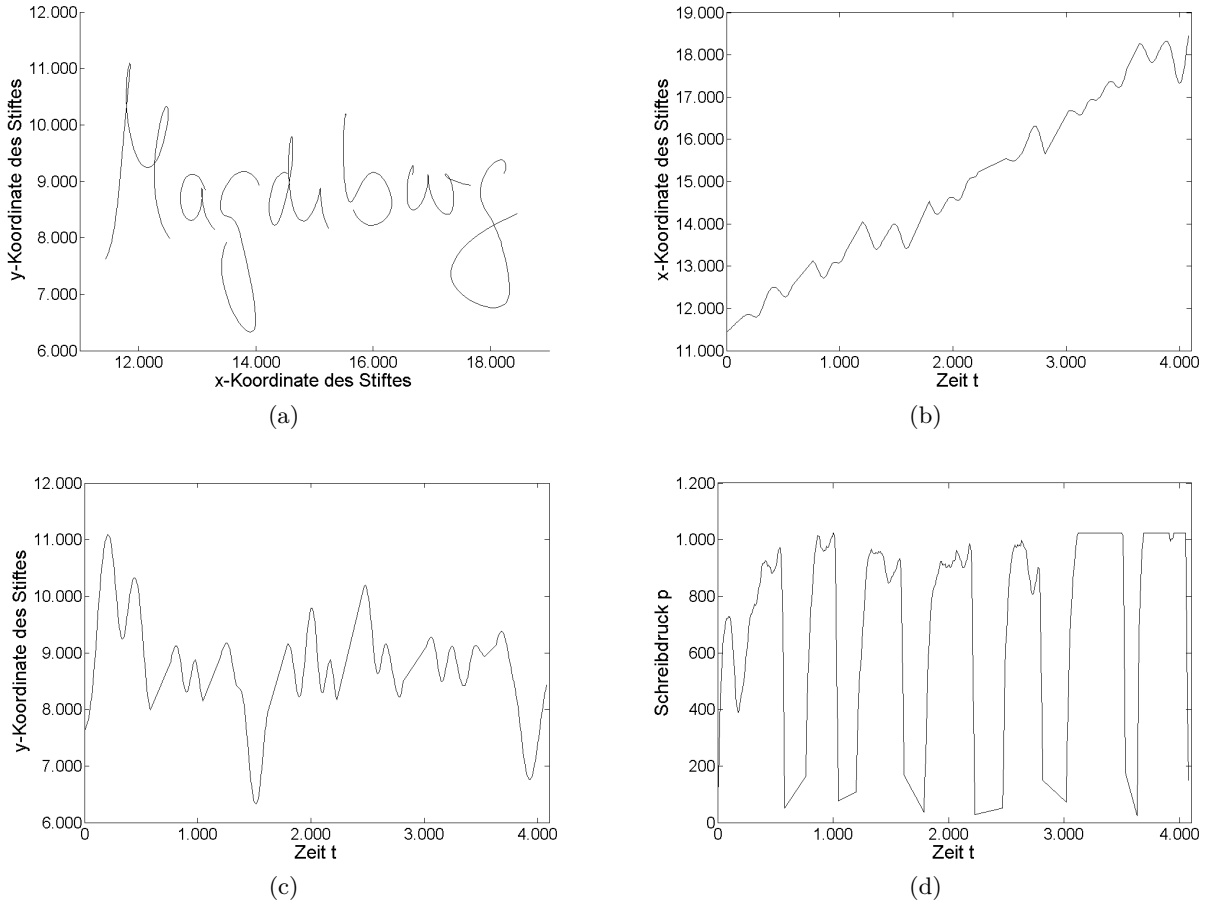


Abbildung 2.7: Darstellung des (a) Schriftbildes, (b) X-Signals, (c) Y-Signals und (d) Drucksignals

Tabelle 2.1: Auszug aus der Wertetabelle des in Abbildung 2.7 gezeigten Beispiel-Sample

Index	X	Y	t	P
0	11439	9798	0	64
1	11454	9764	7	153
2	11466	9740	13	253
3	11479	9707	19	350
4	11495	9666	25	439
5	11511	9615	32	519
6	11528	9554	38	571
7	11545	9482	44	616
8	11562	9398	50	656
9	11579	9301	57	672
10	11596	9192	63	692
11	11613	9070	69	708
12	11629	8936	75	720
13	11645	8792	82	720

Index	X	Y	t	P
14	11660	8634	88	724
15	11675	8469	94	728
16	11689	8296	100	724
17	11705	8118	107	728
18	11720	7937	113	708
19	11735	7757	119	684
20	11751	7580	125	656
21	11767	7409	132	632
22	11782	7247	138	588
23	11797	7094	144	543
24	11811	6953	150	499
25	11823	6825	157	451
26	11834	6712	163	422
...	...	...	...	...

Klassifizierung der Sensoren für die biometrische Handschrift

Es existiert eine Vielzahl von Sensoren, die die Erfassung von Daten für die biometrische Handschrift ermöglichen. Im Bereich der automatisierten Offline-Untersuchung der Handschrift können beispielsweise Digitalkameras oder auch Flachbett-Scanner eingesetzt werden. Diese ermöglichen eine Aufnahme des Schriftbildes in einer hohen Qualität und Geschwindigkeit. Da wir uns in dieser Arbeit mit der Betrachtung des Schreibprozesses beschäftigen, sind diese Geräte jedoch für unsere Zwecke ungeeignet. Wir benötigen, wie im vorherigen Abschnitt beschrieben, eine Sequenz von zeitabhängigen Erfassungspunkten. Im Folgenden geben wir eine Übersicht über verschiedene Sensoren, die in der Lage sind, entsprechende Daten für die Handschrift zu erfassen. Dazu haben wir die folgende Klassifikation basierend auf der verwendeten Technik entwickelt.

Funktionalität im Stift. Für die Erfassung dynamischer Handschriftendaten existieren spezielle Stifte, die die aufgezeichneten Rohdaten speichern oder direkt an den Computer übertragen. Einfache Modelle ermöglichen dabei die zeitabhängige Aufnahme von Position und Druck, neuere Generationen können zum Teil zusätzlich auch die Stiftwinkel erfassen. Der Vorteil dieser Stifte liegt darin, dass der Nutzer auf Papier schreiben kann, was seinem üblichen Schreibverhalten entspricht. Zusätzlich stehen ihm später zusätzlich zur digitalen Repräsentation die angefertigten Notizen in der gewohnten geschriebenen Form auf Papier zur Verfügung. Obwohl diese Stifte ursprünglich für die Aufzeichnung von handschriftlichen Notizen und Zeichnungen entwickelt wurden, konnten wir in [SV06] und [SVD07c] zeigen, dass sich die erfassten Daten auch zur biometrischen Authentifikation eignen. Wir haben in den Forschungsarbeiten, die diesen Publikationen zugrunde liegen, nachgewiesen, dass der Logitech iO Personal Digital Pen (siehe Abbildung 2.8, nachfolgend als IOPen bezeichnet) für die handschriftliche Verifikation verwendet werden kann und zum Teil bessere Ergebnisse bezüglich der Equal Error Rate erreicht werden können als mit Grafik- oder Signatur-Tablets.

Abbildung 2.8: Logitech iO Personal Digital Pen

Der IOPen arbeitet mit einem Verfahren, das von der schwedischen Firma Anoto [Ano11] entwickelt wurde. Grundlage ist ein spezielles Papier, das mit scheinbar willkürlich verteilten Punkten bedruckt ist (siehe 2.9a). Wie in Abbildung 2.9b dargestellt, haben diese Punkte eine Größe von 0,1 Millimetern und sind innerhalb eines Rasters von 0,3 Millimetern verteilt. In der

Anordnung der Punkte auf einem Blatt Papier sind verschiedene Informationen und Eigenschaften kodiert, die dieses einzigartig machen. Durch die Erfassung und Analyse der einzelnen Punkte und der Position im Raster und zueinander ist die Aufnahme folgender Informationen möglich:

- horizontale und vertikale Position des Stiftes auf dem Blatt,
- Ursprung des verwendeten Papiers (IDs der Seite, des Blockes und des Herstellers) und
- Verwendung spezieller Felder auf dem Blatt (zum Beispiel Checkbox zum Finalisieren der aktuellen Seite).

Zusätzlich zu diesen Daten stellt die Technik des Stiftes den auf die Stiftspitze ausgeübten Druck in 128 Stufen und einen Zeitstempel für jeden erfassten Aufnahmepunkt zur Verfügung.

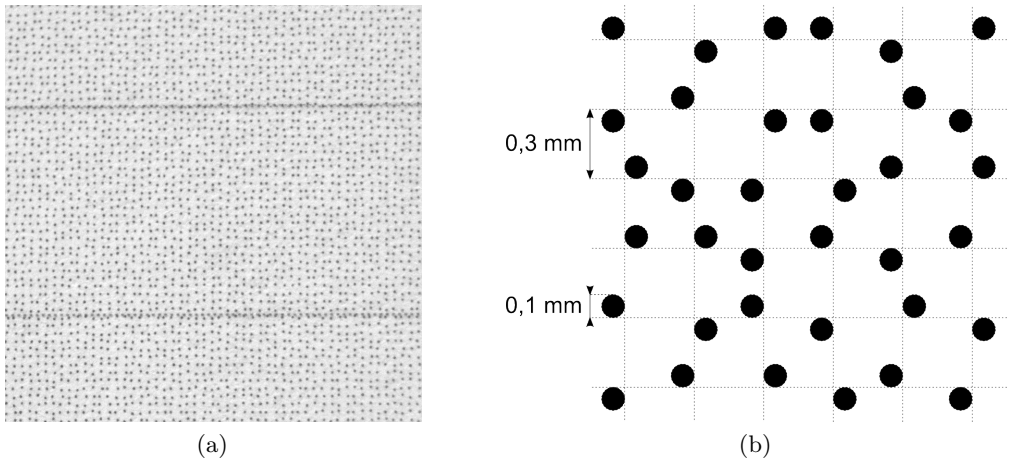


Abbildung 2.9: Beispielhafte Darstellung des Punktrasters, auf dem die Funktionsweise des IOPen beruht: (a)Foto eines Seitenausschnittes, (b) schematische Darstellung

Die Aufzeichnung dieser Informationen erfolgt über eine in den Stift integrierte Kamera, die die Punkte auf dem Papier innerhalb eines 6x6 Rasters erfassen kann. Aufgrund der minimalen Verschiebungen der Punkte innerhalb des Rasters ist eine eindeutige Bestimmung der aktuellen Position der Stiftspitze möglich. Die Position und der Druck werden zusammen mit dem Zeitstempel und den Metadaten zum Papier im internen Speicher des Stiftes abgelegt. Die Speicherkapazität des IOPen erlaubt die Erfassung von ca. 40 A4 Seiten. Bei Bedarf können die Daten über eine USB-Dockingstation an einen Computer übertragen werden.

Funktionalität im Tablet. Signatur-Tablets, als typische Vertreter dieser Kategorie, sind speziell für die Aufnahme dynamischer Daten von Unterschriften entwickelt worden. Sie sind in der Lage, die biometrischen Informationen beispielsweise über Druck- oder Ultraschallsensoren in der Tablet-Oberfläche zu erfassen. Dabei liegt der große Vorteil darin, dass nahezu jeder beliebige Stift verwendet werden kann und durch Dazwischenlegen des zu unterzeichnenden Dokumentes dieses zusätzlich direkt unterschrieben werden kann. Auf diese Weise kann die Echtheit einer Unterschrift auf einem Dokument auch im Nachhinein durch den direkten Vergleich mit

der digital hinterlegten Unterschrift durchgeführt werden. Vielhauer zeigt in [Vie06], dass sich Signatur-Tablets der Firma StepOver ([Ste14]) gut zur biometrischen Benutzerauthentifkation eignen.

Die Drucksensoren des StepOver blueM-Pad III (siehe Abbildung 2.10) ermöglichen zum Beispiel die Erfassung der Stiftposition und des ausgeübten Druckes auf die Stiftspitze. Die Größe der Schreibfläche beträgt ca. acht mal sechs Zentimeter. Dadurch steht auch für zur Unterschrift alternative Schreibinhalte wie Passwörter, PINs, Symbole oder auch kurze Passphrasen ausreichend Platz zur Verfügung. Bei der Übertragung über eine USB-Schnittstelle sind die Daten durch ein offenes Protokoll inklusive einer Prüfsumme gegen Angriffe gesichert.

Abbildung 2.10: StepOver blueM-Pad III

Auch Personal Digital Assistants (PDA) können als zu dieser Gruppe zugehörig betrachtet werden. Sie verfügen über ein druckempfindliches Display, welches mit einem Stift bedient werden kann. Ebenfalls können moderne Tablets oder Mobiltelefone mit Touch-Display hier eingeordnet werden. Allerdings verfügen diese in den meisten Fällen nicht über Möglichkeit der Aufnahme von Daten per Stift sondern sind für die Steuerung mittels Fingerdruck und -gesten ausgelegt.

Funktionalität in Tablet und Stift. Die meisten Sensoren dieser Klasse werden hauptsächlich zur Erstellung beziehungsweise Bearbeitung von Grafiken und Bildern eingesetzt. Die Herausforderungen dieses Anwendungsbereichs begründen auch die häufig sehr hohen Auflösungen der X-/Y-Koordinaten und des Druckes. Der zum Tablet gehörende Stift übermittelt mithilfe geeigneter Technik Informationen, die zur Bestimmung der Position, des Druckes und gegebenenfalls der Stiftwinkel beiträgt. Die erforderliche Verbindung zwischen Tablet und Stift kann dabei beispielsweise durch elektromagnetische Resonanz hergestellt werden. Dies ermöglicht zum einen eine exakte Bestimmung der aktuellen Position der Stiftspitze als auch die Übermittlung des vom Stift gemessenen Druckes, sowie gegebenenfalls die Bestimmung der Winkel. Hardware dieser Kategorie wird zunehmend in Verbindung mit Displays hergestellt. Dies ermöglicht einen nahezu natürlichen Schreibprozess, da hier der Schriftzug gut sichtbar an der Stiftspitze erscheint. Dieser Effekt kann bei einigen Sensoren auch durch den Einsatz von Tintenminen und aufgelegtem Papier erzielt werden. Tablet-PCs sind ein aktuelles Beispiel für diese Kategorie. Abbildung

2.11 stellt das Toshiba Pórtegé M200 dar, das zur Aufnahme der in dieser Arbeit verwendeten Daten genutzt wurde. Weiterführende Information zu diesem Sensor werden in Abschnitt 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank* gegeben.

Abbildung 2.11: Toshiba Pórtegé M200

Zusammenfassend ist in Tabelle 2.2 eine Zuordnung verschiedener zur Erfassung der Handschrift geeigneter Sensoren in die drei genannten Klassen dargestellt. Die gewählte Einteilung lässt eine eindeutige Zuweisung der Hardware zu einer bestimmten Klasse zu.

Tabelle 2.2: Klassifikation unterschiedlicher Sensoren für die dynamische Handschrift entsprechend der verwendeten Technik beziehungsweise dem ursprünglichen Verwendungszwecks

Sensor	Technik basiert auf...		
	Stift	Tablet	Tablet & Stift
Toshiba Tablet PC			x
StepOver Signature Tablet		x	
Logitech IO Pen	x		
PDA	x		
Graphic Tablet			x
Wacom Cintiq21			x
Digi Memo A5			x
Tablet		x	
Touch Screen		x	

2.1.3 Datensicherheit im Umgang mit biometrischen Informationen

Bei der Speicherung biometrischer Daten besteht das Problem, dass dies personenbezogene Informationen, zum Beispiel in Form eines Fingerabdruckes oder einer Unterschrift, beinhalten können. Bei den folgenden Erläuterungen nehmen wir Bezug auf das Datenschutzgesetz des Landes Sachsen-Anhalt (DSG LSA[1]). Das Gesetz regelt den Umgang mit personenbezogenen Daten

[1] Gesetz zum Schutz personenbezogener Daten der Bürger (Datenschutzgesetz Sachsen-Anhalt - DSG LSA) vom 12. März 1992 (GVBl. S. 152) in der Fassung der Bekanntmachung vom 18. Februar 2002 (GVBl. LSA S. 54), zuletzt geändert durch Artikel 1 des Gesetzes vom 27. September 2011 (GVBl. LSA S. 648)

insbesondere in Bezug auf Erhebung, Speicherung, Verarbeitung und Löschung entsprechender Informationen, um eine Beeinträchtigung des Rechts auf informationelle Selbstbestimmung und des Persönlichkeitsrechtes der einzelnen Person zu verhindern. Dabei gibt das DSG LSA keine speziellen Vorgaben zum Umgang mit biometrischen Daten. Allerdings sind dem Gesetz zufolge Informationen *„über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person“* personenbezogene Daten, *„Angaben über die rassische oder ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben“* sind personenbezogene Daten besonderer Art. Schätzt man den Schutzbedarf biometrischer Daten aufgrund dieser Aussagen ein, kann es sich bei diesen Informationen (in Abhängigkeit von der verwendeten Modalität) um personenbezogene Daten beziehungsweise personenbezogene Daten besonderer Art handeln. So lassen die erhobenen Daten einzelner Modalitäten direkt auf die Person schließen, beispielsweise Gesicht, Stimme oder Handschrift (speziell die Unterschrift), oder es sind Rückschlüsse auf die rassische beziehungsweise ethnische Herkunft möglich, zum Beispiel bei den Modalitäten Stimme und Gesicht. Informationen über den gesundheitlichen Zustand einer Person können zudem auch aus biometrischen Informationen abgeleitet werden, hier können ebenfalls Stimme und Gesicht exemplarisch genannt werden. Zusätzlich zu den biometrischen Daten kann es zu Forschungszwecken notwendig sein, so genannte Metadaten über die einzelnen Personen zu erfassen. Dabei kann es sich beispielsweise um Informationen bezüglich Alter, Geschlecht, Bildung oder eben auch rassische beziehungsweise ethnische Herkunft handeln. Metadaten können in der biometrischen Forschung eingesetzt werden, um Gruppen zu bilden, mit dem Ziel, bestimmte Regelmäßigkeiten innerhalb dieser Gruppen zu untersuchen. Denkbar wären hier Gruppen basierend auf dem Alter oder dem Geschlecht der Probanden. Gelingt es einem potentiellen Angreifer, neben den biometrischen Daten auch an die korrespondierenden Metadaten einer Person zu kommen, so stehen ihm zusätzliche wertvolle Informationen zur Fälschung beziehungsweise zum Nachbau der jeweiligen Identität zur Verfügung. So schränkt beispielsweise die Information über das Geschlecht oder die Zugehörigkeit zu einer bestimmten Altersgruppe den Kreis in Frage kommender Personen erheblich ein.

Im Vergleich mit den beiden traditionellen Methoden zur Authentifikation - geheimes Wissen und persönlicher Besitz - entstehen im Bereich der Biometrie ganz andere Anforderungen an den Datenschutz. Da biometrische Eigenschaften auf physischen beziehungsweise verhaltensbasierten Charakteristiken basieren, besteht die Möglichkeit, dass für biometrische Zwecke aufgenommene Daten schützenswerte Informationen über ihren Urheber enthalten. Diese Informationen spielen oft keine wesentliche Rolle für den automatisierten Authentifikationsprozess, werden jedoch aus technischen Gründen bei der Datenaufnahme ebenfalls erfasst. Beispielsweise existieren biometrische Systeme zur Gesichtserkennung, deren zugrunde liegende Algorithmen auf der Analyse der Relationen bestimmter Merkmale zueinander basieren. Dabei ist für das System die Hautfarbe des Benutzers irrelevant, diese wird jedoch naturgemäß bei der visuellen Erfassung der Person während der Enrollment- beziehungsweise Authentifikationsphase erfasst. Wird nun das Bild als Referenz in der Datenbank gespeichert, kann jeder, der Zugriff auf die Daten hat, sensible Informationen des Nutzers in Erfahrung bringen, hier zum Beispiel dessen ethnische Zugehörig-

keit. Die Gefahr des Missbrauchs steigt bei der Speicherung der Daten mehrerer biometrischer Modalitäten, beispielsweise für die Verwendung in multi-modalen Authentifikationssystemen, noch weiter an. Bleiben wir hier beim Beispiel der biometrischen Charakteristik Gesicht, bei der ein Foto oder Video einer unbekannten Person für eine Online-Bildersuche genutzt werden kann. Werden mehrere Personen gefunden, kann anhand der zusätzlichen Informationen, die durch weitere Modalitäten gegeben sind, der in Frage kommende Personenkreis weiter eingeschränkt werden. Ist die Person dann auf diese Weise identifiziert worden, lassen sich gegebenenfalls über deren eigene Website und gegebenenfalls Verlinkungen weitere Informationen (je nach Verfügbarkeit beispielsweise Name, Alter, Adresse, Beruf, Verwandte) zu einem detaillierten Profil zusammenfügen.

Abbildung 2.12 zeigt eine Darstellung potentieller Schwachstellen eines biometrischen Systems im Authentifikationsmodus als Erweiterung des in [RCB01] vorgestellten Schemas. Im Enrollment-Modus sind Schwachstellen analog vorhanden. Potentielle Angriffspunkte stellen der Transfer der Daten zwischen den verbundenen Komponenten als auch die Komponenten selbst dar. So ist es unter Umständen möglich, dass zwischen den einzelnen Komponenten Daten abgefangen und für Zwecke genutzt werden, für die sie ursprünglich nicht gedacht waren. Dies ist beispielsweise in einem Szenario denkbar, in dem sich die biometrischen Referenzdaten auf einen Token im Besitz des Nutzers befinden. Kommt ein versierter Angreifer in den Besitz der Karte, kann er die ungeschützten Referenzdaten auslesen. Handelt es sich dabei beispielsweise um eine Unterschrift beziehungsweise ein Gesicht, können diese für eine Vielzahl von betrügerischen Zwecken eingesetzt werden. Mögliche Risiken bezogen auf die Module des Systems sind zum Beispiel durch eine alternative Parametrisierung oder Manipulation der Funktionalität zu erreichen.

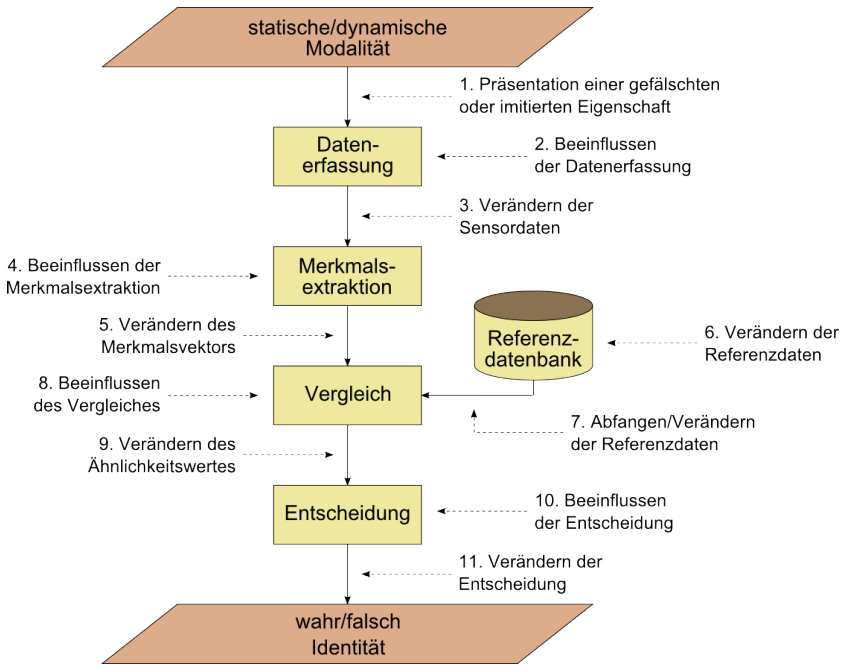


Abbildung 2.12: Schematische Darstellung eines Authentifikationsprozesses in einem biometrischen System und potentieller Angriffsmöglichkeiten (angelehnt an [RCB01])

Eine Möglichkeit, einige der in Abbildung 2.12 gezeigten Schwachstellen zu umgehen, bietet die Kombination der Biometrie mit der Kryptographie. In dieser Arbeit adressieren wir die Verwendung von biometrischen Hash-Funktionen. Dabei stehen sowohl der Schutz der biometrischen Referenzdaten im Zusammenhang mit deren Nutzung im Authentifikationsmodus, als auch die Generierung reproduzierbarer individueller Hash-Werte im Vordergrund. Im folgenden Abschnitt wird auf Grundlagen und Möglichkeiten der Verbindung von Kryptographie und Biometrie eingegangen, wobei das Hauptaugenmerk auf biometrischen Hash-Funktionen liegt.

2.1.4 Biometrisches Hashing

Einen Ansatz, biometrische Daten vor dem Missbrauch zu schützen, stellen biometrische Hash-Funktionen dar. Die Idee des biometrischen Hashing basiert auf Hash-Funktionen in der Kryptographie. Ein Hauptziel kryptographischer Hash-Funktionen besteht darin, identische Eingabedaten immer in gleiche Hash-Werte zu transferieren (siehe beispielsweise [DK07], [Eck05]). Verglichen mit kryptographischen Hash-Verfahren ergibt sich damit ein wesentliches Problem bei biometrischen Hash-Funktionen aus der Varianz der biometrischen Eingabedaten (Intra-Klassen-Varianz). Hier muss folglich aus, nicht zwingend identischen, biometrischen Informationen einer Person als Input immer ein identischer Hash-Wert bestimmt werden. In [SVD08b] haben wir die entsprechenden Eigenschaften für kryptographische Hash-Funktionen identifiziert und denen biometrischer Hash-Funktionen gegenüber gestellt. Tabelle 2.3 gibt die wichtigsten Anforderungen an die Generierung von Hashes sowohl in der Kryptographie als auch in der Biometrie vergleichend wieder.

Tabelle 2.3: Gegenüberstellung der Anforderungen an kryptographische (kh: A → B) und biometrische Hash-Funktionen (bh: A → B) [SVD08b]

Anforderung	**kryptographische Hash-Funktion**	**biometrische Hash-Funktion**
a) Reproduzierbarkeit	Sind zwei Eingabewerte a und a' identisch, so sollen auch die beiden durch ein und dieselbe kryptographische Hash-Funktion kh berechneten Werte $kh(a)$ und $kh(a')$ identisch sein.	Stammen zwei biometrische Eingabedaten a und a' von derselben Person P, so sollen auch die beiden durch ein und dieselbe biometrische Hash-Funktion bh berechneten Werte $bh(a)$ und $bh(a')$ identisch sein.
b) Kollisionsresistenz	Sind zwei Eingabewerte a und a' ungleich, dann müssen auch die durch eine kryptographische Hash-Funktion kh berechneten Hash-Werte $kh(a)$ und $kh(a')$ ungleich sein.	Stammen zwei biometrische Eingabedaten a und a' von zwei unterschiedlichen Personen P und P', dann müssen auch die durch eine biometrische Hash-Funktion bh berechneten Hash-Werte $bh(a)$ und $bh(a')$ ungleich sein.
c) Unumkehrbarkeit	Es sollte rechnerisch nicht möglich sein, aus dem durch die kryptographische Hash-Funktion kh erzeugten Hash-Wert $kh(a)$ wieder den Ausgangswert a zu bestimmen.	Es sollte rechnerisch nicht möglich sein, aus dem durch die biometrische Hash-Funktion bh erzeugten Hash-Wert $bh(a)$ wieder den Ausgangswert a zu bestimmen.
d) Bit-Sensibilität	Kleine Änderungen der Eingabedaten a sollten zu möglichst großen Veränderungen der Ausgabedaten $kh(a)$ führen.	Änderungen der Eingabedaten a sollen nur dann keinen Einfluss auf die Ausgabedaten $bh(a)$ haben, wenn diese von ein und derselben Person stammen.

Die in Tabelle 2.3 aufgeführten Anforderungen können für kryptographische als auch für biometrische Hash-Funktionen nicht in jedem Fall erfüllt werden. Da beispielsweise von kryptographischen

Hash-Algorithmen oft Eingabedaten einer nicht festgelegten Größe auf Hashes mit - zumeist wesentlich geringerer - fixer Länge abgebildet werden, besteht theoretisch die Möglichkeit von Kollisionen.

Die Berechnung von stabilen individuellen biometrischen Hashes unterscheidet sich deutlich von der kryptographischen Hash-Generierung in den Eigenschaften a) Reproduzierbarkeit, b) Kollisionsresistenz und d) Bit-Sensibilität (siehe Tabelle 2.3). Eine wesentliche Anforderung an kryptographische Hash-Funktionen ist, dass identische Hashes nur aus identischen Eingabewerten berechnet werden. Demgegenüber besteht die Herausforderung in der Biometrie in unscharfen Eingabedaten verursacht durch Intra-Klassen-Varianz und Inter-Klassen-Ähnlichkeit. Aufgrund der Intra-Klassen-Varianz ist es erforderlich, dass eine biometrische Hash-Funktion aus in einem gewissen Rahmen variierenden Daten einer einzelnen Person identische Hash-Werte generieren kann. Im Vergleich zum kryptographischen Hashing entspricht dies demzufolge einer Abweichung bei der Bit-Sensibilität und der Kollisionsresistenz bezüglich einer Person mit dem Ziel, die Hash-Reproduzierbarkeit für diese Person zu erreichen. Andererseits begründet die Inter-Klassen-Ähnlichkeit, dass eine biometrische Hash-Funktion auch individuelle Hashes für verschiedene Personen generieren muss. Dies setzt eine hohe Kollisionsresistenz bei einer hohen Anzahl von Personen voraus. Beide Zielsetzungen, hohe Reproduzierbarkeit und hohe Kollisionsresistenz, sind gegenläufig. Dies macht sich bemerkbar, wenn Maßnahmen zur Anpassung an die Intra-Klassen-Varianz, beispielsweise durch Verbreiterung eines Quantisierungsintervalls, neben der gewünschten Erhöhung der Reproduzierbarkeit auch die Kollisionsresistenz verringern.

Gelingt es mit einer biometrischen Hash-Funktion, biometrische Hashes mit sehr hoher Wahrscheinlichkeit zu reproduzieren bei gleichzeitig hoher Kollisionsresistenz, können drei Hauptanwendungsbereiche für dieses Verfahren identifiziert werden: Schutz von biometrischen Referenzdaten (Template Protection), Benutzerauthentifikation und Erzeugung von stabilen individuellen Hash-Werten für kryptographische Anwendungen. Da es bei einigen biometrischen Modalitäten möglich ist, aus den erhobenen Rohdaten persönliche Informationen des Besitzers abzuleiten, handelt es sich zumindest in diesen Fällen um personenbezogene oder -beziehbare Daten. Als Beispiele können an dieser Stelle das Gesicht und die eigenhändige Unterschrift einer Person genannt werden. Aus diesem Grund besteht für die entsprechenden Daten ein gewisser Schutzbedarf, um den Missbrauch durch Dritte zu verhindern. Erfüllt ein biometrischer Hash-Algorithmus die Anforderung der Unumkehrbarkeit (vergleiche Tabelle 2.3, Anforderung c), so bietet sich die Möglichkeit, aus den Referenzdaten einen biometrischen Hash zu generieren. Durch diesen Transfer in den Hash-Raum entstehen neue Referenzdaten, aus denen sich keine Rückschlüsse auf den Urheber der biometrischen Informationen ableiten lassen.

Ziel des Template Protection in der Biometrie ist es, biometrische Referenzdaten (engl. Templates) davor zu schützen, zu anderen als dem vorgesehenen regulären Verwendungszweck genutzt zu werden. Dazu werden die Daten in einer Form verändert, die keinen Rückschluss auf die ursprünglichen Informationen zulässt. Biometrische Verfahren können in diesem Bereich unter anderem wie folgt eingesetzt werden:

Biometrisches Schlüsselmanagement (engl. Biometric Key Management). Methoden die dem Schutz kryptographischer Schlüssel durch die Verwendung eines biometrischen Authentifikationssystems dienen, werden unter dem Begriff Biometric Key Management zusammengefasst. Ist die biometrische Authentifikation erfolgreich, wird dem Nutzer Zugriff auf die geschützten kryptographischen Schlüssel gewährt (siehe zum Beispiel [SRS+99]). In diesem Fall werden durch die Biometrie folglich andere als vertraulich zu behandelnde Informationen geschützt.

Verschlüsselung biometrischer Daten (engl. Encrypted Biometrics). Unter dem Begriff Encrypted Biometrics werden Systeme zur Sicherung sensibler biometrischer Informationen (beispielsweise Referenzdaten) zusammengefasst, durch deren Verschlüsselung mit kryptographischen Verfahren. Das bedeutet, dass biometrische Systeme und sowohl berechtigte als auch unberechtigte Personen nur Zugriff auf die Daten erlangen, wenn sie über die richtigen Schlüssel verfügen. Verfahren dieser Gruppe haben den Nachteil, dass die biometrischen Daten für ihre Nutzung entschlüsselt werden müssen und somit zumindest temporär als Klartexte im System vorliegen. Abhilfe könnten in der Zukunft so genannte homomorphe Kryptographie-Ansätze schaffen, mit denen es möglich sein soll, Operationen zwischen verschlüsselten Daten ohne deren Dekodierung durchzuführen. Momentan steht die Erforschung entsprechender Algorithmen noch am Anfang. Zusätzlich zu den genannten Kritikpunkten muss der sichere Umgang mit den kryptographischen Schlüsseln geregelt sein (Key Management).

Erneuerbare Biometriken (engl. Renewable Biometrics). Da sowohl statische als auch dynamische biometrische Charakteristiken auf biologischen Eigenschaften des menschlichen Körpers beruhen, unterliegen sie der Veränderung durch Alterung oder auch durch mechanischen Einfluss. Um die Performanz eines biometrischen Systems aufrecht zu erhalten, kann daher eine Erneuerung der Referenzdaten erforderlich werden. Eine Erneuerung kann auch notwendig werden, wenn die Referenzdaten einer Person kompromittiert wurden.

Cancelable Biometrics. Verfahren zur Erzeugung von Cancelable Biometrics ermöglichen es, aus den individuellen Merkmalen einer Modalität einer Person unterschiedliche Referenzdaten zu generieren. Diese können beispielsweise für bestimmte Zeiträume oder Anwendungen gültig sein. Um die weitere Verwendung biometrischer Referenzdaten in einer Datenbank zu verhindern, können sie innerhalb dieser Systeme unbrauchbar gemacht werden. Die Möglichkeit der Erzeugung unterschiedlicher Referenzdaten als auch die Eigenschaft der Rückrufbarkeit basiert in den meisten Fällen auf einem Set von Parametern (siehe auch [LJ09]), welche die tatsächliche Biometrie einer Person nach der Aufnahme der Daten verändern. Dabei ergibt sich jedoch ein ähnlicher Nachteil wie bei der Verschlüsselung biometrischer Daten, da auch hier für die sichere Speicherung und den vertraulichen Transfer der Parameter gesorgt werden muss. Die Möglichkeit der Erzeugung mehrerer unterschiedlicher Referenzdaten ohne zusätzliche Parametrisierung ist vor allem bei dynamischen biometrischen Modalitäten gegeben, indem das zugrunde liegende Verhalten verändert wird. Gerade bei der Hand-

schrift sind durch diverse Kombinationen (beispielsweise alphanumerisch mit und ohne Sonderzeichen oder auch Symbole) eine Vielzahl von geschriebenen Inhalten und damit unterschiedlicher Referenzdaten denkbar, wie in verschiedenen Publikationen gezeigt wird (zum Beispiel [KHH02], [Vie06]). Dies ermöglicht zum einen den Ersatz kompromittierter Authentifikationsobjekte als auch andererseits die Verwendung einer Modalität in mehreren Anwendungsszenarien ohne, dass die jeweiligen biometrischen Informationen preisgegeben werden.

Privacy Preserving. Protokolle für Privacy Preserving sind so gestaltet, dass sensitive Informationen ohne deren Preisgabe verarbeitet werden können. Einen Ansatz bietet beispielsweise additively homomorphic Encryption, bei der einzelne Operationen (Addition und Multiplikation) durchgeführt werden können, ohne die betreffenden Daten entschlüsseln zu müssen. Dieses Vorgehen kann in der Biometrie den datenschutzkonformen Umgang mit sensiblen Daten erleichtern. So präsentieren zum Beispiel Erkin et al. in [EFG+09] einen entsprechenden Ansatz für die Gesichtserkennung.

2.1.5 Optimierung

Optimierung von biometrischen Algorithmen im Kontext dieser Arbeit bedeutet die Verbesserung der Performanz in Bezug sowohl auf die Verifikation als auch auf die Hash-Generierung. Dafür haben wir unterschiedliche Verfahren ausgewählt, die in den folgenden Abschnitten beschrieben werden.

Optimierung basierend auf der Merkmalsebene

Zur Verbesserung von Authentifikations- beziehungsweise Hash-Generierungsperformanz kann eine Optimierung der Referenzalgorithmen basierend auf Merkmalsebene sowohl durch das Hinzufügen neuer, als auch das Entfernen vorhandener Merkmale erreicht werden. Zusätzliche Merkmale können einerseits in vielen Fällen zur Verbesserung der Authentifikationsperformanz beitragen. Auf der anderen Seite kann es durch unkritisches Hinzufügen einzelner Merkmale oder durch deren Kombinationen mit anderen Merkmalen zu Einbußen in der Performanz kommen. Anhaltspunkte, welche für einen speziellen Anwendungsbereich besser geeignet sind als andere, können die Ergebnisse einer Analyse einzelner, aber auch des Zusammenspiels mehrerer Merkmale geben. Auf den Ergebnissen der Analyse aufbauend können Merkmale entweder für die Verwendung ausgewählt oder ausgeschlossen werden.

Im Rahmen dieser Arbeit führen wir unterschiedliche statistische Analysen der den Referenzalgorithmen zugrunde liegenden Merkmale durch. Die dabei zum Einsatz kommenden untersuchten Verfahren können entsprechend der Terminologie von John et al. in Wrapper und Filter unterteilt werden [JKP94].

Bei wrapper-basierten Verfahren werden optimale Merkmals-Sets in Abhängigkeit vom verwendeten biometrischen Algorithmus bestimmt (siehe Abbildung 2.13a). Bezogen auf unsere Zielstellung muss dabei für jede potentiell mögliche Zusammenstellung von Merkmalen der

jeweilige Hash-Algorithmus im Verifikations- beziehungsweise Hash-Modus ausgeführt werden. Darauf aufbauend werden diejenige Merkmalsmengen selektiert, die jeweils die beste Performanz aufweisen. Unabhängig vom jeweiligen biometrischen Algorithmus arbeiten andererseits Filter bei der Merkmalsselektion, indem die Relevanz der Merkmale unabhängig vom adressierten Algorithmus beziehungsweise Szenario durch statistische Verfahren analysiert wird (siehe Abbildung 2.13b).

Es ist theoretisch durchaus möglich, für einen biometrischen Algorithmus das optimale Merkmalset zu finden. Dabei ist aber zu bedenken, dass bei einer vollständigen Suche auf einer Menge von k Merkmalen 2^k mögliche Teilmengen existieren, die bei einer Analyse in Betracht gezogen werden müssen. Diese große Anzahl zu betrachtender Möglichkeiten macht diesen Ansatz, in Zusammenhang mit der jeweils zusätzlich durchzuführenden Klassifikation, sehr zeit- und rechenintensiv. Annehmbare Alternativen bieten hier Strategien, die nicht alle, sondern nur vielversprechende Kombinationen von Merkmalen analysieren.

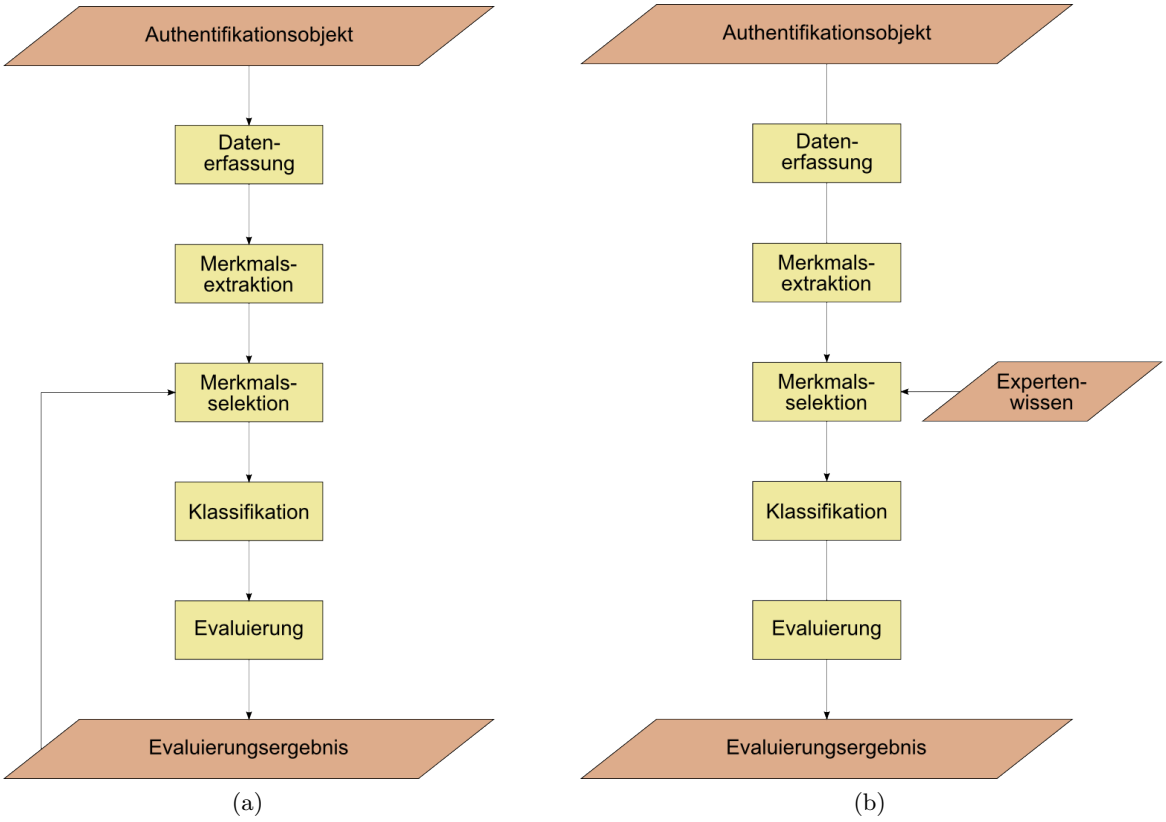


Abbildung 2.13: Schematische Darstellung der Funktionsweise der Merkmalsselektion mithilfe von Wrappers (a) und Filtern (b)

In dieser Arbeit haben wir uns aus einer Vielzahl von existierenden Methoden auf drei Wrapper- und fünf Filter-Verfahren (vergleiche 5.3.2 *Verwendete Wrapper-Methoden* beziehungsweise 5.3.3 *Verwendete Filter-Methoden*) konzentriert, die uns für die Analyse von Merkmalen der biometrischen Modalität der Handschrift als sinnvoll erschienen.

Optimierung durch biometrische Fusion

Ein weit verbreiteter Ansatz, die Sicherheit von Systemen zur Benutzerauthentifikation zu erhöhen, ist die Kombination mehrerer Einzelsysteme. Die Nutzung eines Automaten zum Abheben von Geld vom eigenen Konto stellt eine typische Anwendung eines solchen fusionierten Systems dar. Für den Vorgang ist zum einen der Besitz einer speziellen von der Bank ausgestellten Karte notwendig, zum anderen das Wissen der dazu gehörigen persönlichen Identifikationsnummer (PIN). Nur wenn eine Person beides präsentieren kann, kann sie über das Guthaben des Kontos verfügen. Allerdings besteht auch hier das generelle Problem, dass die Authentifikation basierend auf den beiden Authentifikationsobjekten (Karte und PIN) durchgeführt wird und nicht die Anwesenheit der autorisierten Person sichergestellt wird. Zusätzlich ist es natürlich auch möglich, das geheime Wissen und/oder den persönlichen Besitz mit dem Faktor Biometrie zu kombinieren. Abbildung 2.14 zeigt die Möglichkeiten der Fusion der drei Authentifikationsmethoden geheimes Wissen, persönlicher Besitz und Biometrie. Dabei verdeutlichen die Schnittflächen der Kreise die möglichen Kombinationen der Faktoren. Die Erhöhung der Sicherheit bei der Authentifikation geht dabei durch höhere Anzahl an Faktoren zu Lasten der Komplexität bei der Bedienung. Diese bedingt unter anderem auch die Akzeptanz der Nutzer. Ein System, dessen Sicherheit auf der Präsentation einer Vielzahl von unterschiedlichen Authentifikationsobjekten beziehungsweise -faktoren beruht, wird vermutlich von den Benutzern abgelehnt werden, auch wenn es eine erhöhte Sicherheit verspricht.

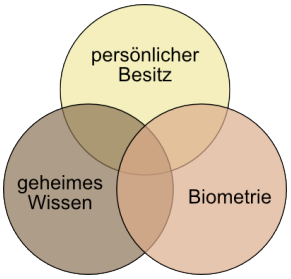


Abbildung 2.14: Fusionsmöglichkeiten der Faktoren geheimes Wissen, persönlicher Besitz und Biometrie

Der Fokus der in dieser Arbeit untersuchten Ansätze zur Kombination biometrischer Komponenten liegt auf der single-modalen Fusion. Alternative Konstellationen der Fusion der Handschrift mit anderen biometrischen Modalitäten sind natürlich ebenfalls möglich (siehe beispielsweise [VS05], [WSV06]), werden hier aber nicht berücksichtigt. Das Konzept der single-modalen Fusion adressiert die Kombination von Komponenten einer einzelnen Modalität, in unserem Fall ist dies die dynamische biometrische Handschrift. Dies hat für den Benutzer den Vorteil, dass er sich nur mit der Handhabung einer einzelnen Modalität und entsprechender Sensoren auseinander setzen muss. Als Fusionskomponenten können hier beispielsweise verschiedene Instanzen, Semantiken oder Algorithmen zum Einsatz kommen (siehe auch [SVF11]).

Bei unseren Forschungen im Bereich multi-biometrische Systeme orientieren wir uns an der Klassifikation, die von Ross, Nandakumar und Jain in [RNJ06] vorgeschlagen wurde. Diese basiert

auf der Art und der Anzahl der verwendeten Fusionskomponenten und bildet die Grundlage unserer Untersuchungen bezüglich der Fusion in dieser Arbeit.

Multi-sensor Fusion. Bei dieser Fusionsklasse werden unterschiedliche Sensoren verwendet, die die Daten einer einzelnen biometrische Modalität erfassen.

Multi-algorithmic Fusion. Multi-biometrische Systeme dieser Klasse verwenden unterschiedliche Algorithmen einer Modalität zur Authentifikation.

Multi-instance Fusion. Verschiedene Instanzen einer einzelnen Modalität werden bei multi-biometrischen Systemen dieser Klasse verwendet. Neben der durch ihre eingeschränkte Anzahl begrenzten Verwendung von multiplen physiologischen Eigenschaften wie Fingern oder Iriden können vor allem von verhaltensbasierten Modalitäten unterschiedliche Instanzen generiert werden.

Multi-sample Fusion. In dieser Klasse multi-biometrischer Systeme werden verschiedene Samples einer einzelnen Modalität für die Authentifikation genutzt. Dies können beispielsweise Aufnahmen unterschiedlicher Positionen des selben Fingers oder mehrere Aufnahmen einer Schriftprobe einer Person sein.

Multi-modal Fusion. Bei der multi-modalen Fusion werden mindestens zwei verschiedene biometrische Modalitäten miteinander kombiniert, um die Authentifikationsperformanz zu erhöhen.

Die fünf genannten Klassen können auch untereinander kombiniert werden. So ist beispielsweise ein multi-instance System denkbar, welches zwei Schreibinhalte mittels zweier unterschiedlicher Sensoren aufnimmt.

Gemäß der einzelnen Prozessmodule eines biometrischen Systems kann die Fusion zweier oder mehr Komponenten an verschiedenen Stellen durchgeführt werden [RNJ06]. Im Folgenden werden die entsprechenden Level der Fusion genannt und kurz beschrieben.

Sensor Level Fusion. Die Fusion auf Sensor Level nutzt die von den einzelnen Sensoren erfassten Daten, kombiniert sie auf geeignete Weise und gibt das Ergebnis als fusionierte Sensordaten an den Merkmalsextraktor weiter. An dieser Stelle kann zum Beispiel bei einem Gesichtserkennungssystem die Frontalaufnahme des Gesichtes mit einer rechts- und einer linksseitigen Aufnahme zu einem einzigen Bild zusammengesetzt werden. Dieses kann dann im weiteren Erkennungsprozess verwendet werden und durch die zusätzlichen Informationen zu besseren Ergebnissen führen.

Feature Extraction Level Fusion. Bei der Fusion auf Feature Extraction Level werden die Merkmalsvektoren der beteiligten Systeme nach deren Extraktion miteinander fusioniert. Dadurch entsteht ein fusionierter Merkmalsvektor, der für den Vergleich mit den Referenzdaten herangezogen wird. Diese Art der Fusion kann beispielsweise durch einfaches Aneinanderhängen der einzelnen Vektoren durchgeführt werden. Bei dieser Methode kann

allerdings, je nach Modalität und Anzahl der jeweils extrahierten Merkmale, ein Merkmalsvektor mit sehr hoher Dimension entstehen, der für weitere Operationen unhandlich werden kann. Im Fall der aktuellen Version des BioHash-Algorithmus werden zum Beispiel 131 Merkmale extrahiert. Bei einer Fusion auf Feature Extraction Level würde sich folglich die Dimension des fusionierten Merkmalsvektors mit jedem beteiligten Algorithmus auf Basis des BioHash-Algorithmus um diese Zahl erhöhen.

Matching Score Level Fusion. Die Fusion auf Matching Score Level kombiniert die individuellen Ähnlichkeits- beziehungsweise Unähnlichkeitswerte, die die einzelnen unabhängigen Vergleicher der beteiligten Systeme ermittelt haben. Der daraus berechnete Wert wird dann dem Entscheidungsprozess übergeben. Der entscheidende Vorteil dieses Fusionsansatzes liegt darin, dass die Anzahl der zu kombinierenden Werte der Anzahl der beteiligten Subsysteme entspricht. Dies vereinfacht die durchzuführenden Operationen und ermöglicht eine einfache, aber effektive Gewichtung der einzelnen Werte vor der Fusion. Ein großer Teil der aktuellen multi-biometrischen Systeme nutzt Fusion auf Matching Score Level.

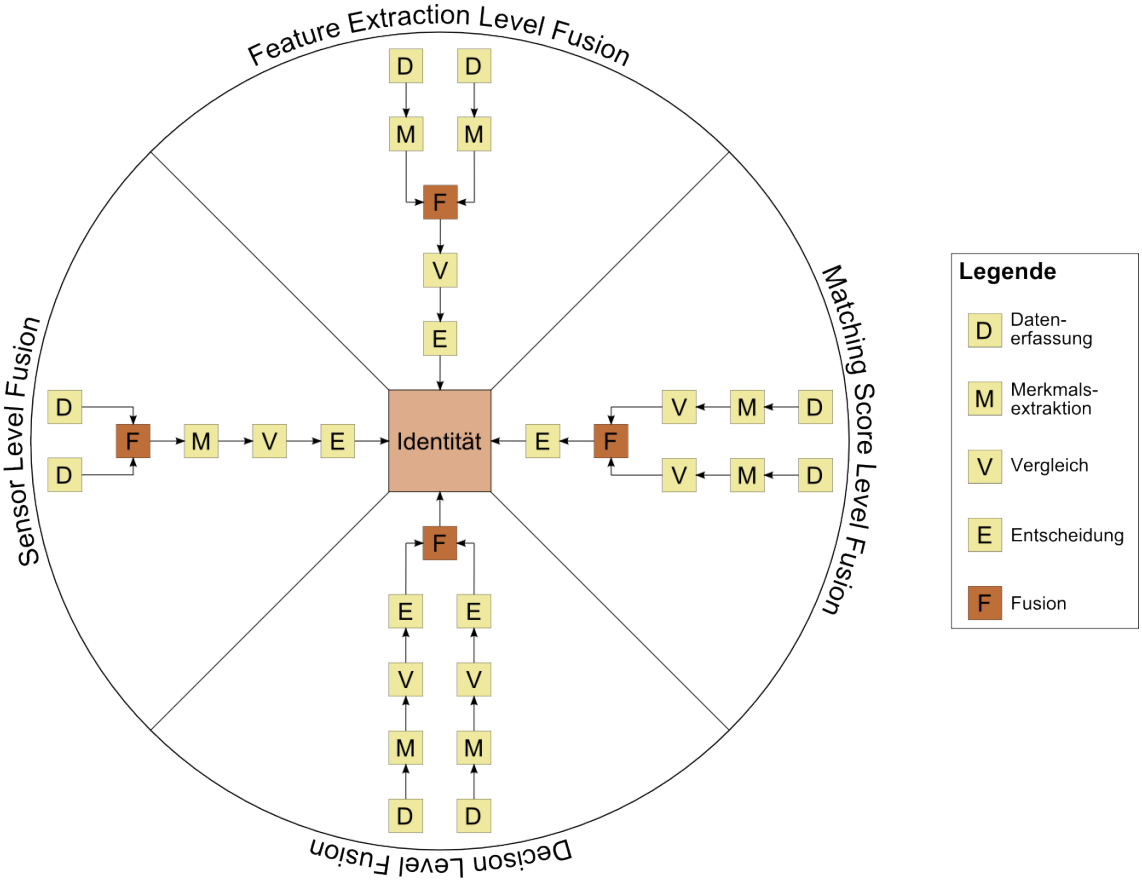


Abbildung 2.15: Klassifikation nach dem Zeitpunkt der Fusion innerhalb des biometrischen Authentifikationsprozesses am Beispiel von zwei Fusionskomponenten (nach [RNJ06])

Decison Level Fusion. Die Ergebnisse der vollständig durchlaufenen Subsysteme werden bei einer Fusion auf Decison Level miteinander kombiniert. Dabei handelt es sich um

die einzelnen Entscheidungen, die beispielsweise bei einer Verifikation durch boolsche Operationen fusioniert werden können. Bei einer Identifkation kann die finale Entscheidung zum Beispiel durch ein (gewichtetes) Ranking der in Frage kommenden Personen erfolgen.

In Abbildung 2.15 sind die vier Level der biometrischen Fusion grafisch dargestellt, die entsprechend des Zeitpunktes der Fusion innerhalb des biometrischen Authentifikationsprozesses gebildet werden können. Beispielhaft ist jeweils die Kombination zweier Fusionskomponenten abgebildet. Für die biometrische Fusion können nach Belieben weitere Komponenten hinzugefügt werden. Dabei sind auch Kombinationen sowohl mittels unterschiedlicher Fusionsklassen als auch -level denkbar.

2.1.6 Biometrische Fehlerraten

Der Vergleich der Authentifikationsperformanz von biometrischen Systemen beziehungsweise Algorithmen wird mittels Fehlerraten durchgeführt. Diese müssen empirisch ermittelt werden, da sie nicht direkt im System gemessen werden können. Um eine geeignete Einstellung des Systems zu erreichen, ist ein hoher Testaufwand notwendig. Aufgrund der Intra-Klassen-Varianz kommt es zu Abweichungen bei biometrischen Daten einer Modalität einer einzelne Person, während durch die Inter-Klassen-Ähnlichkeit Daten unterschiedlicher Personen eine hohe Ähnlichkeit zueinander haben können. Um trotz Intra-Klassen-Varianz eine Authentifikation durchführen zu können, wird mit Schwellwerten gearbeitet. Dadurch kann ein Toleranzbereich definiert werden, in dem die Ähnlichkeit von Referenz- und aktuell präsentierten Daten als ausreichend bewertet. Außerhalb dieses Bereiches erfolgt eine Rückweisung. Je nach Einstellung des Schwellwertes ist folglich mit einem bestimmten Restfehler zu rechnen, bei dem beispielsweise eine autorisierte Person vom System gar nicht oder als eine andere autorisierte Person erkannt wird. Im Folgenden wird basierend auf Mansfield und Wayman [MW02] auf die Ermittlung und Bewertung der in dieser Arbeit verwendeten Fehlerraten False Match Rate, False Non Match Rate und die darauf basierende Equal Error Rate zur Bestimmung der Verifikationsperformanz eingegangen.

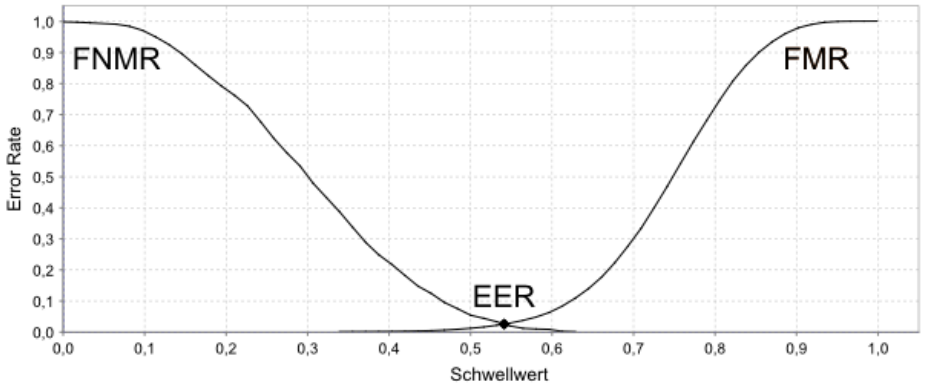


Abbildung 2.16: Beispielhafte Darstellung von FNMR, FMR und EER

False Non Match Rate (FNMR). Die False Non Match Rate gibt die Wahrscheinlichkeit an,

mit der berechtigte Personen von einem biometrischen Algorithmus abgewiesen werden. Berechnet werden kann die False Non Match Rate als das Verhältnis der Anzahl der fälschlichen Rückweisungen und der Summe der berechtigten Versuche.

$$FNMR = \frac{Anzahl\ fälschlicher\ Rückweisungen}{Gesamtanzahl\ berechtigter\ Zutrittsversuche} \tag{2.1.1}$$

In der Abbildung 2.16 ist die Kurve der FNMR in Abhängigkeit zum Schwellwert für einen beispielhaften biometrischen Algorithmus dargestellt. Wird der Schwellwert zum Beispiel auf 0,3 gesetzt, ergibt sich eine False Non Match Rate von ca. 0,5. Daraus lässt sich schließen, dass bei dieser Einstellung berechtigte Personen bei durchschnittlich jedem zweiten Versuch abgewiesen werden. Wird der Schwellwert verringert, erhöht sich damit die Wahrscheinlichkeit, dass eine berechtigte Person vom System nicht akzeptiert wird.

False Match Rate (FMR). Wie häufig unberechtigte Personen von einem biometrischen Algorithmus akzeptiert werden, wird durch die False Match Rate angegeben. Sie wird durch die Division der Anzahl der fälschlichen Akzeptanzen durch die Anzahl der nicht erechtigten Versuche berechnet:

$$FMR = \frac{Anzahl\ fälschlicher\ Akzeptanzen}{Gesamtanzahl\ unberechtigter\ Zutrittsversuche} \tag{2.1.2}$$

Zu einer fälschlichen Annahme einer Person kann es einerseits kommen, wenn die Inter-Klassen-Ähnlichkeit zweier Nutzer hoch genug ist, dass einer zufällig als der andere akzeptiert wird. Auf der anderen Seite kann ein Dritter beabsichtigt durch einen (gezielten) Angriff versuchen, eine ausreichende Inter-Klassen-Ähnlichkeit zu einer autorisierten Person herzustellen.

Der Verlauf der False Match Rate wird in Abbildung 2.16 für einen beispielhaften Algorithmus gezeigt. Bei dieser Konstellation wird im statistischen Durchschnitt etwa jeder zweite unberechtigte Zutrittsversuch erfolgreich sein, die entsprechenden FMR von 0,5 liegt in diesem Fall bei einem Schwellwert von circa 0,74. Wird dieser weiter erhöht, vergrößert sich auch die Wahrscheinlichkeit der Annahme unautorisierter Personen.

Equal Error Rate (EER). Bei Gleichheit von False Non Match Rate und False Match Rate wird der entsprechende Wert als Equal Error Rate (siehe Schnittpunkt von FNMR und FMR in Abbildung 2.16) bezeichnet. Die Equal Error Rate wird häufig als Vergleichswert für biometrische Systeme beziehungsweise Algorithmen herangezogen. In unseren Evaluierungen verwenden wir die EER sowohl zum allgemeinen Vergleich der Ergebnisse der beiden Algorithmen als auch zum Vergleich der Wirksamkeit der Anpassung der Parameter der einzelnen Algorithmen beziehungsweise der Merkmalsselektion. Weiterhin ziehen wir die EER

auch zur Bestimmung von Fusionsparametern heran, die mittels der Verifikationsperformanz der einzelnen Algorithmen für verschiedene Schreibinhalte ermittelt werden.

Die Equal Error Rate ist nicht als der optimale Arbeitspunkt eines biometrischen Systems anzusehen. Der einzustellende Schwellwert hängt vielmehr von der Verwendung des einzelnen biometrischen Systems und der damit verfolgten Ziele ab. Ist es im Interesse des Betreibers, dass die Nutzer bequem, also möglichst beim ersten Versuch, Zugang erlangen, wird er den Schwellwert höher ansetzen. Damit sinkt die Wahrscheinlichkeit der Abweisung autorisierter Nutzer. Da False Non Match Rate und False Match Rate gegenläufig sind hat dies den gleichzeitigen Anstieg der Rate der fälschlicherweise akzeptierten Personen zur Folge. Entsprechend wirkt sich eine Verkleinerung des Schwellwertes gegenteilig aus, nämlich durch die Abnahme angenommener nicht berechtigter und der Zunahme abgewiesener autorisierter Personen. Diese Verschiebung des Schwellwertes würde das System also sicherer machen, auf Kosten der Bequemlichkeit der zum Zugang berechtigten Nutzer. Die Equal Error Rate wird als normierter Wert eingesetzt, der es ermöglicht, biometrische Systeme und Algorithmen beziehungsweise unterschiedliche Instanzen dieser (beispielsweise unter Verwendung unterschiedlicher Parametrisierungen) miteinander zu vergleichen.

2.1.7 Reproduktion und Kollision im Hash-Generierungsmodus

Die entscheidenden Informationen für die Güte eines Hash-Algorithmus stellen die Wahrscheinlichkeiten der genauen Reproduzierbarkeit und der dabei zu erwartenden Kollisionen dar. Gebräuchliche Fehlermaße aus der Biometrie (siehe 2.1.6 *Biometrische Fehlerraten*) können für die Untersuchung der Performanz der Hash-Generierung nicht eingesetzt werden. Wie im vorhergehenden Abschnitt erläutert, können sie dann im Zusammenhang mit biometrischen Hashes genutzt werden, wenn diese als Referenz- beziehungsweise Authentifikationsdaten eingesetzt werden, um die Trennschärfe bei deren Einsatz zur Authentifikation zu bestimmen (siehe zum Beispiel [SVD07a]). Bei der Untersuchung biometrischer Systeme anhand biometrischer Fehlerraten sind keine exakten Übereinstimmungen zwischen hinterlegten Referenz- und aktuell präsentierten Authentifikationsdaten erforderlich. Es wird lediglich überprüft, ob die entsprechenden Hashes eine Ähnlichkeit in einem ausreichenden, vorher definierten Ausmaß zueinander besitzen. Sollen jedoch individuelle identische Hashes aus variierenden biometrischen Daten einer Person für kryptographische Zwecke generiert werden, ist keine Abweichung von den Referenzdaten erwünscht. Um die Performanz der biometrischen Hash-Generierung zu messen, haben wir drei Maße entwickelt [SVD08a] [MSV11a]: die Reproduction Rate (RR), die Collision Rate (CR) und die Collision Reproduction Rate (CRR). Die Reproduction und Collision Rate werden auf Basis einer Analyse der Hashes mittels Hamming Abstand (siehe [Ham50]) bestimmt. Mit diesem Abstandsmaß ist es möglich, die genaue Übereinstimmung beziehungsweise den Grad der Ähnlichkeit von zwei Hash-Vektoren festzustellen. Dazu wird in jedem Vergleich der Hamming Abstand hd zwischen zwei biometrischen Hash-Vektoren b und b' berechnet. Wie in den Gleichungen 2.1.3 und 2.1.4 dargestellt, wird der Hamming Abstand hd zweier Hash-Vektoren b und b' mit derselben Dimensionalität k durch den Vergleich der Elemente mit identischem Index i bestimmt.

Das Ergebnis beträgt entweder 0, wenn beide Werte identisch sind, oder 1 im anderen Fall. Der Abstand zwischen b und b' ergibt sich entsprechend Gleichung 2.1.3 aus der Summe der Ergebnisse der einzelnen Vergleiche. Demzufolge wird sich der Wert für die Hamming Distanz im Intervall von $[0, k]$ befinden. Das bedeutet, dass es bei unseren Betrachtungen keine Rolle spielt, wie groß die Differenz zwischen den Werten der beiden Elemente b_i und b'_i ist.

$$hd(b, b') = \sum_{i=1}^{k} d(b_i, b'_i) \tag{2.1.3}$$

mit:

$$d(b_i, b'_i) = \begin{cases} 0, wenn\ b_i = b'_i \\ 1, sonst \end{cases} \tag{2.1.4}$$

Reproduction Rate. Um zu ermitteln, ob die hinterlegten Referenz-Hashes von den jeweiligen Personen identisch reproduziert werden, werden diese mit den Hashes der Testdaten der selben Personen verglichen. Dazu wird die Hamming Distanz verwendet, die die Anzahl der Elemente angibt, in denen sich beide Hashes unterscheiden. Die Reproduction Rate (RR) wird ermittelt durch das Verhältnis der identisch reproduzierten Hashes (Hamming Distanz gleich Null) zur Anzahl aller Tests.

$$RR = \frac{Anzahl\ identisch\ reproduzierter\ Hashes}{Anzahl\ generierter\ Hashes} \tag{2.1.5}$$

Ziel bei der Optimierung einer biometrischen Hash-Funktion ist es, die Reproduction Rate zu maximieren, wobei in der Biometrie durch die Intra-Klassen-Varianz eine Reproduction Rate mit 100% sehr schwer zu erreichen ist.

Collision Rate. Die Collision Rate (CR) stellt den Anteil an Kollisionen in Bezug auf alle Tests dar. Dazu wird ermittelt, wie oft die Referenz-Hashes registrierter Nutzer fälschlicherweise identisch durch andere Personen reproduziert werden (Hamming Distanz gleich Null). Bei diesen kann es sich entweder um registrierte Nutzer (zufällige Übereinstimmung) oder Angreifer (beabsichtigte Übereinstimmung) handeln. Die Collision Rate wird berechnet, indem die Anzahl der fälschlich identisch erzeugten Hashes durch die gesamte Anzahl der Versuche dividiert wird:

$$CR = \frac{Anzahl\ fälschlicherweise\ identisch\ reproduzierter\ Hashes}{Anzahl\ generierter\ Hashes} \tag{2.1.6}$$

Die Collision Rate soll durch die biometrische Hash-Funktion minimiert werden, um die

Anzahl von Kollisionen möglichst gering zu halten, die durch die Inter-Klassen-Ähnlichkeit verursacht werden. Diese entsteht dabei entweder durch zufällige Ähnlichkeit der Daten registrierter Nutzer zu denen der betroffenen Person oder mittels (gezielter) Fälschungen durch einen Angreifer.

Abbildung 2.17 zeigt beispielhaft ein Histogramm der Hamming Distanz basierend auf 14 Hash-Elementen. Zur Ermittlung der Histogrammwerte wird jeweils das Auftreten der einzelnen Distanzwerte von 0-14 aufsummiert, wobei zwischen Intra- und Inter-Klassentests unterschieden wird. Die Werte für RR und CR werden bei einer Hamming Distanz von Null ermittelt, was jeweils einer kompletten Übereinstimmung der Hashes entspricht. Das heißt, hier wird angezeigt, inwieweit individuelle Hashes durch den echten Nutzer reproduziert werden können (siehe Abbildung 2.17: RR, Intra-Klassentest) beziehungsweise Hashes fremder Nutzer identisch generiert werden können (siehe Abbildung 2.17: CR, Inter-Klassentest).

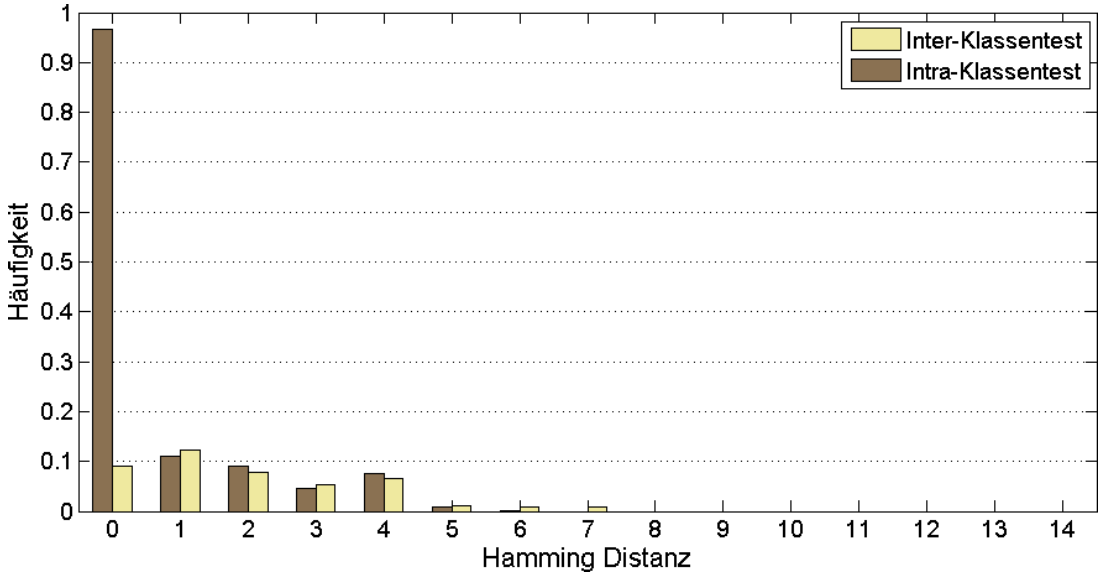


Abbildung 2.17: Beispielhaftes Hamming Distanz Histogramm für ein Set von 14 Merkmalen, RR und CR werden bei einer Distanz von Null ermittelt ($RR \approx 0,95$; $CR \approx 0,08$)

Die Beobachtung aller Distanzwerte kann hilfreich bei der Optimierung der Parameter der betrachteten Algorithmen (siehe Kapitel 4 *Verwendete Algorithmen*) sein, indem deren Veränderungen protokolliert und ausgewertet werden. Dieses Vorgehen wird, wie in Abschnitt 5.2 *Optimierung von Parametern (A.2)* beschrieben, im Rahmen der experimentellen Evaluierung durchgeführt, um einen Teil der Parameter-Sets zu ermitteln ($minCRR$, $RR \geq a\%$, $CR \leq b\%$, siehe Abschnitt 7.2 *Evaluierung: A.2 Optimierung von Parametern*).

Collision Reproduction Rate. Da sich aufgrund der Quantisierung in beiden hier untersuchten Algorithmen Reproduction Rate und Collision Rate gegenseitig beeinflussen, bewirkt das gezielte Verbessern des einen Wertes die Verschlechterung des anderen und umgekehrt. Um einen Trade Off-Wert zum Vergleich der Algorithmen untereinander als auch verschiedener

Optimierungsschritte (zum Beispiel Parametrisierung, Fusion) zu erhalten, haben wir die Collision Reproduction Rate entwickelt (CRR, siehe Gleichung 2.1.7, [MSV11a]).

$$CRR = \frac{CR + (1 - RR)}{2} \tag{2.1.7}$$

Das Ziel ist, einen Indikator für die Performanz der biometrischen Hash-Generierung zu entwickeln, vergleichbar mit der Equal Error Rate bei der Verifikation. In der Formel 2.1.7 sind RR und CR gleich gewichtet, eine Wichtung kann aber bei Bedarf individuell angepasst werden.

Die drei von uns entwickelten Performanzmaße für biometrische Hash-Funktionen (RR, CR und CRR) sind nicht auf die Handschrift beschränkt. Sie können für jedes biometrische Hash-Verfahren eingesetzt werden, da sie auf dem direkten Vergleich der erzeugten Hash-Werte mittels Hamming Distanz beruhen.

2.2 Stand der Technik

In diesem Abschnitt gehen wir auf eine Auswahl relevanter Ergebnisse und Erkenntnisse als Übersicht zum State of the Art ein. Dabei konzentrieren wir uns auf Veröffentlichungen im Kontext unserer Arbeit.

2.2.1 Biometrische Handschrift

Seit einigen Jahrzehnten gibt es in der biometrischen Forschung Untersuchungen zur Verwendung der Handschrift zur Benutzerauthentifikation. Die Beweggründe liegen nahe, da die Handschrift, speziell die Unterschrift, schon seit langer Zeit zur Authentifikation und Willensbekundung eingesetzt wird. Dies setzt ein gewisses Maß an Individualität einzelner Handschriften und an Unterscheidbarkeit zwischen Handschriften unterschiedlicher Personen voraus. In [Mic82] weist Michel darauf hin, dass der Grundstein für die Individualität der Handschrift bereits im Kindesalter gelegt wird. Basierend auf den vorgegebenen Formen und Größenverhältnissen der einzelnen Buchstaben und Zahlen entwickeln Kinder schon beim Erlernen der Schrift persönliche Ausprägungen. Beeinflusst durch den Prozess des Älterwerdens und des Wachstums unterliegt die Handschrift von Heranwachsenden mehr oder weniger umfangreichen Schwankungen, die zusätzlich noch von persönlichen Neigungen und vom Schönheitsempfinden des Einzelnen beeinflusst werden. Auch wenn die Intra-Klassen-Varianz mit Eintritt ins Erwachsenenalter deutlich geringer wird, unterliegt die Handschrift weiterhin Schwankungen. Gründe sind dafür hauptsächlich darin zu suchen, dass am Schreibprozess eine Vielzahl von mentalen und motorischen Prozessen beteiligt sind. Entsprechend ist im Allgemeinen mit fortschreitendem Alter eine Zunahme der Intra-Klassen-Varianz zu beobachten, in Fällen, bei denen es zu geistigen und körperlichen Abbauerscheinungen kommt.

Abhängig von der Art der Erfassung der Handschriftenproben kann zwischen statischen und dynamischen Verfahren unterschieden werden. Erfolgt die Erfassung etwa durch Aufnahme des Ergebnisses des Schreibprozesses (beispielsweise durch Fotografie oder Flachbett-Scanner) handelt es sich bei den aufgenommenen Informationen um statische Handschriftendaten. Die Auswertung kann bildbasierend auf den so entstanden Daten erfolgen. Im Gegensatz dazu wird bei dynamischen Erfassungsverfahren der Schreibvorgang an sich zeitabhängig erfasst. Dies ermöglicht zusätzlich zur Auswertung mittels statischer Merkmale auch die Untersuchung der Daten abhängig von ihrem zeitlichen Verlauf.

Ziel der Erfassung und Untersuchung kann neben der biometrischen Authentifikation auch die Erkennung des geschriebenen Inhaltes sein. In [PM13] präsentieren die Autoren beispielsweise einen Ansatz, basierend auf statischen Handschriftinformationen arabische Schriftzeichen zu erkennen. Im biometrischen Anwendungsbereich untersuchen zum Beispiel Johnson und Guest [JG11] die Auswirkungen von Begrenzungen in der Ausdehnung des verfügbaren Schreibbereiches anhand von Unterschriftfeldern auf Formularen auf die Erkennungsgenauigkeit. Die zugrunde liegenden Daten sind demzufolge statischer Natur.

Plamondon und Lorette präsentieren bereits in [PL89] einen Überblick über unterschiedliche Verfahren zur handschriftbasierten Authentifikation. Die Autoren unterscheiden statische und dynamische Ansätze, wobei letztere nochmal in funktions- beziehungsweise merkmalbasiert unterteilt werden. Ein häufig genutzter Ansatz in der dynamischen handschriftbasierten Biometrie ist die Verwendung von statistischen Merkmalen, die aus der Sequenz von Erfassungspunkten eines Schrift-Samples abgeleitet werden. Jain et al. stellen in [JGC02] ein System vor, welches jeweils die einzelnen Segmente eines Schriftzuges zu einem Segment verbindet. Aus dessen X- und Y-Koordinaten werden dann sowohl statische als auch dynamische Merkmale abgeleitet. Als Segment wird in der Handschriftbiometrie ein (Teil-)Schriftzug vom Aufsetzen des Stiftes bis zu dessen darauf folgenden Absetzen bezeichnet. Vielhauer beschreibt in [Vie06] 69 statistische Merkmale, extrahiert basierend auf den Sensorsignalen X-, Y-Position, Druck, Höhen- und Seitenwinkel. Häufig aus den Signalen abgeleitete statistische Merkmal sind beispielsweise Schreibdauer, Geschwindigkeit oder Beschleunigung (vergleiche auch [PL89], [Vie06] oder [FEL14]). Ein handschriftbasiertes System, das nicht auf statistischen Informationen sondern auf einem Hidden Markov Model zusammen mit einem Viterbi Path-Ansatz beruht präsentieren Ly-Van et al. in [LGD04].

Ein Vorteil der Handschrift besteht darin, dass hier der geschriebene Inhalt bei Bedarf (zum Beispiel bei Kompromittierung, für zusätzliche Applikationen) geändert werden kann. Auf diesen Aspekt gehen wir im folgenden Abschnitt ein.

Semantiken als alternative Schreibinhalte

Dadurch, dass dynamische biometrische Modalitäten auf bewusstem menschlichen Verhalten basieren, wird die Einbringung von zusätzlichen Informationen beziehungsweise Eigenschaften ermöglicht. Als Beispiel sei hier die Änderung der Sprechweise durch Erhöhung der Geschwindigkeit

genannt. Vorteile ergeben sich dabei bei einigen dynamischen Modalitäten durch die bewusste Veränderung der semantischen Bedeutung des Verhaltens durch die jeweilige Person. Dies ist durch alternative Inhalte möglich, zum Beispiel textuell beim Tippverhalten, beim Sprechen oder eben auch bei der Handschrift. In unserer Arbeit verwenden wir den Begriff Semantik für alternative Schreibinhalte. Die dynamische Modalität der Handschrift wird im Bereich der biometrischen Benutzererkennung oft mit der Authentifikation durch die Unterschrift gleich gesetzt. Allerdings wurden in den letzten Jahren auch Untersuchungen zur Verwendung von alternativen Schreibinhalten durchgeführt. Beispielsweise verwendet Schmidt in [Sch99] zusätzlich zur Unterschrift das Word *Grünschnabel*. Kato et al. untersuchen in [KHH02] die Möglichkeit der Verwendung von mehr oder weniger komplexen Zeichnungen für die Authentifikation. Vielhauer et al. zeigen in verschiedenen Publikationen (zum Beispiel [VSM02], [Vie06], [SVD09b]) dass die Verwendung von zur Unterschrift alternativen Schreibinhalten zur Benutzerverifikation ähnliche Ergebnisse liefert wie die Signatur selbst. Zu den Vorteilen des Einsatzes von Alternativen zur Unterschrift zählen sowohl die Anonymisierung, das Hinzufügen des Faktors *geheimes Wissen*, als auch das Ersetzen beziehungsweise Widerrufen des Inhaltes. Durch letzteren Aspekt kann durch das Nutzen eines alternativen Schriebinhaltes eine neue Referenz gespeichert werden, während die bisherige gelöscht werden kann.

2.2.2 Biometrisches Hashing

In der Literatur findet sich eine Vielzahl von Ansätzen, die die Biometrie mit Aspekten der Kryptographie verbinden. Diese reichen vom Schutz von kryptographischen Schlüsseln durch Biometrie bis zur Generierung von Hash-Werten basierend auf einzelnen biometrischen Modalitäten. In den folgenden Absätzen wird auf eine Auswahl der publizierten Methoden eingegangen. Die Selektion stellt dabei keine Wertung der vorhandenen Verfahren dar. In diesem wie auch in den folgenden Abschnitten zum Stand der Technik beginnen wir mit der Nennung eigenen Vorarbeiten im jeweiligen Bereich, sofern vorhanden.

In [Alb94] präsentiert Albert eine allgemeine Herangehensweise zur Ableitung kryptographischer Schlüssel direkt aus einem biometrischen Merkmal. Der Autor nennt dabei als zwei zu bewältigende Hauptprobleme die Intra-Klassen-Varianz bei der Erfassung der biometrischen Daten und dass bei einer Kompromittierung des Merkmals oder des Schlüssels beide nicht mehr durch das entsprechende System genutzt werden können.

Soutar et al. beschreiben in [SRS+99] ein Verfahren, dessen Ziel der Schutz von kryptographischen Schlüssel durch biometrische Authentifikation. Bei einer erfolgreichen Authentifikation, erhält der Nutzer Zugang zum Schlüssel. Systeme dieser Art haben jedoch den Nachteil, dass für die Speicherung der Nutzerdaten und der Schlüssel weiterhin entsprechende Schutzvorkehrungen getroffen werden müssen, wie es bei herkömmlichen kryptographischen Anwendungen auch der Fall ist. Die Autoren beschreiben zusätzlich die Möglichkeit, den kryptographischen Schlüssel in den biometrischen Referenzdaten des Systems zu verbergen. Die erfolgreiche Authentifikation ermöglicht dabei die Extraktion des Schlüssels aus der zugehörigen Referenz. Als Problem stellt

sich dar, dass ein potentieller Angreifer die Position der Information innerhalb der Referenzdaten lokalisieren könnte.

Die Generierung eines kryptographischen Schlüssels auf Basis eines gesprochenen Passwortes schlagen Monrose et al. in [MRLW01] vor. Dafür wird ein 12-dimensionaler Vektor aus Cepstral Koeffizienten sowie ein text- und sprecherunabhängiges Akustik-Modell verwendet. In der weiteren Verarbeitung entsteht eine Segmentierung als Grundlage für drei verschiedene Arten von Merkmalen, die der Generierung so genannter Feature Descriptors dienen. Diese können als Hash-Wert des biometrischen Ansatzes verstanden werden.

Hao et al. nutzen in [HAD05] den Iris Code. Aus einem zufällig erzeugtem Schlüssel wird ein 2048 Bit langer Pseudo Iris Code mittels Error Correction Codes (Hadamard Code und Reed-Solomon Code) erzeugt. Dieser wird mit dem Referenz-Code (2048 Bit) der Iris des Nutzers über eine XOR-Operation verknüpft. Der dabei entstandene Code wird zusammen mit dem Hash-Wert des ursprünglichen Schlüssels gespeichert. Zur Erlangung des Schlüssel wird der aktuell präsentierte Iris Code verwendet. Ebenfalls mittels XOR wird ein Wert bestimmt, aus dem nach der Anwendung der Fehlerkorrektur in 99,5% der Fälle wieder der Schlüssel generiert werden kann. Der von Hao et al. vorgeschlagene Ansatz schützt einen zufällig generierten Schlüssel durch Nutzung von Redundanz zur Kompensation der Intra-Klassen-Varianz des Iris Codes.

In [SSM05] präsentieren Sutcu et al. ein Verfahren, das Referenzdaten für die biometrische Gesichtserkennung basierend auf Hash-Werten generiert. zugrunde liegt die Annahme, dass die Elemente des biometrischen Merkmalsvektors eines Nutzers jeweils um einen bestimmten Wert schwanken. Aufgrund dieser Annahme wird für jeden Nutzer eine bestimmte Anzahl von Gauss-Funktionen generiert. Die so ermittelten Werte werde als Referenzdaten gespeichert. Die Parameter der Transformationen werden auf einem persönlichen Token (zum Beispiel SmartCard) gespeichert, um für den Authentifikationsprozess abrufbar zu sein. Für einen potentiellen Angriff genügen sowohl die Daten in der Datenbank als aucg auf dem Token nicht, um daraus die Originaldaten abzuleiten.

Färberböck et al. beschreiben in [FHK$^+$10] ein Verfahren, um Cancelable Biometrics aus der biometrischen Modalität Iris zu generieren. Ausgangsdaten sind einerseits Aufnahmen des gesamten Auges inklusive Augenlid (Rectangular Iris Image) und zum Anderen Aufnahmen der Iris-Textur (Polar Iris Image) als Ergebnis der Iris-Detektion und der Umwandlung der Textur in Polarkoordinaten basierend auf dem entsprechenden Rectangular Iris Image. Als Transformationsverfahren werden Block Re-Mapping und Texture Warping eingesetzt.

Die Untersuchung der Generierung von Hashes aus unscharfen Eingabedaten beschränkt sich nicht auf die Biometrie. Zum Beispiel existieren Verfahren, die so genannte inhaltsbasierte Hash-Werte generieren (auch content-based Hashing, perceptual Hash, perceptual Fingerprint). Die nachfolgende Auswahl beschränkt sich auf die für uns relevanten Veröffentlichungen aus einer Vielzahl von Publikationen. Die Verfahren können entsprechend der Art der verwendeten Daten (zum Beispiel Audio, Bild, Video) klassifiziert werden.

Mihçak und Venkatesan schlagen in [MV01] ein Verfahren vor, das auf der Simulation der menschlichen Wahrnehmung basiert, um einen Hash-Wert von Audio-Dateien zu generieren. Im ersten Schritt wird aus dem Audiosignal eine Zeit-Frequenz-Repräsentation erzeugt (Modulated Complex Lapped Transform), gefolgt von der Bestimmung von Statistiken basierend auf den hörbaren Signalen. Eine Abbildung von ähnlichen Werten auf einen Punkt erfolgt im Anschluss unter Verwendung von Error Correction Codes. Als Ergebnis entsteht ein binärer Hash-Wert der Audio-Datei, der beispielsweise zur Identifikation von Audio-Dateien eingesetzt werden kann.

Ein System, welches für ein Bild einen binären String fester Länge erzeugt, wird von Monga und Evans in [ME04] vorgestellt. Der zugrunde liegende Algorithmus simuliert die menschliche Wahrnehmung eines Bildes. Dieser verwendet Feature Points, die gegenüber geringen Veränderungen invariant sind. Die bestimmten Merkmale werden durch ein probabilistisches Verfahren quantisiert. Es entsteht ein für das Bild individueller binärer Hash-Wert, der trotz Veränderungen des Bildes bis zu einem gewissen Grad identisch generiert werden kann. Die Einsatzgebiete liegen laut der Autoren in der Katalogisierung, in der Authentifikation und im Watermarking. Als Erweiterung des Ansatzes schlagen Monga et al. in [MBE04] einen Algorithmus vor, der das Image Hashing in die Merkmalsextraktion integriert (intermediate Hash). zusätzlich wird der endgültige Hashwert dann durch ein so genanntes Data Clustering (final Hash) erzeugt.

Die oben beschriebenen Verfahren zeigen, dass die Generierung von individuellen Hash-Werten basierend auf schwankenden Eingabedaten möglich ist. Entsprechend der Aussagen der Autoren verfügen diese Werte einerseits über eine genügende Reproduzierbarkeit beziehungsweise Trennschärfe. Auf der anderen Seite zeigen einige der Publikationen auch Probleme auf, die für entsprechende Hash-Funktionen basierend auf schwankenden Daten zu lösen sind.

Neben dem Schutz biometrischer Referenzdaten stellt die Verwendung biometrischer Hash-Werte zur Benutzerauthentifikation einen weiteren Anwendungsbereich dar. Dabei wird das Ausmaß der Ähnlichkeit des Referenz-Hashes und des aus den aktuell präsentierten Daten generierten Hashes verwendet, um eine Person zu authentifizieren. Vielhauer beschreibt in [Vie06] ein biometrisches Hash-Verfahren, welches Hash-Werte aus dynamischen Handschriftendaten bestimmt. Die Ähnlichkeit der Hashes zueinander wird dann mit Hilfe von Abstandsmaßen (zum Beispiel Canberra Distanz) bestimmt.

Ein weiterer Anwendungsbereich beschäftigt sich mit der Verbindung von kryptographischen Verfahren mit Vorteilen der Biometrie. Von Interesse kann hier sein, dass der Hash aus einer biometrischen Eigenschaft generiert werden kann, welche ständig zur Verfügung steht. Wird beispielsweise der biometrische Hash zur Erstellung von kryptographischen Schlüsseln herangezogen, können diese Schlüssel nach der Präsentation der entsprechenden biometrischen Modalität generiert werden. Das bedeutet, dass auf das Speichern von Informationen auf zusätzlichen Medien verzichtet werden kann.

Das Fuzzy Commitment Scheme von Juels und Wattenberg ([JW99]) verbindet während des Enrollment-Prozesses den Merkmalsvektor der Eingabedaten mit einem Codeword w, das auf

einem Error Correcting Code C basiert. Auf Grundlage eines biometrischen Eingabewertes x und des aus w bestimmten Hash-Wertes $h(w)$ wird die Differenz $x - w$ berechnet. Für eine spätere Authentifikation werden $h(w)$ und $x - w$ als Referenzdaten gespeichert. Werden von einer Person weitere biometrischen Daten x' präsentiert, wird davon $x - w$ subtrahiert. Dadurch wird w' bestimmt, welches die folgenden Eigenschaften besitzt: $w' = w + d$, mit $d = x' - x$. Stammt x' von derselben Person wie x, so sollten beide Werte in einem ausreichendem Maß ähnlich zueinander sein. Daraus folgt, dass w und w' ebenfalls dicht beieinander liegen, da: $x' - x = w' - w$. Ist die Error Correcting Capacity von C größer als der Abstand zwischen w und w', kann w wieder hergestellt werden. In diesem Fall ist die Authentifikation erfolgreich. Obwohl in der Arbeit ein biometrisches Authentifikationssystem beschrieben wird, zeigen Folgearbeiten, dass basierend auf dieser Idee auch kryptographische Ansätze möglich sind.

Maiorana et al. beschreiben in [MCN08] ein Verfahren zum Schutz von Referenzdaten basierend auf der Handschrift. Um der Intra-Klassen-Variabilität beziehungsweise der Inter-Klassen-Ähnlichkeit entgegenzuwirken, verwenden die Autoren die Mittelwerte der Merkmale der Samples der einzelnen Nutzer beziehungsweise aller Nutzer, um einen binären Merkmalsvektor zu erzeugen. Durch die Bestimmung derjenigen Merkmale, die für einen einzelnen Nutzer zu den besten Verifikationsergebnissen führen, ist es möglich, eine nutzerspezifische Parametrisierung zu erreichen.

In [KKM+08] nutzen die Autoren das Fuzzy Commitment Scheme für ein Authentifikationssystem basierend auf der DNA. Ziel ist sowohl der Schutz der Referenzdaten als auch die Generierung kryptographischer Schlüssel. Dazu wird die von Juels und Wattenberg in [JW99] vorgeschlagene Einschränkung auf binäre Elemente auf beliebige Einträge (die hier als Symbole bezeichnet werden) verallgemeinert. Der Algorithmus basiert auf der Nutzung von Short Tandem Repeats (STR), Wiederholungen kurzer Basenpaar-Muster innerhalb eines DNA-Stranges. Die erzeugten Referenzdaten (Genetic Fingerprint Template) besitzen eine Entropie von ca. 85 Bit.

Ein theoretisches Framework zur Generierung von kryptographischen Schlüsseln aus variierenden Daten wird von Dodis et al. in [DRS04] vorgeschlagen. Das vorgestellte Schema basiert auf zwei Strukturen, dem *Secure Sketch* und dem *Fuzzy Extractor*. Während der Sketch Generation wird aus den Eingangsdaten w der so genannte Secure Sketch P erzeugt. Dieser stellt innerhalb des Systems die Referenzdaten dar. Basierend auf den zu w hinreichend ähnlichen Daten w' und dem Secure Sketch P kann im Reconstruction-Prozess w reproduziert werden. Um die Sicherheit des Systems zu gewährleisten, ist es erforderlich, dass der Secure Sketch P so wenig Informationen wie möglich über die Ausgangsdaten w preisgeben kann. Dadurch wird vermieden, dass aus P Rückschlüsse auf w gezogen werden können. Der Generation-Prozess des Fuzzy Extractor ermöglicht es, aus unscharfen Eingabedaten w einen geheimen String R und einen öffentlichen String P zu erzeugen. Im Reproduction-Prozess kann unter der Voraussetzung, dass w und w' ausreichend ähnlich zueinander sind, aus w' und der öffentlichen Information P der geheime String R generiert werden.

Ein auf Secure Sketches basierendes Verfahren zum Schutz der Referenzdaten wird von Sutcu et al. in [SLM07] vorgeschlagen. Grundlage ist die Bestimmung von Merkmalsvektoren V der

biometrischen Modalität Gesicht während des Enrollments. Diese werden verwendet, um für jedes Merkmal den Mittelpunkt und den Bereich zu ermitteln, um die diese für eine beziehungsweise alle registrierten Personen variiert. Weiter wird eine Quantisierungsmatrix M erstellt, basierend auf den lokalen und globalen Extrema der einzelnen Merkmale. Zur Generierung des Sketch P werden die Eingabedaten V_1 quantisiert und für jedes Merkmal der entsprechende Wert ermittelt. Dieser berechnet sich aus der Differenz des korrespondierenden Codewortes eines Wertes und dem Wert selbst. Für die Rekonstruktion wird der aktuelle Merkmalsvektor V_2 ebenfalls quantisiert und zu jedem Element wird das korrespondierende Codewort ermittelt. Vom Codewort wird dann der entsprechende Wert des Sketches P subtrahiert. Die Rekonstruktion ist erfolgreich, wenn die Differenz der Codewörter innerhalb eines festzulegenden Toleranzbereiches liegt.

Ein Problem bei der Generierung von Secure Sketches kann auftreten, wenn aus den generierten Daten Informationen über die Eingabedaten gewonnen werden können. Ist dies zu einem gewissen Grad möglich, kann von einem Angreifer ein Eingabedatum erzeugt werden, mit dem das System über den Secure Sketch die ursprünglichen Eingabedaten generiert. Um dies zu verhindern, schlagen Fang et al. in [FLC10] einen Mix aus verschieden Secure Sketches zur Verringerung des Entropieverlustes vor. Durch diesen Mix ist es möglich, den Entropieverlust von Informationen mit einem höheren Schutzbedarf (zum Beispiel biometrische Daten) auf weniger schützenswerte Daten, beispielsweise Passwörter oder PINs, zu übertragen.

In [FFGO07] stellen Freire et al. eine Methode zur Hash-Generierung basierend auf der dynamischen Unterschrift vor, die auf Vektor-Quantisierung basiert und nutzerspezifische, sich optional überlappende Mengen von Merkmalsvektoren verwendet. Die Auswahl der geeigneten Merkmale erfolgt dabei mithilfe eines genetischen Algorithmus (Integer Coding Genetic Selection) auf Basis eines Testsets. Die Autoren setzen das Verfahren zum Schutz der Referenzdaten ein.

Karabat et al. präsentieren in [KEM11] ein Verfahren, welches die Bestimmung der Kapazität von binären biometrischen Hashes zur Benutzerauthentifikation ermöglicht. Die Methode bestimmt die maximal mögliche Anzahl der Nutzer, die von einem Authentifikationssystem basierend auf biometrischen Hashes erkannt werden können. Dazu wird angenommen, dass jeder biometrische Hash-Vektor als Referenzvektor möglich ist und die Wahrscheinlichkeit jeder Bit-Position identisch ist. Die durch Intra-Klassen-Variation von der Referenz abweichenden Hashes werden als Versionen der Referenz angesehen, die durch Noise entstehen. Evaluiert wird das Verfahren anhand verschiedener State of the Art-Algorithmen der Modalität Gesicht. Für einen der getesteten Algorithmen (siehe auch [NTG04], [NTG06]) ermitteln die Autoren beispielsweise für binäre Hash-Vektoren der Längen 64, 256 und 512 *bit* entsprechende Kapazitäten von rund 2^{26}, 2^{53}, 2^{105} beziehungsweise 2^{208}.

Mit dem Schutz biometrischer Referenzdaten und daraus resultierenden Interoperabilitätsproblemen befassen sich Busch et al. in [BAN+10]. Dabei gehen die Autoren neben generellen Sicherheitsaspekten in der Biometrie auch auf biometrische Systeme nach dem Standard ISO/IEC 24745 ([[IS10]) ein.

In [RU10] beschreiben die Authoren ein System, welches aus Iris-Texturen kryptographische Schlüssel ableitet. Ein Vorteil des beschriebenen Ansatzes liegt darin, dass keine biometrischen Daten gespeichert werden müssen. Dazu wird ein Interval Mapping durchgeführt, welches auf Merkmalen basiert, die durch ein Pixel-Block-Verfahren aus der Iris-Textur erzeugt wurden.

Das Ziel der Entwicklung des Semantic Web besteht darin, die Bedeutung von Informationen für Computer messbar und damit automatisch verwertbar zu machen. Dwivedi et al. presentieren in [DKDS11] einen Ansatz, biometrische Modalitäten im Semantic Web zu nutzen. Im Vordergrund stehen dabei die Erhöhung der Sicherheit und der Schutz vor Angriffen. Zu diesem Zweck schlagen die Autoren die Verwendung von Cancelable Biometrics in Kombination mit Access Policies vor.

Draper et al. präsentieren in [DKMV07] ein Verfahren zur Verschlüsselung von biometrischen Daten basierend auf dem Fingerabdruck mit dem Ziel des Template Protection. Es ermöglicht den Vergleich der aktuell präsentierten Daten mit den Referenzdaten, ohne letztere zu entschlüsseln. Zur Transformation der aus den Rohdaten extrahierten Merkmale werden spezielle Funktionen (Slepian-Wolf-Codes, siehe [SW73]) eingesetzt.

In [BCK09] stellen die Autoren eine Möglichkeit zur anonymen Identifikation in Kombination mit so genannten Cancellable Biometrics basierend auf dem Fingerabdruck vor. Die vorgestellte Methode verhindert zusätzlich festzustellen, ob eine Person mehrmals an einem biometrischen System identifiziert wurde oder nicht. Die Idee ist die zeitabhängige Umwandlung der biometrischen Daten in Referenzdaten durch eine Menge von Transformationen.

Jassim et al. beschreiben in [JAS09] ein Verfahren zum Schutz biometrischer Referenzdaten eines Gesichterkennungssystems. Sie verwenden orthonormale Random Projections, um aus dem Merkmalsvektor, welcher aus den biometrischen Daten mittels Wavelet-Transformation erzeugt wurde, die Referenzdaten zu bestimmen. Die dazu notwendige zufällige orthonormale Matrix wird über eine Rotationsmatrix erzeugt. Die Referenzdaten werden aus der Matrix, dem Merkmalsvektor und einem optionalen zufälligen Blinding-Vektor berechnet. Der Effekt der Erneuerbarkeit der Referenzdaten im Fall der Kompromittierung kann durch die Verwendung einer alternativen Rotationsmatrix und/oder eines anderen Blinding-Vektors erreicht werden.

In [RCCB07] präsentieren Ratha et al. unterschiedliche Verfahren zur Erstellung verschiedener Referenzdaten aus einer biometrischen Modalität. Sie adressieren damit die Bereiche Cancellable Biometrics und Tracking von Authentifizierungsverhalten einzelner Personen über mehrere biometrische Systeme und deren Datenbanken (Cross Matching). Die Autoren vergleichen empirisch die Performance der untersuchten Methoden bezüglich Cancellability und der Vermeidung von Cross Matching.

Eine Umsetzung von Cancellable Biometrics mittels des so genannten BioHash wird von Teoh et al. in [TKL08] beschrieben. BioHash kombiniert ein biometrisches Template (zum Beispiel Fingerprint [TNG04a], Face [TNG04b] oder Palmprint [CTGN04]) in Form eines geordneten Merkmalsvektors mit nutzerspezifischen Zufallszahlen. Diese Zufallszahlen werden auf einem Token gespeichert. Die grundlegende Idee ist, dass die Reproduktion des Merkmalsvektors ohne

die korrespondierenden Zufallszahlen (zum Beispiel auf einem physischen Medium) nicht möglich ist.

In [NNJ12] beschreiben Nagar et al. einen multi-biometrischen Ansatz für ein biometrisches Cryptosystem. Sie verwenden eine Feature Level Fusion basierend auf den Verfahren von Fuzzy Valt und Fuzzy Commitment mit dem Ziel, die Informationen mehrerer Modalitäten einer Person in einem Secure Sketch zu speichern und damit die Referenzdaten zu schützen. Die Autoren zeigen anhand von zwei Datenbanken basierend auf den Modalitäten Fingerabdruck, Gesicht und Iris, dass durch diesen Ansatz neben der erhöhten Sicherheit der Referenzen auch eine Verbesserung der Erkennungsgenauigkeit möglich ist.

Als Bereich der Crypto-Biometrics können biometrische Session Keys (auch One-Time Biometric Keys) genutzt werden, um Daten sicher zu transferieren. Im Folgenden werden Verfahren beschrieben, die den Einsatz biometrischer Session Keys ermöglichen.

Erkin et al. präsentieren in [EFG+09] ein Gesichtserkennungssystem. Das Protokoll basiert auf einem Eigenface-Erkennungsalgorithmus und der Verwendung von zwei homomorphen Public-Key-Verschlüsselungsverfahren. Im ersten Schritt sendet die erste der beiden beteiligten Parteien die homomorph verschlüsselten (mittels Paillier beziehungsweise Damgard, Geisler and Krøigaard Kryptosystem Verfahren) Parameter des Gesichtes, welche als Referenzdaten verwendet werden, ohne mehr als die dazu erforderlichen Parameter der Datenbank zu erfahren. Die zweite Partei hat Zugriff auf diese Referenzdatenbank, erhält aber keinerlei Informationen über das Bild oder das Ergebnis des Erkennungsprozesses. Erreicht wird dies durch die Verschlüsselung der biometrischen Daten, die es ermöglicht, die verschlüsselten Authentifizierungs- und Referenzdaten ohne Entschlüsselung zu vergleichen.

Ueshige und Sakurai schlagen in [US06] ein Framework für biometrische Session Keys vor, welches das Problem der sicheren biometrischen Authentifizierung über unsichere Übertragungswege adressiert. Dazu werden so genannte One Time Transformations (OTT) genutzt, die von einer vertrauenswürdigen dritten Partei (OTT Function Generator) generiert werden. Nach einer erfolgreichen Authentifizierung des Clientsystems und des biometrische Authentifizierungsservers beim OTT Function Generator, generiert dieser eine OTT, die an beide Parteien gesandt wird. Während der Client die Merkmalsdaten der aktuell präsentierten Modalität mittels der OTT transformiert, generiert der Server mit derselben OTT die korrespondierenden Referenzdaten.

In [LLCM07] präsentieren Lee et al. ein Gesichtserkennungssystem basierend auf One Time Templates. Um die Referenzdaten während des Enrollment-Prozesses zu transformieren, werden für jeden Nutzer und jede Anwendung (ermöglicht so auch Cancellable und Renewable Biometrics) eine zufällige orthogonale Matrix A und ein zufälliger Vektor b verwendet. Diese beiden Hilfsdaten werden auf einem physischen Token des jeweiligen Nutzers gespeichert. Während einer Authentifizierung präsentiert der Nutzer seine biometrischen Daten und das entsprechende Token. Basierend auf den Hilfsdaten A und b wird der aktuelle Merkmalsvektor auf dieselbe Weise transformiert wie die Referenzdaten. Um die One-Time-Funktionalität zu erreichen, werden die

Referenzdaten und die Transformationsparameter A und b nach einer erfolgreichen Authentifizierung erneuert.

Hsu et al. schlagen in [HLP12] die Verwendung von SIFT (scale-invariant feature transform), welches auf homomorpher Verschlüsselung basiert, zum Schutz der Privatsphäre vor. In der Evaluierung des Systems zeigen die Autoren neben Beispielen aus anderen Bereichen, dass die Nutzung auch für die biometrische Modalität Gesicht funktioniert. Verbesserungen in Bezug auf die Sicherheit des Systems werden beispielsweise in [SS14] von Schneider und Schneider vorgeschlagen. Besonderer Augenmerk wird hier darauf gelegt, dass beim Vergleich der verschlüsselten Informationen keine Interaktion zwischen den teilnehmenden Parteien notwendig ist, was bei ähnlichen Ansätzen in diesem Bereich häufig der Fall ist.

Umfangreiche Übersichten zu Forschungen im Kontext des Schutzes biometrischer Daten werden in der Literatur detailliert zum Beispiel von Jain et al. [JNN08], Cavoukian und Stoianov [CS09] als auch Rathgeb und Uhl [RU11] gegeben. Die genannten Arbeiten enthalten zusätzlich auch Untersuchungen zur Sicherheit der beschriebenen Systeme.

2.2.3 Optimierung

Neben der Erweiterung der Anzahl der verwendeten Merkmale existieren auch eine Vielzahl anderer Ansätze zur Optimierung biometrischer Verfahren. Die Analyse und Selektion von Merkmalen, die für den Einsatzzweck besonders geeignet sind, stellen eine Möglichkeit zur Optimierung biometrischer Systeme dar.

Eine Optimierung eines Systems zur Authentifikation mittels Fingerabdruck durch einen genetischen Algorithmus stellen Scheidat et al. in [SEV06] vor. Neben einer eigenen Datenbank wurde das Verfahren mit den verfügbaren Daten der Fingerprint Verification Contests der Jahre 2000, 2002 und 2004 evaluiert. Im besten Fall wurde durch die Verringerung der Equal Error Rate von 0,0620 auf 0,0369 eine Verbesserung von circa 40% erreicht.

Kumar und Zhang zeigen in [KZ05], dass mittels Merkmalsselektion die Erkennungsperformanz beeinflusst und gesteigert werden kann. Für die Analyse der Merkmale setzen sie ein auf Korrelation basierendes Verfahren (CFS, correlation based feature selection) ein und untersuchen die Klassifikationsperformanz der Merkmale.

In [HBF09] präsentieren Hollingsworth et al. ein Verfahren für ein irisbasiertes System. Hier werden potentielle fragile Bits des Iris-Codes durch eine Maskierung selektiert, was eine bessere Trennung der Verteilung von Übereinstimmungen beziehungsweise Nicht-Übereinstimmungen bewirkt.

Für die biometrische Modalität Gesicht schlagen Fratric und Ribaric in [FR11] eine Merkmalsextraktion vor, mit dem Ziel die Erkennungsrate zu verbessern. Das Verfahren basiert auf den positiven Eigenschaften einer Diskriminanzanalyse und einer lokalen Merkmalsextraktion und wird von den Autoren als lokale binäre Diskriminanzanalyse (LBLDA, engl. local binary linear

discriminant analysis) bezeichnet.

Basierend auf drei Unterschriftsdatenbanken, jeweils unterteilt in drei Altersgruppen (< 26, $26 - 60$ und > 60), zeigen Erbilek und Fairhurst, dass durch Merkmalsanalyse und -selektion bessere Ergebnisse in der Authentifikation zu erzielen sind [EF12]. Die Autoren verwenden zwei Ansätze, die zum einen auf der Korrelation der Merkmale untereinander und andererseits auf einer Distance Discriminant-Analyse basieren. Die Ergebnisse zeigen, dass die im Test besten Fehlerraten mit einer Anzahl von jeweils zehn bis elf Merkmalen erzielt werden, die Anzahl der ursprünglichen Merkmale beträgt 32.

Izquierdo et al. untersuchen in [IDVR14] Optimierungsansätze für ein biometrisches System basierend auf *Acoustic Images*. Die Bilder der Testpersonen werden mittels Radar- und Sonartechnik aufgenommen. Durch die Verwendung von vier Frequenzen und vier Positionen entstehen 16 Instanzen. Die Autoren zeigen, dass die Verwendung von nur 12 Acoustic Images zu ähnlichen Equal Error Rate führt wie die von 16.

In den letzten Jahren hat in der Biometrie die Erforschung von so genannten multi-biometrischen Systemen immer mehr zugenommen. Multi-biometrische Systeme kombinieren verschiedene biometrische Systeme oder Teile davon (im Folgenden auch als Komponenten oder Fusionskomponenten bezeichnet) miteinander. Ziel ist es, die Erkennungsgenauigkeit zu verbessern und damit die Sicherheit biometrischer Systeme zu erhöhen. Die Verwendung mehrerer biometrischer Komponenten kann die Probleme verringern, die durch Intra-Klassen Variabilität und Inter-Klassen Ähnlichkeit entstehen, da sich die verwendeten Einzelkomponenten gegenseitig ergänzen.

In [VS05] verbessert sich die Identifikationsrate um etwa 30% durch die Kombination von Handschrift und Sprache. Eine multi-sensor Fusion der Daten eines Stiftes zur Handschrifterfassung und eines Signatur-Tablet haben Scheidat et al. in [SVO07] vorgeschlagen. Dabei wurde beim Schreiben auf dem Drucksensitiven Tablet der Stift verwendet. Dadurch standen zwei unterschiedliche dynamische Repräsentationen desselben Schriftzuges zur Verfügung. Basierend auf diesem Vorgehen konnte im besten Fall eine Verbesserung um ca. 36% im Vergleich zum besten Ergebnis der Einzelsysteme erreicht werden.

In [GFA+07] wird ein multi-algorithmisches System vorgestellt, das auf vier unterschiedlichen Handschrift-Algorithmen basiert. Hier konnte durch die Fusion eine Verbesserung von etwa 42% erreicht werden.

Scheidat et al. schlagen in [SVD09b] die paarweise Fusion von unterschiedlichen Schreibinhalten (Semantiken) vor. Im beschriebenen System führt die Fusion zu einer Verbesserung von bis zu circa 57%, verglichen mit der Authentifikationsperformanz der beteiligten Einzelsysteme.

In [SVD07b] beschreiben Scheidat et al. eine Fusion, die auf jeweils zwei nacheinander aufgenommenen Schriftproben mit gleichem Inhalt basieren. Die Autoren berichten von einer Verbesserung von etwa 17% gemessen am besten Ergebnis ermittelt für die jeweils einzelnen Samples.

Chang et al. haben in [CBF03] ein System vorgestellt, das eine Kombination von 3D Range Data

und 2D-Bildern für die Gesichtserkennung nutzt. Die durch die Fusion erzielte Verbesserung beträgt ca. 1% verglichen mit der Erkennungsrate des besten beteiligten Einzelsystems.

Kisku et al. präsentieren in [KGS10] ein multi-biometrisches System, in dem eine Kombination aus Gesicht und Fingerabdruck zur Benutzerauthentifikation genutzt wird. Die Autoren berichten von einer Verbesserung von ca. 7%.

Einen generischen Fusionsansatz für die Modalität Iris schlagen Rathgeb et al. in [RUW11] vor. Dieses Selective Bit Fusion genannte Verfahren wird auf der Bitlevel-Ebene mit dem Ziel durchgeführt, neben der Erkennungsgenauigkeit auch die Verarbeitungsgeschwindigkeit zu erhöhen.

Eine eine multi-modale Fusion auf Feature Level präsentieren Nagar et al. in [NNJ12]. Die Autoren kombinieren die biometrischen Modalitäten Fingerabdruck, Iris und Gesicht. Auf diesen basierend wird zur sicheren Speicherung der Referenzdaten ein einzelner Secure Sketch pro Person erzeugt. Neben einer Verbesserung der Sicherheit zeigen die Ergebnisse der Fusion eine höhere Authentifikationsperformanz als die einzelnen Modalitäten.

2.2.4 Alterung und Auswirkungen der Alterung biometrischer Daten

Wie bereits in der Motivation dieser Arbeit dargelegt (siehe 1.1), unterliegt die Bevölkerung in vielen Industrieländern einem demografischen Wandel. Durch die biologische Alterung sind die physischen und mentalen Fähigkeiten einer ständigen Veränderung unterworfen, welche für jeden Menschen individuell ist. Dies bedingt im Kontext biometrischer Forschung auch eine Auseinandersetzung mit der Alterung biometrischer Daten. Wichtige Aspekte dabei sind Untersuchungen bezüglich der altersbedingten Veränderung biologischer Eigenschaften und deren Einfluss auf die verschiedenen biometrischen Anwendungsszenarien, die Fähigkeit der alleinigen/unterstützten Bedienung biometrischer Systeme durch bestimmte Altersgruppen oder die automatische/manuelle Erneuerung von Referenzdaten, die einen Alterungseffekt aufweisen (so genanntes Template Update).

Im Folgenden wird auf eine Auswahl von Studien und Publikationen eingegangen, die sich mit dem Einfluss des Alterns auf biometrische Modalitäten befassen. Ein Großteil der Untersuchungen entstand im Zusammenhang mit der Einführung biometrischer Pässe in Europa. Um einen Überblick über die relevanten Einflüsse der Alterung zu geben, haben wir eine Vielzahl an biometrischen und Schriftforensischen Studien und Publikationen unter diesem Gesichtspunkt ausgewertet. In den folgenden Abschnitten haben wir die wichtigsten Erkenntnisse zusammengefasst.

Entsprechend der berücksichtigten Studien funktioniert die Gesichtserkennung bei Kindern zum Teil sehr schlecht. Basierend auf dem in [ide05] beschriebenen System zur Gesichtserkennung gibt es Probleme mit der Erkennung von Gesichtern bei Kindern etwa bis zum vollendeten 13. Lebensjahr. Die niederländische Studie *Evaluation Report Biometrics Trail 2b or not 2b* [Min05] berichtet von einer problematischen Erkennung von Kindern unter zwölf Jahren, da sich bei ihnen charakteristische Gesichtsmerkmale noch nicht oder nicht genügend ausgebildet haben. Givens et al. berichten in [GBDB02] von einem statisch signifikanten Effekt der Alterung im Kontext der

Gesichtserkennung. Allgemein gesehen seien demzufolge ältere Menschen einfacher zu erkennen als jüngere Personen, wie mehrere Untersuchungen ergaben ([GBDB02], [PGM+03], [Fra03]). Dabei wurde festgestellt, dass sich die Identifikationsrate mit zunehmenden Alter erhöht. Fraser [Fra03] berichtet von zwei bis drei Prozent alle zehn Jahre, Phillips et al. [PGM+03] sprechen von circa fünf Prozent in demselben Zeitraum. Weiter wird die durchschnittliche Identifikationsrate von Phillips et al. [PGM+03] für 18 bis 22 Jährige mit etwa 62%, für 38 bis 42 Jährige mit 74% und für 73 bis 74 Jährige mit 82% beziffert. Auf der anderen Seite berichtet der UK Passport Service in seinem *Biometrics Enrolment Trail* [UK 05], dass die Verifikationsrate des untersuchten Gesichtserkennungssystems für Personen unter 60 Jahren besser ausgefallen ist, als für ältere Menschen. O'Toole et al. berichten in [OPV+99], dass sich die charakteristischen Merkmale eines Gesichtes mit zunehmenden Alter entwickeln. Basierend auf dieser Annahme sind aus Sicht eines biometrischen Systems die Gesichter von Kindern ähnlicher als die von älteren Menschen. Dementsprechend sollte die Erkennung mit zunehmenden Alter einfacher werden, zusammen mit abnehmender Ähnlichkeit. Dieser Zusammenhang spiegelt sich auch in den meisten recherchierten Studien wieder, wobei [UK 05] eine Ausnahme bildet. Dort wird beschrieben, dass Menschen über 60 eine geringere Verifikationsrate aufweisen als jüngere Testpersonen. Als Gründe werden unter anderem Schwierigkeiten in der Positionierung zum biometrischen System und in dessen Bedienung genannt.

Die meisten der von uns recherchierten Publikationen berichten davon, dass die Erfassung brauchbarer Aufnahmen von Fingerabdrücken von Kindern unter sechs Jahren sehr schwierig ist. Speziell war es nicht möglich, ein Bild eines Fingerabdruckes in ausreichender Qualität von Kindern unter vier Jahren aufzunehmen. In den wenigen Fällen, in denen dies gelungen ist, war es meistens der Daumen aufgrund seiner größeren Fläche. Elliott et al. berichten in [EKS09], dass es häufig notwendig war, bei Testpersonen im Alter von sechs bis zwölf den Daumen beziehungsweise den Mittelfinger zu erfassen. Die Autoren der Studie *BIODEVII* [[BI08] fanden für Kinder von sechs bis zwölf Jahren heraus, dass in ihrem Fall eine höhere Sensorauflösung (mindestens 500 dpi) zu einer ausreichenden Qualität der Abdrücke bei beiden Zeigefingern führte. Die Qualität der so entstandenen Bilder wurde mit einem NFIQ[2] von 1 bis 3 angegeben. Die meisten Studien berichten, dass die Qualität der Fingerabdrücke sich mit zunehmenden Alter bis zum 14. Lebensjahr erhöht, bis circa 60 Jahre relativ stabil bleibt und für ältere Menschen dann wieder abnimmt. Einerseit zeigen die Untersuchungen, dass die Qualität der Fingerabdrücke für die Altersgruppe 15-60 höher ist als für die Gruppe der sechs bis 14 Jährigen. Andererseits wurde die schlechteste Qualität für die Fingerabdrücke der Altersgruppe 61-80 ermittelt. Entsprechend Elliott et al. [EKS09] war es in den meisten untersuchten Fällen in der Altersgruppe 0-11 nicht erfolgreich, einen oder zwei Fingerabdrücke mit einem NFIQ von 1 bis 3 zu erfassen. Das gleiche gilt auch für die Altersgruppe der über 60 Jährigen. Ein Zusammenhang zwischen der Hautfeuchtigkeit der Fingerspitzen und dem Alter der Testpersonen wurde in [EKS09], [Sic05] und [BSI05] nachgewiesen. Es wird berichtet, dass die Finger älterer Menschen trockener sind als die jüngerer.

[2]NFIQ (NIST Fingerprint Image Quality) stellt ein Qualitätsmaß für biometrische Fingerabdrücke dar, wobei 1 für *excellent*, 2 für *very good*, 3 für *good*, 4 für *fair* und 5 für *poor* steht. Detaillierte Informationen zu NFIQ sind für den interessierten Leser beispielsweise in [TWW04] zu finden.

Trockenere Haut führt bei der Aufnahme von Fingerabdrücken dazu, dass die dabei entstehenden Bilder weniger Kontrast aufweisen und daher in der Qualität als schlechter eingestuft werden.

Ein weiterer Aspekt für die Qualität ist die Abnutzung der Haut, da die Qualität der Aufnahmen teilweise von der Intensität der Struktur der Oberfläche des Fingers abhängt. Bei älteren Personen kann es dazu kommen, dass die Erhebungen der Abdrücke weniger intensiv sind [BSI05]. Basierend auf Authentifikationstests beschreiben die Studien [MKCK01], [BSI05] und [MEKW07] Abhängigkeiten der Systemperformanz von den untersuchten Altersgruppen. In diesem Zusammenhang wird auch von einer geringeren False Rejection Rate bei jüngeren Personen berichtet. Die Wahrscheinlichkeit, einen NFIQ-Wert schlechter als 3 zu erhalten ist für Menschen über 65 Jahre sehr hoch [EKS09], wodurch eine fehlgeschlagene Verifikation bei Personen dieser Altersgruppe leicht möglich ist. Diese Aussagen werden auch durch den UK Passport Service in seinem *Biometrics Enrolment Trail* [UK 05] bestätigt: Hier sind ältere Personen bei der Verifikation mittels Fingerabdruck weniger erfolgreich als jüngere. Die Genauigkeit verringert sich folglich mit zunehmenden Alter der Testpersonen, da sie von der Qualität der Aufnahmen abhängig ist. Elliott et al. haben in [EKS09] festgestellt, dass die Verwendung von Aufnahmen von Fingerabdrücken mit hoher Qualität allgemein zu besseren Ergebnissen führt als die Nutzung qualitativ schlechterer Bilder. Eine Abhängigkeit zwischen Bildqualität und der Anzahl von False Matches wird auch in [ME06] berichtet. Entfernt man also Aufnahmen schlechter Qualität, wird sich die Anzahl von False Non Matches verringern. In der Studie *Biometric Data Experimented in Visas* [[BI08] wird gezeigt, dass die Fehlerrate für Kinder 2.3% beträgt, während sie für Erwachsene und ältere Menschen 2.4% beziehungsweise 4% beträgt.

Die Muster der Iris werden schon vor der Geburt gebildet und verändern sich, mit Ausnahme von Krankheit oder Verletzung des Augapfels, ein Leben lang nicht mehr ([NTN02], [Dau08]). Trotz dieser Stabilität zeigen die Ergebnisse der betrachteten Studien Unterschiede zwischen den Altersgruppen in Bezug auf die Erfassung und Erkennung. Die Experimente des UK Passport Service für den *Biometrics Enrolment Trail* [UK 05] ergaben, dass die Aufnahme der Iris für behinderte Menschen schwer war. Sie waren oft nicht in der Lage, sich vor dem Sensor richtig zu positionieren, geduldig und lang genug in die richtige Richtung zu sehen. Schwierigkeiten gab es ebenfalls bei hörgeschädigten Personen, da sie nicht immer den gegebenen Anweisungen folgen konnten. Ebenfalls wurde teilweise die benötigte Zeit für die Datenaufnahme als zu lang empfunden. Die Studie berichtet weiterhin, dass die allgemeine Rate für erfolgreiche Enrollments beziehungsweise die Rate für ein erfolgreiches Enrollment im ersten Versuch für junge Personen höher war als für ältere Menschen.

Corby et al. untersuchen in [CSS+06] das Enrollment der Iriden von Kindern im Alter von 1,5 bis 8 Jahren. Das Ergebnis zeigte, dass dies in 24% der Fälle nicht möglich war. In den meisten Fällen fiel es vor allem jüngeren Teilnehmern schwer, ihre Augen lange genug in der richtigen Position zu halten. Zum Vergleich, nur bei 6% der Erwachsen schlug das Enrollment fehl. Eine Schlussfolgerung der Autoren ist, dass eine effektive biometrische Erkennung basierend auf der Iris erst ab drei Jahren möglich ist. Eine Abhängigkeit der Erkennungsraten vom Alter

wurde während der Studie *BioP II - Untersuchung der Leistungsfähigkeit von biometrischen Verifikationssystemen* [BSI05] festgestellt. Dabei wurden für die 30 bis 50 Jährigen die besten Ergebnisse erzielt, während die Testpersonen über 50 sowie die unter 20 am schlechtesten abschnitten. Die Untersuchung zeigte ebenfalls, dass Brillenträger signifikant höhere False Recognition Rates aufwiesen als die Testpersonen ohne Brille. Ursachen dafür können im Reflektions- und Brechungsverhalten der Brillengläser vermutet werden. Darüber hinaus kann nach Park und Lee [PL05] die Erkennungsgenauigkeit, abhängig vom verwendeten Algorithmus, durch die Augenlider oder die Augenbrauen beeinflusst werden. Diese können sich mit zunehmenden Alter ebenfalls verändern. Dieser Umstand wird auch von Ma et al. in [MTWZ04] beschrieben. Die Autoren berichten, dass etwa 57,7% der False Non Matches durch die Verdeckung durch Augenlider und Augenbrauen bedingt werden. Die Erfolgsrate der Verifikation ist nach [UK 05] für ältere Menschen über 55 Jahre geringer als für Personen, die jünger sind. In letzte Zeit beschrieben einige Publikationen Beobachtungen über die Alterung von biometrischen Referenzdaten. Zum Beispiel berichten Fenker et al. in [FOB13] von einer Erhöhung der False Non Match Rate auf circa 150% über einen Zeitraum von drei Jahren.

Bisher wurde der Einfluss der Alterung auf die Handschrift noch nicht so umfangreich im Kontext der biometrischen Authentifikation untersucht, wie dies bei den oben genannten Modalitäten der Fall ist. Aus diesem Grund haben wir vor allem medizinische und forensische Publikationen, die sich mit der Handschrift befassen, herangezogen, um einen Überblick über mögliche Effekte zu erhalten, die durch die biologische Alterung hervorgerufen werden können. Ebenso wie die anderen drei betrachteten Modalitäten verändert sich die Handschrift mit dem Alter. Aber im Gegensatz zu den Modalitäten Gesicht, Fingerabdruck und Iris wird die Handschrift in den meisten Fällen als dynamische Biometrie verwendet. Aus diesem Grund ist sie nicht nur abhängig von physischen Eigenschaften des Inhabers sondern auch von dessen mentalen Fähigkeiten. Birren und Botwinick beschreiben in [BB51] altersabhängige Änderungen in der Schreibgeschwindigkeit. Demzufolge verringert sich die Geschwindigkeit des Schreibvorgangs bei Personen Mitte der 50. Entsprechend anderer Publikationen benötigen ältere Menschen mehr Zeit zum Schreiben [SPB96], schreiben mit geringerer Geschwindigkeit und weniger Druck [RW06] und produzieren mehr instabile Bewegungen [SPB96] verglichen mit jüngeren Menschen. Auf der anderen Seite schreiben junge Menschen mit höherer Geschwindigkeit und verfügen über einen allgemein weicheren Schreibprozess, was sich durch weniger Fluktuationen in der Stiftgeschwindigkeit ausdrückt [MTS+99].

In [Gue06] untersucht Guest mögliche Abhängigkeiten von biometrischen Unterschriftensystemen basierend auf dynamischen Merkmalen der Handschrift. Der Autor fand dabei keine signifikanten Abhängigkeiten zwischen den untersuchten Altersgruppen bei der Nutzung eines biometrischen Systems zum Enrollment und zur Verifikation. Es konnten aber Unterschiede bezogen auf die Altersgruppen und den Merkmalen gefunden werden, die sich auf die Schreibdauer und Stiftdynamik beziehen (zum Beispiel Geschwindigkeit, Beschleunigung). Mit zunehmenden Alter der Testpersonen sinkt die Geschwindigkeit und Beschleunigung des Stiftes. Dabei nahmen die Schreibdauer und die Anzahl der Stiftabsetzer zu. Zusätzlich wurde festgestellt, dass sich die

Reproduzierbarkeit der Unterschrift nicht signifikant mit zunehmenden Alter verändert. Erbilek und Fairhurst untersuchen in [EF12] die Auswirkungen des physischen Alterns anhand von drei biometrischen Unterschiftendatenbanken. Die Ergebnisse zeigen hier, dass die Fehlerraten für die Altersgruppe > 60 im Vergleich zu den Gruppen < 26 und $26 - 60$ immer am besten ausfallen. Im dritten Fall schneiden diese ebenso am besten ab, hier kann aufgrund der zur Verfügung stehenden Daten nur mit der Gruppe $26 - 60$ verglichen werden. Die Autoren begründen ihre Beobachtung damit, dass ältere Personen mehr Übung und Erfahrung im Schrieben haben als jüngere. Dies wird beispielsweise von Galbally et al. in der Form bestätigt, dass sie davon berichten, dass eine höhere Übung der Schreibbewegungen zu besseren Fehlerraten führen [GFMP11]. Wie Michel in [Mic82] berichtet, ist die Unterschrift resistenter gegenüber Veränderungen als andere Schreibinhalte. Der Grad des Einflusses von Alter und durch Krankheiten hervorgerufener Phänomene auf die Unterschrift ist ebenfalls geringer. Als Grund gibt Michel die relativ häufige Verwendung der Unterschrift an, wodurch deren Schreibprozess weitestgehend automatisiert wird.

Wie die in den vorhergehenden Abschnitten genannten Beispiele zeigen, sind Auswirkungen zwischen der biologischen Alterung und der Performanz biometrischer Systeme nachweisbar. Die Art und der Umfang dieser Effekte sind dabei von der biometrischen Modalität und den jeweiligen Altersgruppen abhängig. Diese Aspekte sollten bei der Entwicklung, Evaluierung und beim Einsatz biometrischer Systeme berücksichtigt werden. Im normalen Anwendungsfall ist davon auszugehen, dass zwischen der Erfassung der Enrollment-Daten und der Verifikationsdaten eine gewisse Zeitspanne vergeht. Treten in diesem Zeitraum markante Veränderungen auf, beispielsweise hervorgerufen durch normale biologische Alterung oder Krankheit beziehungsweise Verletzung, ist unter Umständen eine Benutzung des biometrischen Systems für die betreffende Person nicht mehr möglich.

Es werden in der Literatur Verfahren präsentiert, die eine Aktualisierung der Templates ermöglichen. Scheidat et al. schlagen in [SMV07] verschiedene Ansätze vor, die Strategien zur Verwaltung von Daten im Cache adaptieren, beispielsweise *First in First Out (FIFO)* oder *Least Recently Used (LRU)*. Rattani et al. präsentieren ein *graph min-cut algorithm* genanntes Verfahren, welches zu Ziel hat, Daten mit einer signifikanten Intra-Klassen-Varianz zu erfassen, um die Referenzdaten zu aktualisieren [RMR13].

3 Herausforderungen, Anforderungen und Lösungsansätze

In diesem Kapitel wird zusammenfassend dargelegt, welche aktuellen Herausforderungen es im Bereich der Hash-Generierung basierend auf der dynamischen Handschrift gibt. Darauf aufbauend legen wir Lösungsansätze zur Erfüllung dieser Herausforderungen dar und präsentieren die für diese Arbeit zur weiteren Untersuchnug getroffene Auswahl.

3.1 Herausforderungen und Anforderungen

Die dynamische Handschrift verfügt über großes Potential zum Einsatz in biometrischen Systemen. Der große Vorteil liegt hier in der einfachen Abgabe von Schriftproben und der potentiell hohen Akzeptanz durch die Anwender. Diese begründet sich darin, dass die Handschrift seit Jahrhunderten zur Sicherung der Authentizität und zur Bekundung des Willens eingesetzt wird, beispielsweise bei der Unterzeichnung von Verträgen. Dies birgt für die biometrische Forschung aber auch den Nachteil, dass eine Vielzahl der Probanden davor zurückschrecken, ihre tatsächliche Unterschrift abzugeben. Hier bietet sich die Verwendung von alternativen Schreibinhalten an, allerdings mit der Anforderung, dass diese eine ähnlich hohe Performanz wie die Unterschrift aufweisen, bezogen auf den jeweiligen Verwendungszweck.

Für die beiden in dieser Arbeit fokussierten Anwendungsbereiche, Verifikation und Hash-Generierung, sollen Möglichkeiten zur Optimierung gefunden, untersucht und evaluiert werden. Dazu werden wir beispielhaft zwei biometrische Algorithmen für die dynamische Handschrift betrachten, welche die beiden genannten Applikationsmodi beherrschen. Nachteilig wirkt sich in diesen zwei Bereichen der Biometrie eine relativ hohe Intra-Klassen-Variabilität aus. Da die dynamische Handschrift eine Biometrie ist, die auf dem Verhalten beruht, können einzelne Proben einer Person teilweise stark voneinander abweichen. Ein biometrischer Algorithmus für die dynamische Handschrift sollte folglich in der Lage sein, trotz dieser Variationen in beiden Modi Personen zu erkennen beziehungsweise zwischen unterschiedlichen Benutzern zu differenzieren.

Auf der einen Seite sind für die Verifikation einer Person Abweichungen der einzelnen korrespondierenden Elemente des erzeugten Hash-Vektors im Rahmen eines vorgegebenen Bereiches möglich. Ist die Ähnlichkeit geringer, kann die Identität der Person folglich nicht bestätigt werden. Die Hash-Generierung setzt andererseits eine identische Reproduktionen der Hash-Vektoren für eine Person voraus. Die Anforderungen an die Vergleiche zwischen den Daten unterschiedlicher Personen sind für beide Modi jeweils entsprechend. Bei der Hash-Generierung wird jede exakte

Übereinstimmung zwischen Hashes unterschiedlicher Personen als Kollision registriert. Weichen die Hash-Vektoren zweier Nutzer nur um ein einzelnes Element ab, handelt es sich nicht um eine Kollision. Das hat zur Folge, dass für beide Modi unterschiedliche Performanzmaße eingesetzt werden müssen. Ebenfalls ist davon auszugehen, dass Verifikation und Hash-Generierung durch unterschiedliche Ansätze zur Optimierung abweichend beeinflusst werden. Hier ist neben einer entsprechenden Datenbasis auch eine Methodologie zu entwerfen, die dem Rechnung trägt. Dies umfasst auch die Wahl beziehungsweise Entwicklung und Umsetzung geeigneter Performanzmaße für beide Operationsmodi.

Damit stellt sich das Hauptziel dieser Arbeit als Verbesserung der Performanz beider Algorithmen dar. In diesem Zusammenhang umschreibt der Ausdruck *Performanz* die Verringerung der Fehleranfälligkeit sowohl im Verifikations- als auch im Hash-Generierungsmodus. Bei den dabei untersuchten Fehlern handelt es sich zum einen um die mögliche Nichterkennung berechtigter beziehungsweise Falscherkennung nicht berechtigter Personen während der Verifikation. Bei der Hash-Generierung andererseits untersuchen wir Fehler die auf der Falschgenerierung von Hashes einer Person beziehungsweise durch Kollisionen von Hashes unterschiedlicher Personen beruhen. Diese Fehler sollen im Rahmen dieser Arbeit basierend auf geeigneten Verfahren zur Merkmalserweiterung, Parameteroptimierung, Merkmalsanalyse und -selektion, Fusion und der Untersuchung von Alterungsaspekten auf Basis der Algorithmen Biometric Hash und Secure Sketch für die biometrische Modalität der dynamischen Handschrift empirisch überprüfbar minimiert werden.

3.2 Mögliche Lösungsansätze

Um die Performanz biometrischer Systeme zu verbessern, existieren verschiedene Ansatzpunkte. Weit verbreitet ist das Hinzufügen von zusätzlichen Merkmalen. Basierend auf unseren Erfahrungen ist dabei in den meisten Fällen mit einer Verbesserung der Verifikationsperformanz zu rechnen. Allerdings können im Kontext der Hash-Generierung neue Merkmale bedeuten, dass diese, zusätzlich zu den bereits vorhandenen, nicht stabil in den Hash-Raum transferiert werden. Damit bleibt die Wahrscheinlichkeit der Reproduzierbarkeit individueller Hashes im günstigsten Fall gleich beziehungsweise verschlechtert sich andernfalls. Um dem entgegenzuwirken, kann eine Analyse der Merkmale durchgeführt werden.

Eine Möglichkeit, hier Abhilfe zu schaffen, bieten geeignete Verfahren zur Analyse der Merkmale und eine darauf aufbauende Selektion. Diese Ansätze untersuchen die Merkmale, inwiefern sie allgemein beziehungsweise für einen speziellen Verwendungszweck einsetzbar sind. Im Allgemeinen können solche Untersuchungen in Abhängigkeit sowohl von den verwendeten Verfahren als auch von den Merkmalen untereinander durchgeführt werden. Basierend auf den Ergebnissen dieser Analysen können dann Merkmale aus dem Set entfernt werden, die die Performanz des jeweiligen Modus beziehungsweise Algorithmus negativ beeinflussen oder eventuell auch gar keinen Einfluss auf diesen haben.

Weitere Optimierungen der Verifikations- und Hash-Generierungsperformanz können basierend

auf dem jeweiligen Algorithmus aufbauen. So können beispielsweise durch eine Anpassung von Parametern des Algorithmus (sofern solche existieren) in Abhängigkeit vom anvisierten Anwendungsszenario Verbesserungen erreicht werden. Die Parameter können beispielsweise Einfluss auf die Erzeugung der Merkmale nehmen oder den Vergleich von Referenz- und aktuell präsentierten Authentifikationsdaten nehmen. Je nach Algorithmus kann dieser gegebenenfalls auch auf Implementierungsebene entsprechend der Zielanwendung angepasst werden. Denkbar sind hier zum Beispiel alternative Abstandsmaße zur Bestimmung des Matching Score.

Auch die biometrische Fusion kann zur Untersuchung möglicher Optimierungsansätze zur Verbesserung der Performanz von Authentifikation beitragen, wie in diversen unserer Vorarbeiten und Publikationen Dritter gezeigt wird. Ob dies auch für die Generierung von biometrischen Hashes gilt, muss in entsprechenden Untersuchungen noch gezeigt werden. Die Fusion kann mittels mehrerer Modalitäten oder aber auch nur mit einer einzelnen durchgeführt werden.

Auch eine Veränderung der Qualität der einem Verfahren zugrunde liegenden Sensordaten kann bei der Optimierung hilfreich sein. So können Bilder mit höherer Auflösung für eine bessere Erkennungsrate bei einem gesichtbasierten System sorgen. Für die biometrische Modalität der Handschrift können beispielsweise aktuelle Tablets mit höherer sensorischer Auflösung von X und Y beziehungsweise höherer Taktfrequenz zu besseren Authentifikationsergebnissen führen. Entsprechende Ansätze können aber zur Folge haben, dass gegebenenfalls bessere Hardware zum Einsatz kommen muss, die Algorithmen angepasst werden beziehungsweise ältere Referenzdaten durch neue ersetzt werden müssen.

3.3 Ausgewählte Lösungsansätze

Als biometrische Verfahren haben wir den Biometric Hash-Algorithmus nach Vielhauer et al. [VSM02] und den Secure Sketch-Algorithmus für die Handschrift nach Scheidat et al. [SVD09a] [MSV12] (basierend auf [SLM07]) ausgewählt. Beide Verfahren bieten die erforderlichen Betriebsmodi Verifikation und Hash-Generierung. Als Grundlage für die zwei Algorithmen dient ein identisches Set von 131 statistischen Merkmalen. Verbale und formale Beschreibungen von Biometric Hash und Secure Sketch für die dynamische Handschrift sind im Abschnitt 4.1 *Biometric Hash-Algorithmus für die dynamische Handschrift* beziehungsweise im Abschnitt 4.2 *Secure Sketch-Algorithmus für die dynamische Handschrift* zu finden. Anhand dieser Referenzalgorithmen untersuchen wir verschiedene Ansätze zur Optimierung von biometrischen Verfahren im Kontext der beiden potentiellen Anwendungsszenarien Benutzerverifikation und Generierung von Hashes. Im Folgenden beschreiben wir kurz die von uns gewählten Verfahren zur Optimierung beider Algorithmen. Dabei handelt es sich um eine beispielhafte Auswahl aus einer Vielzahl von Möglichkeiten. Weiterhin gehen wir auf die von uns gewählten beziehungsweise entwickelten Maße zur Bestimmung der Performanz ein. Zusätzlich dazu werden wir ebenfalls Untersuchungen zur Kurzzeitalterung biometrischer Daten und der daraus resultierenden Effekte motivieren.

3.3.1 Aufgabenstellung A.1 bis A.4 - Optimierung

Für die Optimierung von biometrischen Algorithmen haben wir entsprechend der Aufgabenstellungen *A.1 - A.4* eine Auswahl von Verfahren aus vier verschiedenen Hauptbereichen getroffen, die sich für die Anwendung sowohl bezüglich der von uns eingesetzten Algorithmen als auch der beiden Anwendungsszenarien Benutzerverifikation und Hash-Generierung eignen. Dabei handelt es sich um das Hinzufügen neuer Merkmale, die Parametrisierung der Algorithmen, die Untersuchung der statistischen Merkmale und die Fusion biometrischer Komponenten.

A.1 Hinzufügen zusätzlicher Merkmale

Im Rahmen unserer Untersuchungen sind dazu im ersten Schritt dieses Teils der Optimierung zu den anfänglich zur Verfügung stehenden 69 [Vie06] weitere 62 Merkmale konzipiert und implementiert worden. Diese bilden die Grundlage für jeweils beide Betriebsmodi von Biometric Hash und Secure Sketch. Die neu hinzugefügten 62 Merkmale werden im Abschnitt 4.3 *Statistische Merkmale für beide Referenzalgorithmen* beschrieben. Die Präsentation und Auswertung der Evaluierungsergebnisse sowohl der schrittweise konkatenierten Merkmal-Sets als auch des Verlaufs der Performanzmaße Equal Error Rate und Collision Reproduction Rate bei zunehmender Anzahl von Merkmalen erfolgt in Abschnitt 7.1 *Evaluierung: Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale.*

A.2 Optimierung von Parametern

Im Rahmen der Optimierung der Parameter wird für beide Algorithmen jeweils ein skalarer Wert betrachtet, von dem die Abbildung der Werte beim Mapping abhängig ist. Beim Biometric Hash-Algorithmus ist dies der Tolerance Factor TF während es sich beim Secure Sketch-Verfahren um den Expansion Factor EF handelt. Da bei einem Teil der weiteren hier untersuchten Optimierungsstrategien die bezüglich der Parametrisierung gewonnenen Erkenntnisse einfließen sollen, beschränken wir uns auf jeweils einen Parameter pro Algorithmus. Anderenfalls würde sich mit jedem weiteren Parameter die Anzahl der in der Evaluierung zu berücksichtigenden Test-Setups vervielfachen. Die Bestimmung der Parameter wird in Abschnitt 5.2 *Optimierung von Parametern (A.2)* erläutert, während die Parameterwerte und die daraus resultierenden Ergebnisse in Abschnitt 7.2 *Evaluierung: A.2 Optimierung von Parametern* vorgestellt und diskutiert werden.

A.3 Analyse und Selektion von Merkmalen

Bei der Untersuchung der statistischen Merkmale konzentrieren wir uns neben dem Hinzufügen neuer auf das Auswählen beziehungsweise Entfernen vorhandener Merkmale. Eine statistische Analyse der Merkmale mit dem Ziel, die für den jeweiligen Verwendungszweck relevanten zu identifizieren und zur weiteren Nutzung zu selektieren, wird dann basierend auf den gesamten 131 Merkmalen durchgeführt. Dabei werden auch die im vorherigen Optimierungsschritt ermittelten Parameter-Sets berücksichtigt. Für die statistische Analyse und die daraus resultierende Selektion

der Merkmale werden Wrapper und Filter (siehe Abschnitt 2.1.5 *Optimierung basierend auf der Merkmalsebene*) verwendet. Dabei untersuchen Wrapper die Merkmale basierend auf dem zugrunde liegenden Algorithmus, während Filter die Merkmale unabhängig vom Algorithmus analysieren. Die einzelnen im Zusammenhang von Verifikations- beziehungsweise Hash-Generierungsmodus der Verfahren Biometric Hash und Secure Sketch erforschten Wrapper- und Filter-Verfahren werden im Abschnitt 5.3.2 vorgestellt. Die Ergebnisse, die für die Merkmalsselektion während der experimentellen Evaluierung ermittelt wurden, werden in Abschnitt 7.3 *Evaluierung: A.3 Analyse und Selektion von Merkmalen* präsentiert und diskutiert. Diese Auswertung wird aus unterschiedlichen Blickwinkeln durchgeführt. Zum einen werden die erreichten Ergebnisse für die einzelnen Konstellationen (zum Beispiel abhängig von Algorithmen, Semantiken, Parametern) miteinander verglichen. Zweitens werden Rankings betrachtet, die abhängig von Filter und Wrapper erstellt wurden. Drittens erfolgt eine Auswertung der Häufigkeit, mit der die durch einzelne Analyseverfahren ermittelten Merkmal-Sets das jeweils beste Ergebnis für die einzelnen Konstellationen ermittelten.

A.4 Biometrische Fusion

Als weitere Möglichkeit der Optimierung untersuchen wir die biometrische Fusion. Dabei haben wir uns auf die Kombination von handschriftbasierten Komponenten konzentriert. Diese so genannte single-modal Fusion nutzt nur eine einzelne Modalität, in unserem Fall die Handschrift. Dafür wurden aus einer Vielzahl von potentiellen Kombinationen beispielhaft vier Möglichkeiten ausgewählt. Die allgemeinen Grundlagen der Fusion werden in Abschnitt 2.1.5 *Optimierung durch biometrische Fusion* beschrieben. Bei unseren Ansätzen zur Optimierung untersuchen wir bei der Fusion beider Referenzverfahren die Kombination der Algorithmen auf Matching Score Level (5.4.2 *Handschriftbasierte multi-algorithmic Fusion auf Matching Score Level*). Die Kombination von zwei Instanzen der Handschrift wird in Abschnitt 5.4.3 *Handschriftbasierte multi-instance Fusion auf Matching Score Level* erläutert. Instanzen sind im Kontext unserer Arbeiten unterschiedliche Semantiken, die miteinander paarweise fusioniert werden. Demgegenüber kombiniert die multi-sample Fusion zwei unterschiedliche Samples einer einzelnen Semantik miteinander, wie wir in Abschnitt 5.4.4 *Handschriftbasierte multi-sample Fusion auf Matching Score Level* beschreiben. Die Ergebnisse der experimentellen Evaluierung der Fusionsstrategien und deren Auswertung sind in Abschnitt 7.4 *Evaluierung: A.4 Biometrische Fusion* zu finden.

3.3.2 Aufgabenstellung B Alterung biometrischer Daten

Während der Untersuchungen bezüglich möglicher Optimierungsstrategien im Rahmen dieser Arbeit fiel uns vor allem im Kontext der biometrischen Fusion auf, dass die erzielten Verbesserungen zum Teil geringer ausfielen als in unseren früheren Arbeiten. Als eine mögliche Ursache haben wir nach eingehender Betrachtung die Alterung der biometrischen Referenzdaten identifiziert. Im Gegensatz zu früheren Untersuchungen zur Fusion verwenden wir hier zur Parametrisierung und Evaluierung Daten, die einen beziehungsweise zwei Monate älter sind als die hinterlegten Referenzen. In vorherigen Arbeiten stammten die Daten immer auf einer einzelnen Aufnahme-

sitzung. Aus diesem Grund beleuchten den Einfluss der kurzfristigen Alterung sowohl auf die Authentifikation, als auch auf die Generierung biometrischer Hashes. Das Test-Setup und die Methodologie zur Evaluierung der Auswirkungen der Alterung beschreiben wir in 6.2.5 *Setup zu B Alterung biometrischer Daten*. Die experimentellen Ergebnisse und die daraus resultierenden Schlussfolgerungen stellen wir in Abschnitt 7.5 *Evaluierung: B Alterung biometrischer Daten* vor.

3.3.3 Performanzmaße für die biometrische Hash-Generierung

Der Erfolg der untersuchten Optimierungsansätze wird für Verifikation und Hash-Generierung mittels unterschiedlicher biometrischer Performanzmaße evaluiert. Für die Bestimmung der Erkennungsgenauigkeit der Verifikation benutzen wir die in der Biometrie weit verbreitete Equal Error Rate (EER). Diese wird aus der False Match Rate (FMR) und der False Non Match Rate (FNMR) bestimmt. Die EER ist dabei nicht das Optimum für ein biometrisches System, sondern stellt einen normalisierten Vergleichswert dar. Für die Bewertung, wie erfolgreich die Hash-Generierung ist, haben wir die Maße Reproduction Rate (RR) und Collision Rate (CR) entwickelt und in unseren Untersuchungen eingesetzt. Die Collision Reproduction Rate (CRR) ist der Trade Off beider Werte, welcher als Vergleichswert herangezogen wird. Detailliertere Informationen zu den Raten sind in den Abschnitten 2.1.6 *Biometrische Fehlerraten* beziehungsweise 2.1.7 *Reproduktion und Kollision im Hash-Generierungsmodus* zu finden.

4 Verwendete Algorithmen

Als Referenzalgorithmen werden wir in dieser Arbeit den Biometric Hash-Algorithmus [Vie06] und den Secure Sketch-Algorithmus [SVD09a], [MSV12] für die dynamische Handschrift verwenden. In diesem Kapitel beschreiben wir beide Algorithmen im Kontext der Anwendungsszenarien Verifikation und Hash-Generierung.

Unsere Untersuchungen basieren auf dem Secure Sketch als Vorstufe des Fuzzy Extractors, da für die Umsetzung eines Fuzzy Extractors ein beliebiges kryptographisches Verfahren genutzt werden kann, welches auf den vom Secure Sketch-Algorithmus erzeugten Werten aufsetzt. Nach dem Einsatz einer solchen kryptographischen Hash-Generierung kann im Allgemeinen keine Aussage getroffen werden, welche Auswirkung einzelne Merkmale auf den erzeugten Hash-Wert haben. Diese Information ist für die einzelnen Optimierungsansätze von wesentlicher Bedeutung. Aus diesem Grund wird die Untersuchung der Reproduzierbarkeit der biometrischen Hashes als auch Verifikationsperformanz anhand des basierend auf dem Secure Sketch rekonstruierten Merkmalsvektors vorgenommen. Dabei gehen wir davon aus, dass ein kryptographisches Verfahren aus durch den Secure Sketch-Algorithmus identisch rekonstruierten Merkmalsvektoren auch identische Werte generieren wird.

4.1 Biometric Hash-Algorithmus für die dynamische Handschrift

Der Biometric Hash-Algorithmus wurde zur Generierung von stabilen individuellen Hash-Vektoren basierend auf der biometrischen Modalität der dynamischen Handschrift entwickelt ([VSM02], [Vie06]). Die zur Verfügung stehende Implementierung ermöglicht zum einen die Erzeugung von biometrischen Hash-Werten und zum anderen die biometrische Benutzerauthentifikation basierend auf den erzeugten Hashes.

Die Generierung eines biometrischen Hashes mithilfe des BioHash-Verfahrens erfolgt in den zwei Stufen Enrollment und Hash-Generierung. Diese sind in der Abbildung 4.1 dargestellt und werden im Folgenden beschrieben. Eine formale Beschreibung der beiden Prozesse nehmen wir daran anschließend vor.

Wie bei jedem biometrischen Verfahren dient der Enrollment-Prozess der Registrierung der Referenzdaten eines Nutzers j im biometrischen System. Dazu werden n Handschriftendaten dieser Person erfasst. Die einzelnen Samples bestehen, wie in Abschnitt 2.1.2 *Handschrift als biometrische Modalität* beschrieben, jeweils aus einer Sequenz von physikalischen Daten, die zu diskreten Zeitpunkten von einem geeigneten Sensor zur Erfassung der Handschrift (zum Beispiel Tablet PC, Signatur-Tablet) aufgenommen werden. Der Sensor ist im Allgemeinen in

der Lage, zeitabhängig die Stiftposition ($x(t)$, $y(t)$), den Aufsetzdruck ($p(t)$) und Höhen- ($\phi(t)$) und Seitenwinkel ($\theta(t)$) des Stiftes zu erfassen.

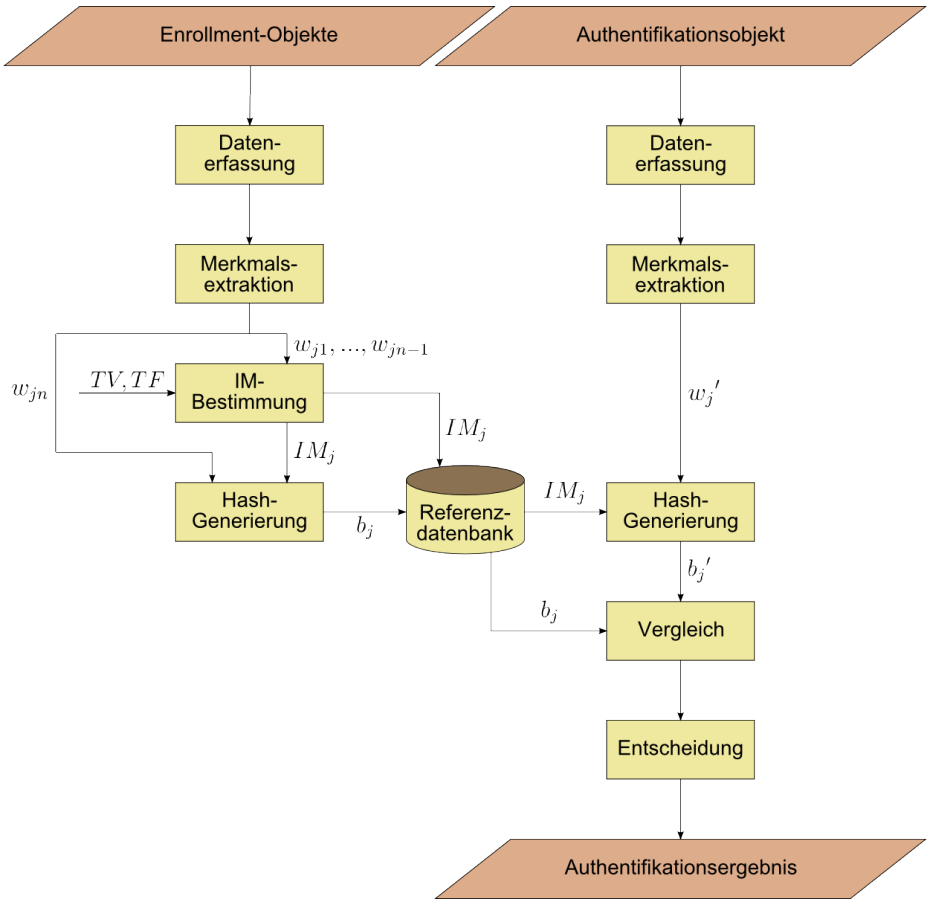


Abbildung 4.1: Schematische Darstellung des Biometric Hash-Algoritmus

Die Rohdaten der n während des Enrollment aufgenommen Samples werden genutzt, um jeweils k statistische Merkmale zu extrahieren, welche in den Merkmalsvektoren $w_{j1}, \ldots, w_{jn}$ gespeichert werden. Mit Ausnahme des Vektors w_{jn} werden aus diesen so genannte Hilfsdaten (engl. Helper Data) in Form der Intervallmatrix IM_j erzeugt, bei der personenspezifische Intra-Klassen-Schwankungen berücksichtigt werden. Der Einfluss der Schwankungen auf die Intervallmatrix kann durch spezielle Werte parametrisiert werden. Der Tolerance Factor TF beeinflusst einerseits als skalarer Wert alle Merkmale global, während auf der anderen Seite durch die einzelnen Elemente des Tolerance Vector TV jedes Merkmal gezielt individuell beeinflusst werden kann. Dafür wird der TV entweder einzeln für jeden Nutzer (*lokaler TV*) oder für eine Gruppe von Nutzern (*globaler TV*) bestimmt. Eine solche Nutzergruppe kann sich beispielsweise aus allen im System registrierten Personen oder auch aus einer (Trainings-)Menge von nicht registrierten Probanden zusammensetzen. Basierend auf der Intervallmatrix IM_j und dem verbleibenden Merkmalsvektor w_{jn} wird im Anschluss der für eine Authentifikation notwendige Referenz-Hash-Vektor b_j bestimmt. Dieser entsteht durch Abbildung der einzelnen statistischen Merkmale auf die korrespondierenden Hash-Werte mithilfe von IM_j. Verknüpft mit der Identität des Schreibers j werden Hash-Vektor b_j und Intervallmatrix IM_j als Referenz in der Datenbank gespeichert.

Für die spätere Generierung eines Hashes wird der aktuell präsentierte Schriftzug erfasst und aus den Rohdaten wird der Vektor $w_j{}'$ mit den statistischen Merkmalen berechnet. Unter Verwendung der individuellen Intervallmatrix IM_j und der Mapping-Funktion wird ein biometrischer Hash-Vektor $b_j{}'$ generiert. Beim Vergleich kann der Grad der Übereinstimmung beider Hashes bestimmt werden, beispielsweise mittels Distanzfunktion. Eine erfolgreiche Authentifikation verlangt lediglich einen Wert innerhalb eines vorgegebenen Toleranzbereiches. Im Hash-Generierungsmodus wird hingegen eine komplette Übereinstimmung beider Hashes verlangt.

Zur formalen Beschreibung beider Algorithmen verwenden wir die folgende Notation: Anzahl der Nutzer M wobei der Index j auf einen bestimmten Nutzer verweist, Anzahl der Merkmale N mit dem Index i als Bezug auf ein einzelnes Merkmal und f_{ij} ist der Wert eines Merkmals. Da die Berechnungsschritte für alle Merkmale identisch sind, werden im folgenden die einzelnen Gleichungen für einen festen Index i betrachtet.

Im ersten Schritt wird für das jeweilige Merkmal i die Breite des entsprechenden Intervalls bestimmt:

$$\Delta I_{ij} = I_{High\ ij} - I_{Low\ ij} \tag{4.1.1}$$

Die zur Bestimmung von $I_{High\ ij}$ und $I_{Low\ ij}$ notwendigen initialen Werte $I_{InitHigh\ ij}$ und $I_{InitLow\ ij}$ werden als Maximum beziehungsweise Minimum über die Anzahl S_j der Trainigs-Samples des jeweiligen Nutzers j berechnet:

$$I_{InitHigh\ ij} = \max_{s=1,S_j} (f_{ijs}) \tag{4.1.2}$$

$$I_{InitLow\ ij} = \min_{s=1,S_j} (f_{ijs}) \tag{4.1.3}$$

Aus der Differenz dieser beiden Werte wird $\Delta I_{Init\ ij}$ als weiterer Input für die Berechnung von $I_{High\ ij}$ (siehe Gleichung 4.1.5) und $I_{Low\ ij}$ (siehe Gleichung 4.1.6) bestimmt:

$$\Delta I_{Init\ ij} = I_{InitHigh\ ij} - I_{InitLow\ ij} \tag{4.1.4}$$

Dazu werden $I_{High\ ij}$ und $I_{Low\ ij}$ als obere beziehungsweise untere Intervallgrenzen wie folgt berechnet.

$$I_{High\ ij} = \lceil I_{InitHigh\ ij} + tv_{ij} * \Delta I_{Init\ ij} * tf \rceil \tag{4.1.5}$$

$$I_{Low} = \begin{cases} \lfloor I_{InitLow\ ij} - tv_{ij} * \Delta I_{Init} \ * tf \rfloor \ wenn \ \lfloor I_{InitLow\ ij} - tv_{ij} * \Delta I_{Init\ ij} * tf \rfloor > 0 \\ 0 \ sonst \end{cases} \tag{4.1.6}$$

Dabei sind tv_{ij} und tf Parameter, die die Breite des Intervals beeinflussen können. Der Tolerance Vector tv wird für jedes Merkmal einzeln bestimmt und genutzt während der Tolerance Factor tf ein einzelner skalarer Wert ist, der benutzer- und merkmalsübergreifend verwendet wird. Beide Parameter können anhand von Trainingsdaten ermittelt werden, wobei die Bestimmung des tv auf der Intra-Klassen-Varianz der Nutzer basiert (siehe dazu auch [Vie06]). Diese kann beispielsweise auf Grundlage von Daten der registrierten Nutzer oder einer zu diesen disjunkten Gruppe von Personen basieren. Ähnlich kann auch bei der Ermittlung des geeigneten tf vorgegangen werden, die mithilfe von Performanzmaßen für den Hash-Generierungs- beziehungsweise den Verifikationsmodus durchgeführt werden kann (siehe auch Abschnitt 6.2.2 *Setup zu A.2 Optimierung von Parametern*).

Im Anschluss wird basierend auf der unteren Intervallgrenze $I_{Low\ ij}$ und der Intervallbreite ΔI_{ij} der Offset Ω_{ij} berechnet:

$$\Omega_{ij} = I_{Low\ ij} \ mod \ \Delta I_{ij} \tag{4.1.7}$$

Die Intervallmatrix IM_j für eine Person j setzt sich zusammen aus dem Vektor der jeweiligen Intervallbreiten ΔI_{ij} aller Merkmale und dem Vektor der zugehörigen Offsets Ω_{ij}.

$$IM_j = (\Delta I_j, \Omega_j) \tag{4.1.8}$$

Die Intervallmatrix IM_j wird als Referenz für die Person j hinterlegt, beispielsweise in einer Datenbank oder auf einer SmartCard. Die in ihr enthaltenen Hilfsdaten werden zur Bestimmung des Biometric Hash verwendet. Dazu erfolgt die Abbildung jedes zuvor berechneten Merkmals f_{ij} in den Hash-Raum mittels der entsprechenden Intervallbreite ΔI_{ij} und dem Off Set Ω_{ij} aus der Intervallmatrix IM_j:

$$b_{ij} = \left\lfloor \frac{(f_{ij} - \Omega_{ij})}{\Delta I_{ij}} \right\rfloor \tag{4.1.9}$$

Für die Verwendung des Biometric Hash-Algorithmus zur Authentifikation wird basierend auf den statistischen Merkmalen eines während des Enrollment-Prozesses aufgenommenen Samples ein Biometric Hash ermittelt, der als Referenz zusätzlich zur Intervallmatrix IM gespeichert wird. Dieser wird mit dem Biometric Hash der aktuell präsentierten Schriftprobe unter Verwendung eines Schwellwertes verglichen.

4.2 Secure Sketch-Algorithmus für die dynamische Handschrift

Das theoretische Konzept des Fuzzy Extractors basierend auf dem Secure Sketch wurde von Dodis et al. in [DRS04] vorgestellt. Die allgemeine Grundidee des Verfahrens des Secure Sketch ist die Erzeugung von Referenzdaten (Secure Sketch) aus einer Menge von Eingangsdaten, mit deren Hilfe aus genügend ähnlichen Informationen eben dieser Input wieder rekonstruiert werden kann.

Für biometrische Daten kann diese Vorgehensweise folgendermaßen beschrieben werden: Aus dem Merkmalsvektor w_j der Referenzdaten eines Nutzers j wird öffentliches Wissen in Form des Secure Sketch s_j extrahiert (siehe Abbildung 4.2). Dieser Vorgang wird als Sketch-Generierung bezeichnet. Basierend auf einem weiteren Merkmalsvektor $w_j{}'$ und dem Secure Sketch s_j soll während des Reproduktionsprozesses der ursprüngliche Merkmalsvektor w_j wieder hergestellt werden. Ein wichtiger Sicherheitsfaktor dabei ist, dass allein aus dem Sketch s_j keine Informationen zur Person oder zu deren biometrischen Daten ableitbar sein dürfen.

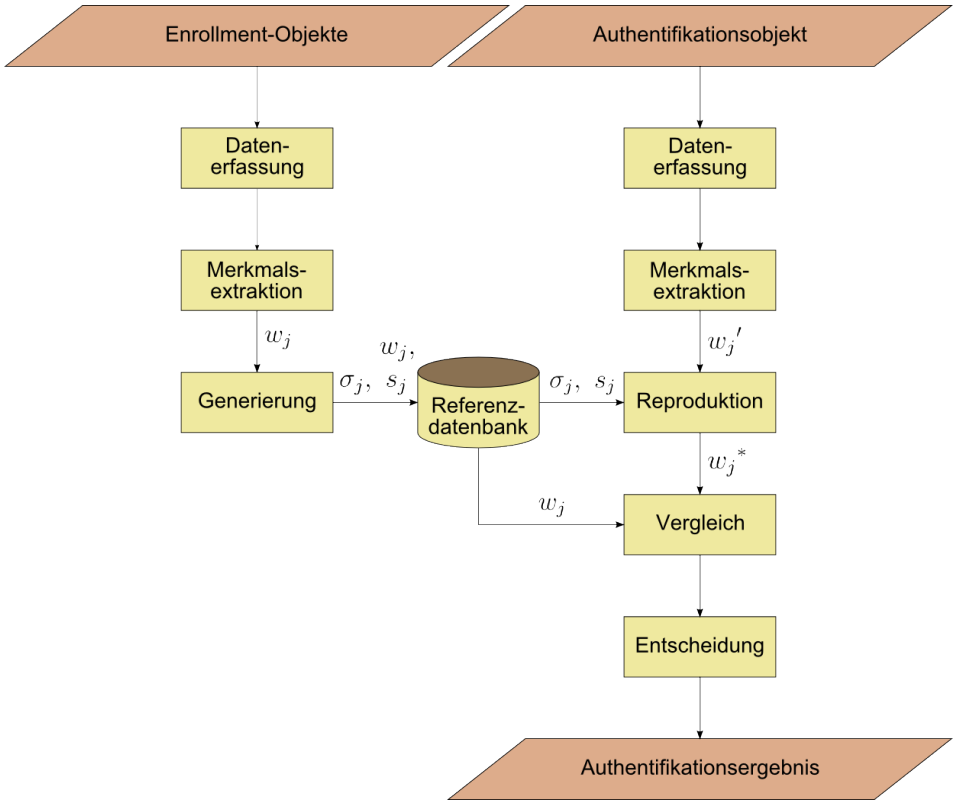


Abbildung 4.2: Schematische Darstellung der Funktionsweise des Secure Sketch-Algorithmus

Das im Folgenden vorgestellte Verfahren wurde von uns basierend auf dem von Sutcu et al. in [SLM07] vorgestellten System zur biometrischen Verifikation anhand von Gesichtern entwickelt [SVD09a], [MSV12]. Dabei wurde darauf geachtet, dass der Algorithmus neben der Verifikation auch zur biometrischen Hash-Generierung eingesetzt werden kann. Basierend auf dem vom Verfahren bestimmten Secure Sketch kann mittels kryptographischer Hash-Funktion ein Hash generiert werden. Ist der Algorithmus in der Lage, für jede Person einen stabilen individuellen

Secure Sketch zu berechnen, kann daraus folglich auch ein eindeutiger kryptographischer Hash erzeugt werden. Allerdings ist es danach nicht mehr realisierbar, zu bestimmen, welche Hash-Komponenten durch einzelne Veränderungen, wie beispielsweise Optimierungen, angepasst wurden. Also ist eine direkte Gegenüberstellung der entsprechenden Hashes nicht möglich. Um einen elementweisen Vergleich zwischen einzelnen Optimierungsschritten zu ermöglichen, werden wir im Rahmen dieser Arbeit nicht den aus dem Secure Sketch mittels kryptographischem Verfahren generierten Hash betrachten, sondern den Secure Sketch selbst. Zur Vereinfachung werden wir für die Ausgabe beider Algorithmen weiterhin synonym den Term *Hash* verwenden.

Um die Varianz der Merkmalswerte zu verringern, wird eine Quantisierung mit einem festen globalen Quantisierungsschritt δ_i durchgeführt, dabei ist S_j die Anzahl der Trainings-Samples des Nutzers j. Der Wert var_{ij} wird zur Bestimmung des individuellen Quantisierungsschrittes verwendet.

$$var_{ij} = \max_{s=1,S_j}(f_{ijs}) - \min_{s=1,S_j}(f_{ijs}) \tag{4.2.1}$$

$$\delta_i = max\left(\min_{j=1,M}(var_{ij}), 1\right) \tag{4.2.2}$$

Als das jeweilige Ausgangsdatum wird für jeden Nutzer j der Durchschnitt des kleinsten und des höchsten Wertes des aktuellen Merkmals berechnet:

$$avg_{ij} = 0.5 \cdot \left(\max_{s=1,S_j}(f_{ijs}) + \min_{s=1,S_j}(f_{ijs})\right) \tag{4.2.3}$$

Im nächsten Schritt werden diese Durschnittswerte mittels des globalen Quantisierungsschrittes δ_i quantisiert, dabei ist ef der Expansion Factor des individuellen Quantisierungsintervalls σ_{ij}.

$$\sigma_{ij} = \left\lceil \frac{ef \cdot var_{ij}}{\delta_i} \right\rceil \tag{4.2.4}$$

$$w_{ij} = \left\lfloor \frac{avg_{ij}}{\delta_i} \right\rfloor \tag{4.2.5}$$

Durch die Veränderung des Expansion Factor ef kann das Verhalten des Algorithmus sowohl im Verifikations- als auch im Hash-Generierungsmodus beeinflusst werden. Aus diesem Grund nutzen wir den Expansion Factor zur Untersuchung der parameterbasierten Optimierung des Secure Sketch-Algorithmus.

Die Referenz eines Nutzers ist der Vektor w_j. Das skalare Codeword c_{ij} wird berechnet aus der

Quantisierung von w_{ij} basierend auf σ_{ij}:

$$c_{ij} = round\left(\frac{w_{ij}}{2\sigma_{ij}+1}\right) \cdot (2\sigma_{ij}+1) \tag{4.2.6}$$

Der Sketch s als Hilfsdatum wird berechnet als die Differenz aus quantisiertem Durchschnittswert und dem dazu gehörendem Codeword:

$$s_{ij} = w_{ij} - c_{ij} \tag{4.2.7}$$

Da der Sketch s_{ij} verfügbar ist, kann das Codeword c_{ij} aus einem beliebigen Testvektor t_j von Person j ermittelt werden, wenn t_j ausreichend ähnlich zum Vektor der Durchschnittswerte avg_j ist. Für jedes Merkmal i des Vektors t_j wird die Quantisierung mittels Quantisierungsschritt δ_i durchgeführt:

$$w'_{ij} = \left\lfloor \frac{t_{ij}}{\delta_{ij}} \right\rfloor \tag{4.2.8}$$

Anschließend wird der Sketch subtrahiert und die nutzerabhängige Quantisierung durchgeführt:

$$c'_{ij} = round\left(\frac{w'_{ij} - s_{ij}}{2\sigma_{ij}+1}\right) \cdot (2\sigma_{ij}+1) \tag{4.2.9}$$

$$w_{ij} = c'_{ij} + s_{ij} \tag{4.2.10}$$

Basierend auf der Annahme, dass die Vektoren w_j und w'_j zueinander in ausreichendem Maße ähnlich sind, müssen die durch die Transformation bestimmten Vektoren c_j und c'_j identisch sein. In diesem Fall kann die Referenz w_{ij} entsprechend Gleichung 4.2.10 als Summe des rekonstruierten Codewords und des Sketches bestimmt werden.

4.3 Statistische Merkmale für beide Referenzalgorithmen

Die verwendeten Merkmale sind für beide Algorithmen identisch. Vielhauer schlägt in [Vie06] 69 statistische Merkmale für die Handschrift vor. Diese geben zu einem Teil statische Eigenschaften der Schrift wieder, wie beispielsweise das Seitenverhältnis. Andere Merkmale, wie zum Beispiel die Geschwindigkeit, basieren auf der Dynamik der Schrift. Die von Vielhauer entwickelten Merkmale werden detailliert in [Vie06] beschrieben. Im Verlauf der Erstellung dieser Arbeit wurden 62 weitere Merkmale entworfen, umgesetzt und der Merkmalsmenge hinzugefügt. Diese sind Bestandteil der Optimierung und werden in Abschnitt 6.2.1 näher beschrieben. Eine tabellarische Übersicht der

insgesamt 131 Merkmale, die den Evaluierungen der beiden Algorithmen zugrunde liegen, wird im Anhang A in den Tabellen A.1 und A.2 gegeben.

5 Ansätze zur Optimierung

Die Optimierung eines biometrischen Systems kann verschiedene Ziele verfolgen. Neben der Verkürzung der Laufzeit liegt das Interesse vor allem in der Verbesserung der Klassifizierungsperformanz des Systems. Vom Blickpunkt des hier vorgestellten biometrischen Hash-Algorithmus bedeutet das einerseits die Maximierung der Reproduzierbarkeit von individuellen Hashes bei gleichzeitig hoher Kollisionsresistenz. Auf der anderen Seite geht damit auch eine Verbesserung der Performanz der Authentifikation einher. Ziel hier ist es, die Erkennung registrierter Personen zu verbessern, während nicht autorisierte Personen zuverlässig abgelehnt werden. Aufgrund der Intra-Klassen-Varianz und der Inter-Klassen-Ähnlichkeit biometrischer Daten ist dies eine Herausforderung, die in dieser Arbeit mithilfe verschiedener Optimierungsansätze angenommen wird. Dieses Kapitel stellt unsere Ansätze zur Optimierung der Performanz von Verifikations- und Hash-Generierungsmodus im Kontext der Aufgabenstellungen *A.1 - A.4* vor. Diese werden evaluiert entsprechend Setup und Methodologie, welche in Kapitel 6 *Aufbau und Methodologie der experimentellen Evaluierung* beschrieben werden. Die Präsentation und Diskussion der Ergebnisse folgt in Kapitel 7 *Präsentation und Diskussion der experimentellen Ergebnisse*.

5.1 Hinzufügen zusätzlicher Merkmale (A.1)

Da die in dieser Arbeit untersuchten Algorithmen auf der Verwendung von statistischen Merkmalen beruhen, liegt hier ein Ansatzpunkt unserer Untersuchungen. Zum einen kann ein biometrischer Algorithmus durch das Hinzufügen von weiteren statistischen Merkmalen beeinflusst werden. Aber auch das Entfernen bestimmter Merkmale kann andererseits zu einer Verbesserung der Performanz im Verifikations- und Hash-Generierungsmodus führen.

Im Rahmen dieser Arbeit haben wir zusätzliche 62 statistische Merkmale konzipiert und umgesetzt. Die Auswirkungen der neuen Merkmale auf die Performanzmaße FNMR/FMR/EER beziehungsweise RR/CR/CRR haben wir untersucht. Die Ergebnisse werden in Kapitel 7 *Präsentation und Diskussion der experimentellen Ergebnisse* vorgestellt und diskutiert.

In diesem Abschnitt gehen wir auf die einzelnen statistischen Merkmale ein, die wir für die zwei in dieser Arbeit fokussierten Algorithmen verwendet haben. Die ersten 69 Merkmale basieren dabei auf Vorarbeiten (siehe [Vie06]), die verbleibenden 62 Merkmale haben wir im Rahmen dieser Arbeit entwickelt und umgesetzt. Sie entstanden im Zusammenhang mit der Betreuung einer Diplomarbeit (Merkmale $n_{70} - n_{89}$, [Guz07]) und von zwei studentischen Projekten (Merkmale $n_{90} - n_{103}$, [FHP08] und Merkmale $n_{104} - n_{131}$, [HHH09]) an der Otto-von-Guericke-Universität Magdeburg. Beide Algorithmen, Biometric Hash und Secure Sketch, basieren auf derselben Menge

an Merkmalen, welche aus der dynamischen Handschrift extrahiert werden. Die statistischen Merkmale selbst können dabei entweder als dynamisch oder statisch angesehen werden. Sie werden als *dynamisch* klassifiziert, wenn sie basierend auf einer zeitlichen Abhängigkeit bestimmt werden. Werden sie hingegen mittels zeitunabhängiger Eigenschaften ermittelt, werden sie der Klasse *statisch* zugeordnet. Im folgenden werden die Merkmale aufgelistet und beschrieben. Dabei erfolgt die Beschreibung neben der verbalen Darstellung zusätzlich entweder formal, grafisch oder durch Angabe von Pseudocode. Der verwendete Pseudocode entspricht dabei den Vorgaben von Cormen et al. in [CLRS09]. Beim Design der neuen Merkmale haben wir entsprechend der Vorgaben von Vielhauer in [Vie06] darauf geachtet, dass deren Werte nicht-negative Ganzzahlen sind. Daher wird, angelehnt an die Bestimmung der Merkmale n_1 - n_{69}, bei nicht-ganzzahligen Zwischenergebnissen vor dem Abrunden mit 100 beziehungsweise 1000 multipliziert. Dieses Vorgehen sichert das teilweise Beibehalten von Informationen der Nachkommastellen. Der in Abbildung 5.1 dargestellte beispielhafte Schriftzug bildet im Folgenden die Grundlage für die Veranschaulichung eines Teils der Merkmale.

Abbildung 5.1: beispielhafter Schriftzug zur Veranschaulichung der Merkmale

Eine Liste aller verwendeten Merkmale ist mit Bezeichnung und kurzer Beschreibung im Anhang A *Statistische Merkmale* und bei Makrushin et al. in [MSV12] zu finden.

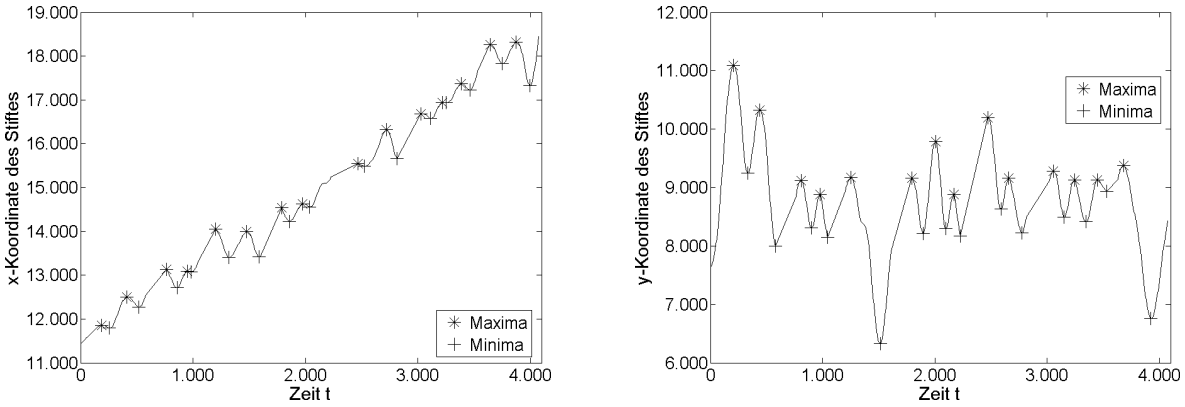
Merkmale n_1 - n_{69}

Die Merkmale n_1-n_{69} stammen aus Vorarbeiten von Vielhauer et al., die sich mit der Entwicklung, Evaluierung und Untersuchung des biometrischen Hashes befassen. Aus diesem Grund verweisen wir an dieser Stelle auf die entsprechende Literatur [Vie06], [VSM02].

Merkmale n_{70} - n_{75}

In Abhängigkeit von den individuellen Schreibweisen unterschiedlicher Personen können gleiche oder ähnliche Schreibinhalte differierende Aussehen aufweisen. Das kann beispielsweise zum einen an der Art liegen, wie Buchstaben begonnen oder beendet werden oder, wie sie miteinander verbunden sind. Diesen Aspekt spiegeln zum Teil Extremwerte im x- beziehungsweise y-Signal wider. Die Merkmale n_{70} bis n_{75} werden mithilfe der Koordinaten x(t) und y(t) bestimmt, indem jeweils die lokalen Maxima beziehungsweise Minima von x(t) und y(t) im Schreibsignal ermittelt

werden. Die Merkmale n_{70} und n_{72} geben dabei die Anzahl der lokalen Maximumwerte für x(t) beziehungsweise y(t) innerhalb des gesamten Schriftzuges an, und die Anzahl der lokalen Minima für x(t) und y(t) wird durch die Merkmale n_{71} beziehungsweise n_{73} repräsentiert. Die Abbildungen 5.2a und 5.2b zeigen die jeweiligen Extrempunkte für x(t) und y(t). Das Listing 5.1 stellt vereinfacht die Bestimmung der Anzahl der jeweiligen Kategorien der Extrempunkte dar. Dabei ergeben sich die einzelnen Merkmale nach dem Durchlauf der Schleife wie folgt aus den Zählvariablen: $n_{70} \hat{=}$ maxXPoints, $n_{71} \hat{=}$ minXPoints, $n_{72} \hat{=}$ maxYPoints und $n_{73} \hat{=}$ minYPoints.



(a) (b)

Abbildung 5.2: Merkmale n_{70} bis n_{73}: Beispielhafte Darstellung der Maximum- und Minimum-Punkte für (a) x(t)-Signal (Anzahl Maxima: 15, Anzahl Minima: 15) beziehungsweise (b) y(t)-Signal (Anzahl Maxima: 14, Anzahl Minima: 14)

Listing 5.1: allgemeine Beschreibung der Ermittlung der Anzahl der lokalen Extrempunkte der x(t)- und y(t)-Signale für die Merkmale n_{70} - n_{75}

```
maxXPoints = minXPoints = maxYPoints = minYPoints = 0
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
for i=1 to s
   // für jeden Punkt, Test auf Minimum/Maximum im X-Signal
   if isLocalMaximumX(Points[i])
      // erhöhe Zähler
      maxXPoints++
   else if isLocalMinimumX(Points[i])
      // erhöhe Zähler
      minXPoints++
   // für jeden Punkt, Test auf Minimum/Maximum im Y-Signal
   if isLocalMaximumY(Points[i])
      // erhöhe Zähler
      maxYPoints++
   else if isLocalMinimumY(Points[i])
      // erhöhe Zähler
      minYPoints++
```

Das Verhältnis zwischen den lokalen Minima beziehungsweise Maxima der x(t)- und y(t)-Signale wird durch die Merkmale n_{74} und n_{75} beschrieben. Ihre Berechnung wird wie folgt durchgeführt:

$$n_{74} = \left\lfloor \frac{n_{70}}{n_{72}} * 1000 \right\rfloor \tag{5.1.1}$$

$$n_{75} = \left\lfloor \frac{n_{71}}{n_{73}} * 1000 \right\rfloor \tag{5.1.2}$$

Merkmale n_{76} - n_{85}

Ob und wie oft sich die Spur einer Schriftprobe selbst schneidet, hängt neben dem geschriebenen Inhalt ebenfalls vom individuellen Stil der Person ab. So kann zum Beispiel das kleine L ohne (l) oder mit (ℓ) Schnittpunkt geschrieben werden. Auch wie breit oder hoch geschrieben wird, hängt neben inhaltlichen Aspekten von der individuelle Schriftcharakteristik einer Person ab. Die räumliche Ausdehnung beeinflusst auch die Schnittpunkte und deren Positionen zueinander.

Die Merkmale n_{76} - n_{85} nutzen die vorgenannten Eigenschaften und basieren auf Untersuchungen von verschiedenartigen Schnittpunkten des Schriftzuges. Die Anzahl der Schnittpunkte innerhalb der Schreibspur, wie sie beispielsweise durch Schleifen entstehen, werden für das Merkmal n_{76} ermittelt (siehe Abbildung 5.3a).

Die Merkmale n_{77} bis n_{83} werden mithilfe eines Rasters bestimmt, das im von der Bounding Box vorgegebenen Rahmen über den Schriftzug gelegt wird. Das Raster entsteht durch vertikale Teilung der Bounding Box nach dem Muster 1/8, 3/8, 5/8 und 7/8 (Linien X_1, X_2, X_3, X_4 in Abbildung 5.3b) beziehungsweise durch horizontale Teilung der Bounding Box nach dem Muster 1/6, 3/6 und 5/6 (Linien Y_1, Y_2, Y_3 in Abbildung 5.3b). Die Wahl dieser Aufteilung wird durch positive Erfahrungen bei Untersuchungen verschiedener Rasterungen in Vorarbeiten motiviert. Für die Merkmale n_{77} bis n_{83} werden jeweils die Schnittpunkte der Schreibspur mit den Linien des Rasters bestimmt, wie in Abbildung 5.3b gezeigt wird. Die Reihenfolge der Merkmale entspricht dabei der Anzahl der Schnittpunkte mit den Geraden X_1 bis X_4 gefolgt von Y_1 bis Y_3.

Eine andere Struktur des Rasters wird zur Bestimmung der Merkmale n_{84} und n_{85} genutzt. Hier werden die Diagonalen (Linien D_1, D_2 in Abbildung 5.3c) der Bounding Box herangezogen und jeweils deren Schnittpunkte mit dem Schriftzug berechnet. Für die Merkmale n_{84} und n_{85} wird die Anzahl der Schnittpunkte der Schrift mit D_1 beziehungsweise D_2 verwendet.

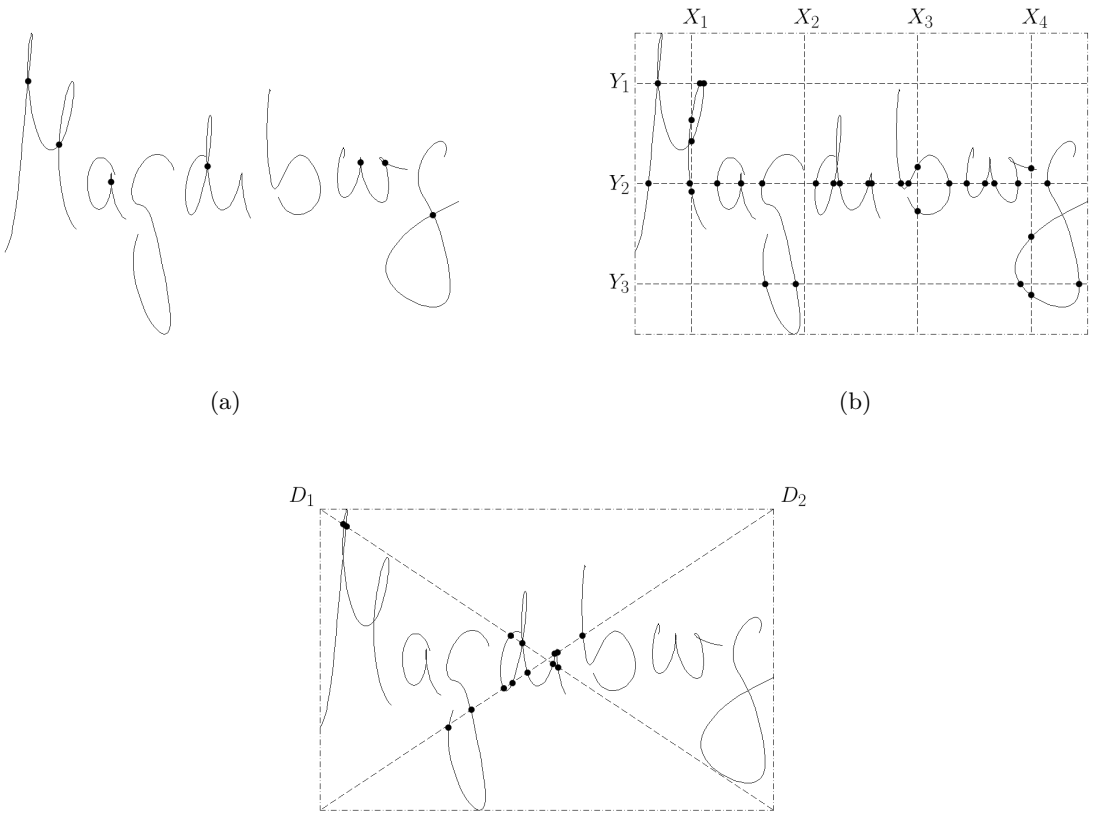


Abbildung 5.3: Schnittpunkte (a) innerhalb des Schriftzuges (Merkmal n_{76}), (b) des Schriftzuges mit dem Raster (Merkmale n_{77} bis n_{83}) und (c) des Schriftzuges mit den Diagonalen der Bounding Box (Merkmale n_{84} und n_{85})

Merkmale n_{86} - n_{89}

Ebenfalls abhängig vom individuellen Schriebstil und vom Schreibinhalt sind die Ausdehnungen der Schrift in horizontaler und vertikaler Richtung. Vor allem die vertikale Ausdehnung der Schrift wird beeinflusst von der Ausnutzung der Ober-, Mittel- und Unterzone [Mic82]. Die Mittelzone entspricht der Höhe der Kleinbuchstaben mit Kurzlänge wie beispielsweise *a, c, e*. Von Oberlänge wird gesprochen, wenn Buchstaben über die Mittel- in die Oberzone reichen, wie das zum Beispiel bei *b, d, h* aber auch bei vielen Großbuchstaben der Fall ist (zum Beispiel *A, B, C*). Analog ragen Buchstaben mit Unterlänge nach unten aus der Mittelzone in die Unterzone (zum Beispiel *g, j, p*). Buchstaben mit so genannter Langlänge nutzen alle drei Zonen aus, Beispiele dafür sind *f, G, ß*. Um der Ausdehnung der Schrift Rechnung zu tragen, basieren die Merkmale n_{86} bis n_{89} auf Abstandsverhältnissen zwischen markanten Punkten innerhalb des Schriftzuges und seiner Länge (Merkmal n_{31}, PathLength) beziehungsweise des geometrischen Schwerpunktes (Centroid) jeweils in Pixel.

Wie in Abbildung 5.4 zu sehen ist, handelt es sich dabei um die räumlichen Koordinaten der

Punkte von Start und Ende der Schreibspur (zeitlich erster beziehungsweise letzter Punkt der Punktesequenz), vom globalen Maximum und Minimum bezüglich x(t) beziehungsweise y(t) sowie vom Centroid. Die einzelnen Merkmale werden entsprechend der Formeln 5.1.4 bis 5.1.7 berechnet, wobei die Funktion $Distanz(P1, P2)$ dem Euklidischen Abstand der jeweils angegebenen Punkte in Pixel entspricht (siehe Gleichung 5.1.3).

$$Distanz(P1, P2) = \sqrt{(x_{P1} - x_{P2}) + (y_{P1} - y_{P2})} \tag{5.1.3}$$

$$n_{86} = \left\lfloor \frac{Distanz(StartPoint, EndPoint)}{n_{31}} * 1000 \right\rfloor \tag{5.1.4}$$

$$n_{87} = \left\lfloor \frac{Distanz(globalMax(x(t)), globalMin(x(t)))}{n_{31}} * 1000 \right\rfloor \tag{5.1.5}$$

$$n_{88} = \left\lfloor \frac{Distanz(globalMax(y(t)), globalMin(y(t)))}{n_{31}} * 1000 \right\rfloor \tag{5.1.6}$$

$$n_{89} = \left\lfloor \frac{Distanz(StartPoint, Centroid)}{Distanz(EndPoint, Centroid)} * 1000 \right\rfloor \tag{5.1.7}$$

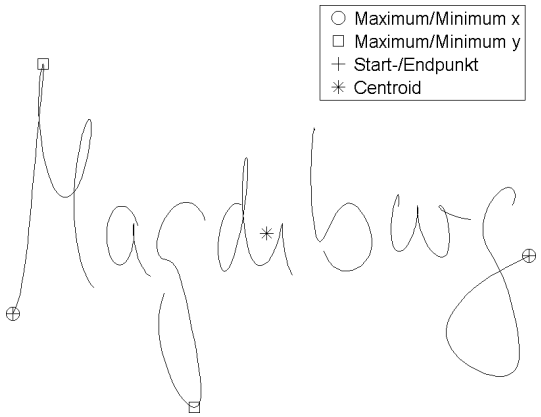


Abbildung 5.4: Start- und Endpunkt, Minimum- und Maximumpunkte bezüglich x(t) beziehungsweise y(t) und Centroid (Merkmale n_{86} bis n_{89})

Merkmale n_{90} - n_{93}

Die Schreibsignale bieten die Möglichkiet der Betrachtung der Veränderung von X-, Y-Position, Druck und Winkeln aber auch daraus abgeleiteter Werte über die Schreibdauer. Diese Signale

verfügen neben globalen auch über lokale Extremwerte, welche gemäß unserer Beobachtungen in initialen Untersuchungen sowohl individuell für den jeweiligen Schreiber als auch für den geschriebenen Inhalt sein können. Die Idee dieser Merkmalsgruppe ist es, die lokalen Extremwerte einzelner Kenngrößen wie Positionen X und Y, Druck, Geschwindigkeit und Beschleunigung auf jeweils einen einzelnen signifikanten Wert abzubilden.

Dazu werden, wie in Listing 5.2 beschrieben, die verwendeten Werte in einem Feld in zeitlicher Reihenfolge aufgelistet. Anschließend wird das Feld kopiert, das Duplikat gespiegelt und elementweise zum Ausgangsfeld addiert. Durch diese Vorgehensweise werden die Elemente entweder am mittleren Index (ungerade Anzahl Elemente) oder in der Mitte (gerade Anzahl Elemente) gespiegelt. Dieser gespiegelte Teil des Feldes wird gelöscht und der Vorgang solange wiederholt, bis nur ein einzelner Wert übrig bleibt. Das Ergebnis wird dann durch die Anzahl der Ausgangswerte dividiert, mit 100 multipliziert und dann ganzahlig abgerundet. Abbildung 5.5 stellt den oben beschriebenen Vorgang beispielhaft an einem fünfstelligen Feld dar.

Listing 5.2: allgemeine Beschreibung der Funktionsweise von fmap

```
function fmap(Array_A)
   // Array_A: Array von Extrempunkten
   s = Array_A.length
   // erzeuge Array_B durch elementweise Addition von
   // Array_A und Array_A in umgekehrter Reihenfolge
   for i=1 to s
      Array_B[i] = Array_A[i] + Array_A[s-(i-1)]
   // bestimme j für gerade bzw. ungerade Anzahl von Elementen
   if (s%2) != 0
      j = s/2
   else
      j = (s+1)/2
   // kopiere Array_B bis j in Array_C
   for i=1 to j
      Array_C[i] = Array_B[i]
   // Rekursion bis nur noch ein Element vorhanden
   if Array_C.length > 1
      fmap(Array_C)
   // Rückgabewert berechnen
   r = floor((Array_C[1] / s) * 100)
   // Rückgabe
   return r
end function
```

Auf diese Weise werden die lokalen Extremwerte einzelner Signale und daraus abgeleiteter Merkmale auf jeweils einen Merkmalswert abgebildet. Dabei handelt es sich um die X-Position (Merkmal n_{90}), die Y-Position (Merkmal n_{91}), den Druck (Merkmal n_{92}) und die Beschleunigung (Merkmal n_{93}). In Abbildung 5.5 ist das Mapping für ein gegebenes Array von Werten beispielhaft dargestellt.

	2	9	6	7	3	4	1	*Start-Array*
+	1	4	3	7	6	9	2	*umkehren*
=	3	13	9	14	9	13	3	*addieren und abschneiden*
+	14	9	13	3				*umkehren*
=	17	22	22	17				*addieren und abschneiden*
+	22	17						*umkehren*
=	39	39						*addieren und abschneiden*

$$fmap(array) = \left\lfloor \frac{39*100}{7} \right\rfloor = 557$$

Abbildung 5.5: Beispielhafte Darstellung der Funktionsweise des Mappings mehrerer Werte auf einen einzelnen (Merkmale n_{90} bis n_{93})

Merkmal n_{94}

Abhängig sowohl vom geschriebenen Inhalt als auch von den individuellen Gewohnheiten des Schreibers, beispielsweise durch Verwendung von Schreib- oder Blockschrift, entstehen Ab- und Aufsetzpunkte. Zwischen diesen berührt das Schreibgerät nicht die Oberfläche. Die Teilschriftzüge von einem Auf- zum nächsten Absetzpunkt bezeichnen wir als Segmente. Diese können sowohl durch einzelne Zahlen, Buchstaben oder Wörter als auch durch individuelle Kombinationen aus mehreren Zeichen oder Teilen dieser bestehen. Diese Eigenschaft wird zur Ermittlung des Merkmals n_{94} genutzt, welches die durchschnittliche Entfernung angibt, während der der Stift innerhalb des Schreibprozesses nicht die Oberfläche berührt hat. Listing 5.3 zeigt den Pseudocode für die Ermittlung des Merkmals.

Listing 5.3: allgemeine Beschreibung der Ermittlung des Merkmals n_{94}

```
count = sumDist = 0
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
for i=1 to s-1
   // für jeden gefundenen Absetzpunkt
   if isPenUpPoint(Points[i])
      // ermittle den Abstand zum nächsten Aufsetzpunkt
      dist = euklidianDistance(Points[i], Points[i+1])
      // addiere aktuellen Abstand auf
      sumDist = sumDist + dist
      // erhöhe Zähler
      count++
// durchschnittliche Entfernung mit abgesetztem Stift
averageDist = sumDist / count
// Ergebnis berechnen
result = floor(averageDist * 1000)
```

Merkmale n_{95} - n_{97}

Bei einer konstanten Erfassungsrate des Schreibsensors häufen sich die durch diesen aufgenommenen Punkte in verschiedenen Bereichen des Schriftzuges unterschiedlich stark. Dies hängt sowohl von der individuellen Schreibgeschwindigkeit als auch vom geschriebenen Inhalt ab. Ebenso spielt es eine Rolle, ob eine Person in der räumlichen Ausdehnung eher weit auseinander schreibt oder eine enge Schreibweise bevorzugt. In Abbildung 5.6 sind die Erfassungspunkte der exemplarischen Schriftprobe dargestellt. Deutlich zu sehen sind die unterschiedlichen Abstände zwischen den einzelnen Punkten, die bei konstanter Erfassungsrate aus unterschiedlichen Schreibgeschwindigkeiten resultieren.

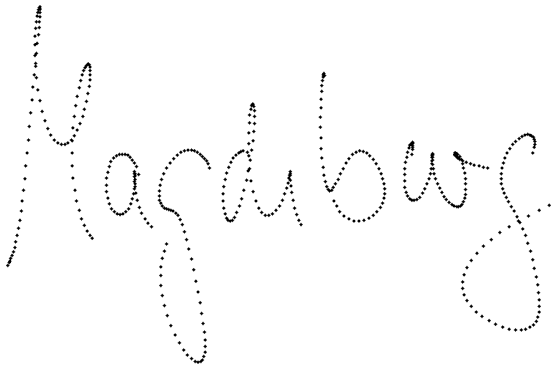


Abbildung 5.6: Darstellung der Aufnahmepunkte eines beispielhaften Schriftzuges (Merkmale n_{95} bis n_{97})

Als Indikatoren der Häufung von Aufnahmepunkten sind die Merkmale n_{95} - n_{97} entstanden. Basierend auf einer kreisförmigen Umgebung mit einem Radius von r wird für jeden einzelnen Punkt die Anzahl der Nachbarpunkte bestimmt. Die Werte für die Merkmale n_{95} und n_{96} entsprechen der Anzahl der wenigsten beziehungsweise der meisten ermittelten Nachbarpunkte. Merkmal n_{97} entspricht der durchschnittlichen Anzahl von Nachbarpunkten innerhalb von r über alle Punkte. Der jeweilige Radius wird für jeden einzelnen Schriftzug dynamisch bestimmt. In Vorarbeiten hat sich gezeigt, dass ein Radius von einem Drittel der Länge der Diagonale der Bounding Box gute Ergebnisse liefert.

Listing 5.4 zeigt die Berechnung der drei Merkmale basierend auf der Bestimmung der Anzahl der Nachbarpunkte innerhalb eines Radius. Dabei ergibt sich nach dem Durchlauf Merkmal n_{95} aus Variable *minPoints*, n_{96} aus *maxPoints* und n_{97} aus *avgPoints*.

Listing 5.4: allgemeine Beschreibung der Ermittlung der Merkmale n_{95} - n_{97}

```
radius = sqrt(BoundingBox.Width^2 + BoundingBox.Height^2) / 3
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
minPoints = s
maxPoints = sumPoints = 0
for i=0 to s
   count = 0
```

```
   for j=0 to s
      if i!=j
         // bestimme euklidische Distanz zwischen aktuellem Punktepaar
         dist = euklidianDistance(Points[i],Points[j])
         // liegt Punkt j im Radius setze Zähler um 1 hoch
         if dist<=radius
            count++
   // aktuelle Anzahl kleiner als minPoints -> ersetzen
   minPoints = min(minPoints, count)
   // aktuelle Anzahl größer als maxPoints -> ersetzen
   maxPoints = max(maxPoints, count)
   // aktuelle Anzahl aufummieren
   sumPoints = sumPoints + count
avgPoints = floor(sumPoints / s)
```

Merkmale n_{98} - n_{103}

Beim Schreiben entstehen innerhalb eines Schriftzuges häufig Schnittpunkte der Schreibspur mit sich selbst (siehe beispielsweise auch Abbildung 5.3a). Betrachten wir diese Punkte innerhalb der aufgenommenen Datensequenz, so stellen sich diese als Kreuzungen der jeweils geraden Verbindungslinien zwischen zwei aufeinander folgenden Punkten zweier Punktepaare dar.

Im Beispiel in Abbildung 5.7 wird der Schnittpunkt basierend auf den beiden Paaren p_i, p_{i+1} und p_j, p_{j+1} erzeugt. Die von beiden Geraden gebildeten Winkel können als Anhaltspunkt für die Individualität der Schreibweise herangezogen werden. Da bei sich schneidenden Geraden die sich gegenüberliegenden Winkel (Scheitelwinkel) gleich groß sind [GKHK69], genügt es, jeweils die nebeneinanderliegenden Winkel (Nebenwinkel, α und β in Abbildung 5.7) zu betrachten. Weiterhin beträgt die Summe der zwei Nebenwinkel immer 180°, daher beschränken wir uns bei der nachfolgenden Betrachtung auf einen der beiden. In unseren Berechnungen verwenden wir den jeweils kleineren Winkel (in Abbildung 5.7 α).

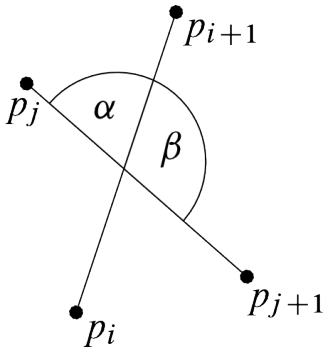


Abbildung 5.7: Beispielhafte Darstellung der Ermittlung eines Schnittpunktes und der zugehörigen Winkel (Merkmale n_{98} bis n_{103})

Die Werte der Merkmale n_{98}, n_{99} und n_{100} geben jeweils den durchschnittlichen Winkel zwischen *0°-30°*, *30°-60°* beziehungsweise *60°-90°* für alle Schnittpunkte an. Die Merkmale n_{101}, n_{102} und

n_{103} geben die prozentuale Anzahl der Winkel in den jeweiligen Bereichen im Vergleich zu allen Schnittpunkten an.

Listing 5.5: allgemeine Beschreibung der Ermittlung der Merkmale n_{98} - n_{103}

```
sum_30 = sum_60 = sum_90 = 0
count_30 = count_60 = count_90 = 0
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
for i=1 to (s-3)
   for j=i+2 to (s-2)
      // isIntersection: Überprüfung, ob sich aktuelle Geraden
      // zwischen zwei Punktepaaren schneiden
      if isIntersection((Points[i], Points[i+1]), (Points[j], Points[j+1]))
         // bestimme Größe der Nebenwinkel
         [a,b] = getAngles((Points[i], Points[i+1]), (Points[j],
            Points[j+1]))
         // ermittle kleineren Winkel
         c = min(a, b)
         // befindet sich Wert im Intervall (0,30], (30,60] bzw. (60,90]
         // summiere Wert auf und erhöhe Zähler um 1
         if (c>0) and (c<=30)
            sum_30 = sum_30 + c
            count_30++
         else if (c>30) and (c<=60)
            sum_60 = sum_60 + c
            count_60++
         else if (c>60) and (c<=90)
            sum_90 = sum_90 + c
            count_90++
// durchschnittliche Winkelgröße pro Kategorie
angle_30 = floor(sum_30 / count_30)
angle_60 = floor(sum_60 / count_60)
angle_90 = floor(sum_90 / count_90)
// Anzahl aller Schnittpunkte
count_all = count_30 + count_60 + count_90
// prozentualer Anteil an allen Schnittpunkten pro Kategorie
percent_30 = floor((count_30 * 100 * 1000) / count_all)
percent_60 = floor((count_60 * 100 * 1000) / count_all)
percent_90 = floor((count_90 * 100 * 1000) / count_all)
```

Das Listing 5.5 verdeutlicht die Bestimmung der sechs Merkmale basierend auf den Winkeln der Schnittpunkte der Schriftspur mit sich selbst. Dabei entsprechen diese wie folgt den einzelnen Variablen: $n_{98} \hat{=}$ angle_30, $n_{99} \hat{=}$ angle_60, $n_{100} \hat{=}$ angle_90, $n_{101} \hat{=}$ percent_30, $n_{102} \hat{=}$ percent_30 und $n_{103} \hat{=}$ percent_90.

Merkmale n_{104} - n_{106}

Ergänzend zu Merkmal n_{94} werden für diese drei Merkmale ebenfalls die Abstände in Pixel zwischen den einzelnen Ab- und Aufsetzpunkten des Stiftes innerhalb eines Samples untersucht. Dazu werden der maximale (n_{104}, Variable *maxDist* in Listing 5.6), der minimale (n_{105}, Variable *minDist*) und der durchschnittliche (n_{106}, Variable *averageDist*) Abstand bestimmt und gespeichert.

Listing 5.6: allgemeine Beschreibung der Ermittlung der Merkmale n_{104} - n_{106}

```
count = sumDist = maxDist = 0
minDist = PathLength
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
for i=1 to s-1
   // für jeden gefundenen Absetzpunkt
   if isPenUpPoint(Points[i])
      // ermittle den Abstand zum nächsten Aufsetzpunkt
      dist = euklidianDistance(Points[i], Points[i+1])
      // addiere aktuellen Abstand auf
      sumDist = sumDist + dist
      // aktuelle Distanz kleiner als minDist -> ersetzen
      minDist = min(minDist, dist)
      // aktuelle Distanz größer als maxDist -> ersetzen
      maxDist = max(maxDist, dist)
      // erhöhe Zähler
      count++
averageDist = sumDist / count
```

Merkmal n_{107}

Auch die Existenz und die Größe von geschlossenen Flächen hängen sowohl von den Schreibgewohnheiten einer Person ab als auch vom Inhalt. Diese vollständig eingeschlossenen Flächen innerhalb eines Schriftzuges werden für dieses Merkmal bestimmt und deren Inhalt berechnet. Die ermittelten Flächeninhalte werden dann mittels fmap-Funktion auf einen einzelnen signifikanten Wert abgebildet (siehe Merkmale n_{90} - n_{93}) und im Merkmal n_{107} gespeichert. Der Toleranzwert, ab dem eine Fläche als geschlossen angesehen wird, kann dabei angepasst werden. In unseren Untersuchungen hat sich dafür ein Wert von 25 Pixels unsere anvisierten Untersuchungen als brauchbar erwiesen. Das bedeutet, entsteht ein geschlossener Bereich durch eine angenommene Verlängerung mit einer maximalen Pixel-Länge von 25, wird diese Fläche bei der Berechnung dieses Merkmals berücksichtigt. Die generelle Vorgehensweise zur Bestimmung des Merkmals n_{107} stellt Listing 5.7 dar.

Listing 5.7: allgemeine Beschreibung der Ermittlung des Merkmals n_{107}

```
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
// Ermittlung der geschlossenen Polygone mit Pixel-Toleranz von 25 und
```

```
// Rückgabe der Punktlisten der einzelnen Polygone in Array polygons
polygons = getAreas(Points, 25)
s = polygons.length
for i=1 to s
   // Berechnung Flächeninhalte und Speicherung in Array areas
   areas[i] = getArea(polygons[i])
// Mapping des Arrays areas auf einzelnen Wert mittels
// fmap-Funktion (siehe Listing 5.2)
result = fmap(areas)
```

Merkmale n_{108} - n_{125}

Die Grundfunktion dieser Merkmale fasst die Bildpunkte des Schriftzuges in sechs Clustern zusammen (siehe Listing 5.8). Dazu werden die Startpunkte folgendermaßen bestimmt: Im ersten Schritt werden alle Punkte entsprechend ihrer Y-Koordinaten sortiert. Dann werden die Punkte geteilt und der größere Teil in einem und der kleinere Teil in einem anderen Array gespeichert. Der dritte Schritt sortiert beide Arrays nach ihren X-Koordinaten. Entsprechend Untersuchungen in Vorarbeiten werden als Startpunkte für die Bildung der Cluster jeweils die Punkte an den Indizes 1/6, 3/6 und 5/6, ausgehend von der Array-Größe, gewählt. Danach wird dann jeweils das Clustering mittels K-Means durchgeführt. Die mit dieser Funktion berechneten Merkmale enthalten für die sechs Cluster die Mittelpunkte in Form der normierten X- und Y-Koordinaten (n_{108} - n_{119}, in Listing 5.8 *startPoints[1].X* bis *startPoints[6].Y*). Zusätzlich wird für jeden Cluster der durchschnittliche Druck der darin enthaltenen Punkte ermittelt und in sechs weiteren Merkmalen (n_{120} - n_{125}, in Listing 5.8 *avgPressureOfCluster[1]* bis *avgPressureOfCluster[6]*) gespeichert.

Listing 5.8: allgemeine Beschreibung der Ermittlung der Merkmale n_{108} - n_{125}

```
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
// sortiere Aufnahmepunkte nach Y-Koordinate
yArray = sortArray(Points, Y)
// teile yArray in zwei Arrays
for i=1 to ceil(s/2)
   yArray1[i] = yArray[i]
j = 1
for i=ceil(s/2)+1 to s
   yArray2[j++] = yArray[i]
// sortiere Arrays nach X-Koordinate
xArray1 = sortArray(yArray1, X)
xArray2 = sortArray(yArray2, X)
// Startpunkte für das Clustering bestimmen
startPoints[1] = xArray1[floor(s/6)]
startPoints[2] = xArray1[floor(s/6)*3]
startPoints[3] = xArray1[floor(s/6)*5]
startPoints[4] = xArray2[floor(s/6)]
```

```
startPoints[5] = xArray2[floor(s/6)*3]
startPoints[6] = xArray2[floor(s/6)*5]
// Clustering basierend auf K–Means und Startpunkten durchführen und
// ermittelte sechs Punkte–Arrays in clusterPoints speichern
clusterPoints = clustering(kmeans, startPoints)
// Durchschnittsdruck der einzelnen Punkte–Arrays bestimmen
s = clusterPoints.length
for i=1 to s
   avgPressureOfCluster[i] = getAvgPressure(clusterPoints[i])
```

Merkmale n_{126} - n_{129}

Analog der Vorgehensweise für die Merkmale n_{108} - n_{125} werden zur Bestimmung der Werte dieser Merkmalsgruppe die Höhenwinkel-Informationen aller Bildpunkte einer Schriftprobe genutzt. Die Startpunkte des Clusterings (K-Means) bilden hier die Punkte an den Indizes 1/8, 3/8, 5/8 und 7/8 des entsprechend der Höhenwinkel sortierten Arrays. Pro ermitteltem Cluster wird der durchschnittliche Höhenwinkel berechnet und in vier Merkmalen (n_{126} - n_{129}, in Listing 7.9 *avgAltitudeOfCluster[1]* bis *avgAltitudeOfCluster[4]*) gespeichert.

Listing 5.9: allgemeine Beschreibung der Ermittlung der Merkmale n_{126} - n_{129}

```
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
// sortiere Aufnahmepunkte nach Druck P
pArray = sortArray(Points, P)
// Startpunkte für das Clustering bestimmen
startPoints[1] = pArray[floor(s/8)]
startPoints[2] = pArray[floor(s/8)*3]
startPoints[3] = pArray[floor(s/8)*5]
startPoints[4] = pArray[floor(s/8)*7]
// Clustering basierend auf K–Means und Startpunkten durchführen und
// ermittelte vier Punkte–Arrays in clusterPoints speichern
clusterPoints = clustering(kmeans, startPoints)
// Durchschnitts–Altitude der einzelnen Punkte–Arrays bestimmen
s = clusterPoints.length
for i=1 to s
   avgAltitudeOfCluster[i] = getAvgAltitude(clusterPoints[i])
```

Merkmal n_{130}

Die Geschwindigkeit an den Wendepunkten einer Schriftprobe werden für die Ermittlung dieses Merkmals herangezogen. Ein Wendepunkt wird durch eine Stelle beschrieben, an der sich die Richtung der Schreibspur bezogen auf die X-Koordinate umkehrt. Basierend auf den Geschwindigkeiten aller Wendepunkte eines Schriftzugs wird per fmap-Funktion ein einzelner eindeutiger Wert berechnet und als Merkmal n_{130} gespeichert (siehe Listing 5.10).

Listing 5.10: allgemeine Beschreibung der Ermittlung des Merkmals n_{130}

```
// Points: Array der Aufnahmepunkte in Reihenfolge ihrer Erfassung
s = Points.length
for i=1 to s-3
   // ermittle Winkel, der durch aktuellen Punkt und seine drei
   // Nachfolger gebildet wird
   angle = getAngle(Points[i], Points[i+1], Points[i+2])
   if angle!=0
      speed = getSpeed(Points[i], Points[i+1], Points[i+2])
      // aktuelle Geschwindigkeit mit 1000 multiplizieren,
      // aufrunden und dem Array inflectionSpeed hinzufügen
      add(inflectionSpeed, ceil(speed * 1000))
// Array inflectionSpeed auf einzelnen Wert abbilden
// (siehe Funktion fmap in Listing 5.2)
result = fmap(inflectionSpeed)
```

Merkmal n_{131}

Der Verlauf des Drucksignals wird neben dem geschriebenen Inhalt auch von den individuellen Gewohnheiten des Schreibers beeinflusst. Dieser unterscheidet sich beispielsweise an Auf- und Absetzpunkten. Um einen Anhaltspunkt für diese Veränderungen zu erhalten, gibt Merkmal n_{131} die Standardabweichung der Druckwerte eines Samples an. Diese wird entsprechend der folgenden Formel 5.1.8 bestimmt:

$$n_{131} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5.1.8}$$

Es ist anzunehmen, dass durch das Hinzufügen neuer Merkmale zwar eine Verbesserung der Verifikation zu erwarten ist, jedoch keine Steigerung der Reproduktion im Hash-Generierungsmodus. Bei letzterem kann beispielsweise beim Biometric Hash-Verfahren keine Verbesserung erfolgen, da neue Merkmale zusätzlich zu den bestehenden auf Hash-Werte abgebildet werden. In dem unwahrscheinlichen aber günstigsten Fall, dass jeder der neuen Hash-Werte immer reproduzierbar ist, wird sich an der aktuellen Reproduktionsrate nichts ändern. Im anderen Fall wird sich die Reproduktionsrate verschlechtern. Unter anderem aus diesem Grund haben wir uns dafür entschieden, neben dem Hinzufügen neuer Merkmale auch eine Analyse aller Merkmale durchzuführen, um diejenigen zu selektieren, die zur Generierung stabiler individueller Hashes beitragen.

5.2 Optimierung von Parametern (A.2)

Eine weitere Möglichkeit, die Verifikations- und Hash-Generierungsperformanz der beiden Referenzalgorithmen zu erhöhen, bieten verschiedene Verfahrensparameter. Dies sind für den Biometric

Hash-Algorithmus der Tolerance Factor TF und der Tolerance Vector TV. Beide Möglichkeiten der Parametrisierung wurden bereits bei der Vorstellung des Verfahrens in Abschnitt 4.1 *Biometric Hash-Algorithmus für die dynamische Handschrift* beschrieben. Während mit dem Tolerance Factor die Intervallbreite für jedes Merkmal durch den selben Wert global beeinflusst wird, steht mit dem Tolerance Vector eine lokale Anpassung zur Verfügung. Er bietet eine elementweise Größenanpassung des Mapping-Intervalls durch einen auf jedes einzelne Merkmal abgestimmten Wert. Für die Evaluierung des Biometric Hash konzentrieren wir uns auf den Tolerance Factor als Wert zur Parametrisierung von Verifikations- und Hash-Generierungsmodus.

Eine ähnliche Möglichkeit wie der Tolerance Factor bietet für den Secure Sketch-Algorithmus der Expansion Factor EF (siehe auch Abschnitt 4.2 *Secure Sketch-Algorithmus für die dynamische Handschrift*). Dieser ermöglicht eine Verkleinerung beziehungsweise Vergrößerung des individuellen Quantisierungsintervalls bei der Bestimmung des Secure Sketch.

Im Rahmen der von uns untersuchten Optimierungsstrategien in Bezug auf die Parameter ist zu beachten, dass der Begriff *optimale Parameter* sich in diesem Zusammenhang nur auf die zur Evaluierung zur Verfügung stehenden Datensätze, das jeweils angestrebte Szenario und die verwendete Methodologie bezieht. Die Verwendung einer anderen Konstellation oder abweichender Zusammensetzung der Datenbank als in Abschnitt 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank* beschrieben, wird aller Voraussicht nach zu anderen optimalen Parametern beziehungsweise Testergebnissen führen.

Die jeweils optimalen Parameter für TF und EF haben wir auf die gleiche Weise ermittelt und werden sie aus diesem Grund hier zusammen beschrieben. Um die für die Testdatenbank und das anvisierte Szenario optimalen Parameter zu finden, nutzen wir zwei unterschiedliche Testsets: Enrollmentdaten (Session 1) und die Daten zur Parameterbestimmung (Session 2, siehe Abschnitt 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank*). Mittels dieser Daten wurden Simulationen der Verifikation und Hash-Generierung durchgeführt, bei denen wir für jeden der beiden Parameter und den korrespondierenden Algorithmus die in Abschnitt 2.1.6 *Biometrische Fehlerraten* eingeführten Maße Equal Error Rate (EER), Reproduction Rate (RR), Collision Rate (CR) und die Collision Reproduction Rate (CRR) bestimmt haben. Die Parameter TF beziehungsweise EF wurden jeweils in Schritten der Größe $0,25$ erhöht. Abbildung 5.8 verdeutlicht unser Vorgehen beispielhaft für die Semantik *Symbol* und den Tolerance Factor. Basierend auf den Ergebnissen für Biometric Hash- und Secure Sketch-Verfahren und fünf Semantiken haben wir verschiedene Parameter-Sets erstellt, die im späteren Verlauf evaluiert werden sollen. Die Erstellung der Parameter-Sets basiert auf bestimmten Eigenschaften, die im Kontext der Optimierung von Verifikation und Hash-Generierung sinnvoll sind. Es werden beispielsweise die Parameterwerte gewählt, für die die geringste Equal Error Rate beziehungsweise Collision Reproduction Rate bestimmt wird. Die Parameter-Sets stellen sich für die beiden Algorithmen wie folgt dar:

Null. Tolerance Factor/Expansion Factor wird pauschal auf den Wert Null gesetzt und entspricht damit der Ausgangssituation ohne gesetzten Quantisierungsparameter. Dieses Szenario

dient während der experimentellen Evaluierung als Referenz zu den Tests mit gesetzten Parameterwerten.

minEER. Hier wird der Tolerance Factor/Expansion Factor genutzt, für den die kleinste Equal Error Rate ermittelt wird.

minCRR. Der Tolerance Factor/Expansion Factor, für den bei der Parameterbestimmung die kleinste Collision Reproduction Rate erzeugt wurde, wird gewählt.

RR $\geq$ a%. Derjenige Tolerance Factor/Expansion Factor wird gewählt, für welchen bei der Parameterermittlung eine Reproduction Rate von mindestens a% berechnet wurde. Die Wahl von a hängt dabei von der jeweiligen Performanz bezüglich Reproduction Rate ab.

CR $\leq$ b%. Tolerance Factor/Expansion Factor wird auf den Wert gesetzt, der bei der Parameterbestimmung eine Collision Rate von maximal b% erreicht hat. Die Wahl von b hängt dabei von der jeweiligen Performanz bezüglich Collision Rate ab.

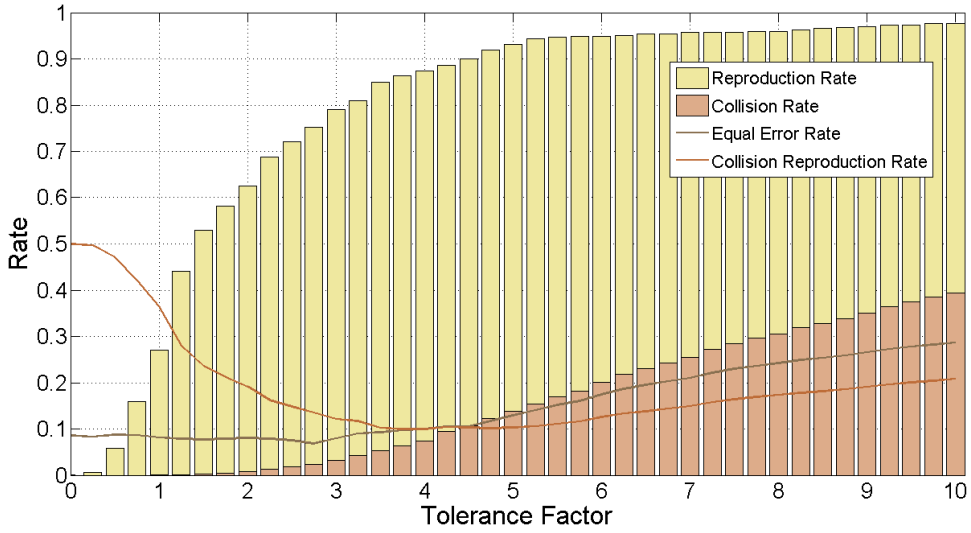


Abbildung 5.8: Beispielhafte Darstellung der Vorgehensweise bei der Bestimmung des optimalen Tolerance Factor für die Semantik Symbol

5.3 Analyse und Selektion von Merkmalen (A.3)

Ein Problem bei einer Vielzahl biometrischer Verfahren ist die große Menge an statistischen Merkmalen, die die Grundlage für die Authentifikation beziehungsweise Hash-Generierung bilden. Diese sind in vielen Fällen intuitiv ausgewählt und nicht zuvor bezüglich ihrer Eigenschaften untersucht worden. Das legt die Vermutung nahe, dass einzelne oder aber auch Kombinationen bestimmter Merkmale für die genannten Anwendungsbereiche ungeeignet sein können. Im ungünstigsten Fall verschlechtert das Hinzufügen eines zusätzlichen Merkmals die Erkennungs- beziehungsweise Hash-Generierungsperformanz des Algorithmus. Diese Auswirkung motiviert

eine entsprechende Entfernung des Merkmals aus der relevanten Menge. Eine andere Möglichkeit besteht darin, dass ein neues Merkmal keinen Einfluss auf das System ausübt. In dieser Situation kann durch das Entfernen des Merkmals zumindest zu einer Ersparnis von Rechenzeit und -ressourcen führen.

Analog zu den optimalen Verfahrensparametern (siehe Abschnitt 5.2) sind im Rahmen unserer Untersuchungen optimale Teilmengen der Menge von ursprünglich 131 Merkmalen solche, die basierend auf den hier vorgestellten Algorithmen, Anwendungsszenarien und Testsets ermittelt und evaluiert wurden. Sollte sich mindestens ein Aspekt dieser Konstellation ändern, müssen die Merkmalsanalysen und Selektionen entsprechend neu durchgeführt werden.

Um jeweils optimale Teilmengen der statischen Merkmale bezüglich der Performanz der biometrischen Algorithmen für Verifikation und Hash-Generierung zu selektieren, haben wir verschiedene Verfahren zur Analyse der Merkmale durchgeführt. Unsere Ansätze bezüglich der Analyse und Selektion von Merkmalen der dynamischen Handschrift zur Verbesserung der Verifikations- als auch der Hash-Generierungsperformanz haben wir für den Biometric Hash-Algorithmus in [MSV11a] beziehungsweise [MSV11b] und für den Secure Sketch in [MSV12] veröffentlicht. Entsprechend der Terminologie von John et al. ([JKP94]) lassen sich die dabei untersuchten Selektionsverfahren in Wrapper und Filter unterteilen (siehe Abschnitt 2.1.5 *Optimierung basierend auf der Merkmalsebene*).

5.3.1 Varianzbasierte Vorselektion

Eine relativ einfache Möglichkeit der Merkmalsselektion ist der Ausschluss von Merkmalen, die einen konstanten Wert aufweisen. Zu diesen Merkmalen können solche zählen, die auf vom Sensor nicht erfassten Informationen beruhen. Werden beispielsweise die Winkel des Stiftes nicht vom Sensor erfasst, weisen die daraus bestimmten Werte einen konstanten Standardwert auf (zum Beispiel 0). Merkmale dieser Art sollten daher generell ausgeschlossen werden, da ihr Informationsgehalt für den jeweilig untersuchten Verwendungszweck Null ist. Entsprechend sind diese Daten auch für die Unterscheidung zwischen unterschiedlichen Personen nicht brauchbar, da sie keinerlei Inter-Klassen-Varianz aufweisen. Einige Verfahren zur statistischen Analyse und Selektion von Daten könnten diese Merkmale als besonders stabil und unabhängig von anderen Merkmalen ansehen und als gute Wahl in das reduzierte Merkmal-Set übernehmen. Aus den genannten Gründen sollten diejenigen Merkmale ausgeschlossen werden, die bei allen registrierten Personen den gleichen Wert aufweisen. Dies lässt sich mittels Analyse der Varianz recht einfach feststellen. Dabei wird die Verteilung der Werte eines Merkmals um den Mittelwert ermittelt. Ist keine Varianz nachweisbar, kann das Merkmal aussortiert werden. Wie oben erwähnt, kann ein Grund für den fehlenden Informationsgehalt eines Merkmals sein, dass der eingesetzte Sensor keine entsprechenden Daten liefert. Aus diesem Grund sollte die Merkmalsselektion für einen Algorithmus erneut durchgeführt werden, wenn sich der eingesetzte Sensor ändert.

Tabelle 5.1 zeigt die Merkmale, die keine Varianz aufweisen. Dabei lässt sich die Wahl der Merkmale 19, 20, 21, 22, 24, 25, 127, 128 und 129 damit begründen, dass diese auf den Stiftwinkeln

Altitude und Azimuth basieren, die nicht mit dem zugrunde liegenden Sensor erfasst werden können. Die Wahl der Merkmale 57, 108, 109 und 120 deutet darauf hin, dass diese gegebenenfalls nicht dazu geeignet sind, basierend auf unseren Test-Daten Einfluss auf die biometrische Verifikation beziehungsweise Hash-Generierung zu nehmen.

Tabelle 5.1: Liste der statistischen Merkmale, die aufgrund identischer Merkmalswerte in allen Samples aussortiert wurden

n_i	Name	Beschreibung
19	MaxAltitude	Maximum absolute altitude of pen occured during writing (or -1 if n/a)
20	MinAltitude	Minimum absolute altitude of pen occured during writing (or -1 if n/a)
21	MaxAzimuth	Maximum absolute azimuth of pen occured during writing (or -1 if n/a)
22	MinAzimuth	Minimum absolute azimuth of pen occured during writing (or -1 if n/a)
24	AverageAzimuth	Average Azimuth of pen projected on writing plane
25	AverageAltitude	Average Altitude of pen above the writing plane
57	PenUPress	Average PenUp Pressure normalized to 1 * 1000
108	Cluster0X	normalized X-coordinate of cluster 0
109	Cluster0Y	normalized Y-coordinate of cluster 0
120	Cluster0XYAvgPrs	average pressure of xy cluster 0
127	Cluster1Alti	average altitude of pressure cluster 1
128	Cluster2Alti	average altitude of pressure cluster 2
129	Cluster3Alti	average altitude of pressure cluster 3

Bei den Untersuchungen im Rahmen dieser Arbeit haben wir zwar die Varianz der Merkmale betrachtet, werden aber die darauf aufbauende Vorselektion nicht durchführen. Vielmehr werden wir im Verlauf der experimentellen Evaluierung überprüfen, ob die durch uns gewählten Verfahren diese Merkmale aussortieren. Für die Merkmalsanalyse haben wir die bereits in Grundlagenabschnitt 2.1.5 *Optimierung basierend auf der Merkmalsebene* vorgestellten Wrapper und Filter ausgewählt, welche im folgenden noch einmal zusammenfassen dargestellt werden.

5.3.2 Verwendete Wrapper-Methoden

Um das optimale Merkmal-Set zu finden, müssen alle möglichen sinnvollen Kombinationen der Merkmale untersucht werden. Daraus folgt ein erheblicher Zeit-und Rechenaufwand, wenn man beachtet, dass neben der Zusammenstellung aller Subsets auch für jedes einzelne die Klassifikation durchgeführt werden muss.

Aus diesem Grund haben wir uns mit den Wrappers Sequential Forward Selection (*sfs*) und Sequential Backward Selection (*sbs*) (siehe auch [JKP94], [GE03]) für die Verwendung von zwei Verfahren entschieden, die vom Blickpunkt der Komplexität weniger aufwändig sind. Da diese nicht sämtliche Kombinationen untersuchen, können sie nicht für das Selektieren der optimalen Merkmalsuntermenge garantieren. Für den *simple* Wrapper haben wir uns aufgrund seiner Einfachheit und - im Vergleich mit den anderen beiden - höheren Geschwindigkeit entschieden.

Sequential Forward Selection (*sfs*). Bei der *Sequential Forward Selection* wird mit einem leeren Set begonnen, das durch schrittweises Hinzufügen der besten Merkmale erweitert wird. Im Kontext dieser Arbeit ist das im aktuellen Schritt jeweils beste Merkmal das, welches beim Hinzufügen zum bestehenden Merkmal-Set die Fehlerrate basierend auf den aktuellen Testdaten minimiert.

Sequential Backward Selection (*sbs*). Die Strategie der Sequential Backward Selection beginnt hingegen mit dem kompletten Merkmal-Set und entfernt die schlechtesten Merkmale Schritt für Schritt. Das schlechteste Merkmal in jedem Durchlauf ist das, dessen Entfernung aus dem Set zur größten Verbesserung der Fehlerrate führt. Beide Verfahren (sfs und sbs) haben damit eine Komplexität von $O(k^2)$.

Simple Wrapper (*simple*). Beim Simple Wrapper-Verfahren wird die Klassifikation mit jedem der k Merkmale einzeln durchgeführt. Das Set der selektierten Merkmale setzt sich dann aus den i Merkmalen zusammen, die für sich genommen die geringsten i Fehlerraten besitzen.

Die drei Verfahren *sfs*, *sbs* und *simple* können allerdings nicht das Finden des für den jeweiligen Einsatzzweck optimalen Merkmals-Sets garantieren. Sie bieten jedoch im Vergleich zur vollständigen Suche basierend auf allen möglichen Konstellationen einen akzeptablen Aufwand an Zeit und Rechenkapazität, wie wir in initialen Untersuchungen festgestellt haben.

5.3.3 Verwendete Filter-Methoden

Im Bereich der Filter wurden drei Hauptansätze für die Merkmalsselektion untersucht: ANOVA (*anova*, *anova-2class*), Korrelation (*correlation*) und Entropie (*entropy*, *joint-entropy*) (vergleiche [GE03]). Dies stellt nur eine kleine Auswahl möglicher Methoden dar, die wir basierend auf Voruntersuchungen aufgrund der viel versprechenden Ergebnisse getroffen haben.

Für die Erläuterung der Filter werden die nachfolgend beschriebenen Variablen verwendet, welche zur Verdeutlichung in Gleichung 5.3.1 zusammenfassend dargestellt werden:

$$
\begin{array}{cccccccc}
y & x & & & & x_i & & \\
\Downarrow & \Downarrow & & & & \Downarrow & & \\
y^1: & x^1 = & (x_1^1, & x_2^1, & \cdots, & x_i^1, & \cdots, & x_m^1) \\
y^2: & x^2 = & (x_1^2, & x_2^2, & \cdots, & x_i^2, & \cdots, & x_m^2) \\
\vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\
y^j: & x^j = & (x_1^j, & x_2^j, & \cdots, & x_i^j, & \cdots, & x_m^j) \\
\vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\
y^n: & x^n = & (x_1^n, & x_2^n, & \cdots, & x_i^n, & \cdots, & x_m^n)
\end{array}
\tag{5.3.1}
$$

Gegeben ist die Menge der Merkmalsvektoren x mit der dazugehörigen Menge an Nutzer-IDs y. Der Index j bezieht sich in diesem Kontext auf die Anzahl der Merkmalsvektoren beziehungsweise der IDs denen diese sich zuordnen lassen. Folglich kann jeder Merkmalsvektor x^j für die zugehörige Nutzer-ID y^j beschrieben werden durch $x^j = (x_1^j,\ x_2^j,\ \ldots,\ x_m^j)$. Dabei ist m die Dimension des Vektors entsprechend der Anzahl der aus den Rohdaten extrahierten Merkmale. Zu beachten ist, dass jedem Nutzer eindeutig ein numerischer positiver Wert (Nutzer-ID) zugeordnet ist. Weiter sind die Elemente aus y nicht zwingend nach der ID sortiert, zudem können sich die IDs wiederholen. Jedes Merkmal an einem beliebigen Index i kann für alle Nutzer wie folgt als Vektor

von Merkmalen dargestellt werden $x_i = (x_i^1,\ x_i^2,\ \ldots,\ x_i^n)$ und der Vektor der Identitäten als $y = (y^1,\ y^2,\ \ldots,\ y^n)$, n ist die Anzahl der Vektoren im Testset. Weiterhin werden wir die Menge aller Samples x^j die einer bestimmten Nutzer-ID k zugeordnet werden können mit x^{j_k} bezeichnen. Die Verwendung der jeweiligen Filter führt für jedes Merkmal i zu einem Rankingwert $R(i)$ basierend auf $\langle x_i, y\rangle$.

Analysis of Variance (*anova/anova-2class*). Das Verhältnis zwischen den Inter-Klassen- und Intra-Klassen-Varianzen kann mithilfe der Analysis of Variance (ANOVA) untersucht werden. Dabei kann man davon ausgehen, dass ein Merkmal eine umso bessere Diskriminierungsperformanz hat, je geringer dessen Varianz σ innerhalb der Klasse ist. Gleichzeitig sollte dieses Merkmal zwischen den Klassen möglichst stark variieren. Dies kann durch die Differenz der Mittelwerte μ_1 beziehungsweise μ_2 der Werte vom $i-ten$ Merkmal der zwei Nutzer beschrieben werden. Der Fall, dass zwischen zwei Nutzern unterschieden werden soll, wird in Formel 5.3.2 dargestellt. Die Anzahl der verwendeten Test-Samples für die beiden Nutzer wird hier mit N_1 beziehungsweise N_2 angegeben. In der Formel beschreiben σ_1 und σ_2 die entsprechende Varianz der Werte von Merkmal i für beide Nutzer.

$$R(i) = \frac{N_1 * N_2 * (\mu_1 - \mu_2)^2}{N_1 * {\sigma_1}^2 + N_2 * {\sigma_2}^2} \tag{5.3.2}$$

Im Normalfall werden in einem biometrischen System jedoch mehr als zwei Personen berücksichtigt. Hier unterscheiden wir zwei Möglichkeiten der Anwendung von ANOVA. Auf der einen Seite wird für jeden Nutzer angenommen, dass alle anderen Nutzer einer gemeinsamen Klasse angehören, während die aktuell betrachtete Person seine eigene Klasse bildet (*anova-2class*). Dementsprechend kann der Rang eines Merkmals mittels Formel 5.3.2 bestimmt werden und der finale Ranking-Wert entspricht der Summe für alle Merkmale. Im zweiten Fall (*anova*) wird ein multivariater ANOVA-Test durchgeführt. Dabei wird die Variation eines Merkmals für eine Person ausgedrückt durch die Summe der Abweichungen aller Samples x^{j_k} des Nutzers mit der ID k vom entsprechenden Mittelwert μ_k. Die Variation eines Merkmals zwischen verschiedenen Personen wird bestimmt durch die Summe der Abweichungen der nutzerspezifischen Mittelwerte μ_k vom globalen Mittelwert μ. Der Wert für das Ranking des Merkmals kann entsprechend der Formel 5.3.3 bestimmt werden. In dieser Formel ist $R(i)$ der Rang von Merkmal i, K die Anzahl der Personen und N_k entspricht der Anzahl der Samples für die Nutzer-ID k.

$$R(i) = \frac{\frac{1}{K-1}\sum_{k=1}^{K} N_k(\mu_k - \mu)^2}{\frac{1}{N-K}\sum_{k=1}^{K}\sum_{j=1}^{N_k}(x^{j_k} - \mu_k)^2} \tag{5.3.3}$$

Korrelation (*correlation*). Dem Vorschlag von Guyon und Elisseeff in [GE03] folgend setzen wir in unseren Untersuchungen die Korrelation zwischen Merkmalen und Nutzern als Qualitätsmaß ein. Mittels Pearson-Korrellationskoeffizient wird der Ranking-Wert entsprechend der Formel 5.3.4 bestimmt. Dabei gehen wir davon aus, dass die Diskriminierung zwischen den Klassen der einzelnen Nutzer durch eine geringe Korrelation zwischen den Merkmalswerten und Klassen begünstigt wird. In Formel 5.3.4 stellt der Pearson-Korrelationskoeffizient den Ranking-Wert $R(i)$ dar, wobei μ_x und μ_y die Mittelwerte des Merkmals i beziehungsweise der Identitäten sind. Beträgt der Wert $R(i)$ 1, stellt dies eine maximale Korrelation dar und dies bedeutet absolute Relevanz des Merkmals i bezüglich der Authentifikation. Wird $R(i)$ dagegen 0, besteht keine Korrelation zwischen dem Merkmal und dem Nutzer, und damit ist das Merkmal i für eine Authentifikation irrelevant. Bei der Betrachtung der Korrelation ist zu beachten, dass dadurch nur lineare Zusammenhänge erkannt werden können.

$$R(i) = \left| \frac{\sum_{j=1}^{N} (x^j - \mu_x)(y^j - \mu_y)}{\sqrt{\sum_{j=1}^{N} (x^j - \mu_x)^2 * \sum_{j=1}^{N} (y^j - \mu_y)^2}} \right| \tag{5.3.4}$$

Entropie (*entropy*). In dieser Variante der Entropie gehen wir von der Annahme aus, dass sich ein für den jeweiligen Einsatzzweck relevantes Merkmal einerseits durch eine hohe Inter-Klassen-Entropie auszeichnet. Gleichzeitig sollte es andererseits über eine geringe Intra-Klassen-Entropie verfügen. Für die Berechnung wird einerseits die Summe der Entropien der durchschnittlichen Merkmalswerte der Nutzer $\mu = (\mu_1,\ \mu_2,\ \ldots,\ \mu_k)$ herangezogen (Inter-Klassen-Entropie), zum anderen wird die Summe der Entropien der Merkmalswerte x^{j_k} der einzelnen Nutzer (Intra-Klassen-Entropie) verwendet. Die Bestimmung des auf dem Verhältnis der Summen der beiden Entropien basierenden Rankingwerts für Merkmal i wird in Formel 5.3.5 dargestellt. Hier sind N_μ die Anzahl der Mittelwerte der Nutzer, N_k die Anzahl der Samples für die Person mit der ID k und K die Anzahl der Nutzer.

$$R(i) = \frac{\sum_{j=1}^{N_\mu} H_j(\mu) * log_2 \frac{1}{H_j(\mu)}}{\sum_{k=1}^{K} \sum_{j=1}^{N_k} H_j(x^{j_k}) * log_2 \frac{1}{H_j(x^{j_k})}} \tag{5.3.5}$$

Joint-Entropie (*joint-entropy*). Die Joint-Entropie bietet eine weitere Möglichkeit zur Erzeugung eines Rankings auf Basis der Qualität der Merkmale. Die empirische Abschätzung des Informationsgehaltes bezüglich der Merkmale und der Nutzerklassen stellt dabei das Qualitätsmaß für die einzelnen Merkmale dar. Basierend auf der Annahme, dass die Werte der Merkmale diskret sind, kann dieser Informationsgehalt entsprechend der Gleichung

5.3.6 für ein Merkmal bestimmt werden.

$$R(i) = \sum_{x_i} \sum_{y} P(X = x_i, Y = y) * log \frac{P(X = x_i, Y = y)}{P(X = x_i) * P(Y = y)} \tag{5.3.6}$$

5.4 Biometrische Fusion (A.4)

Ziel einer multi-biometrischen Fusion ist in den meisten Fällen die Verbesserung der Erkennungsgenauigkeit im Vergleich zu den Resultaten der beteiligten einzelnen Komponenten. Allerdings nimmt in den in Abschnitt 2.1.5 *Optimierung durch biometrische Fusion* vorgestellten Szenarien multi-sensor Fusion, multi-instance Fusion, multi-sample Fusion und multi-modal Fusion mit jeder zusätzlichen Fusionskomponente auch die Komplexität in der Interaktion zwischen dem multi-biometrischen System und dem Benutzer zu. Der Nutzer hat mehr als eine biometrische Eigenschaft zu präsentieren beziehungsweise mehrmals dieselbe. Hinzu kommt jeweils die korrekte Ausrichtung der Eigenschaft zum Sensor. Dies kann unter Umständen leicht zu einer Überforderung des Nutzers führen, wodurch sich die Akzeptanz durch die potentiellen Anwender verringert. Bei einem multi-modalen System basierend auf Fingerabdruck und Handschrift kann es beispielsweise notwendig sein, in der richtigen Reihenfolge einem Fingerabdruck- und einem Handschriftsensor die entsprechende biometrische Modalität zur Verfügung zu stellen.

Auf der anderen Seite sind multi-biometrische Systeme auch in der Lage, Probleme zu umgehen, die durch Schwierigkeiten bei der Datenerfassung oder durch zeitlich begrenzte beziehungsweise permanente Abwesenheit einer biometrischen Charakteristik bedingt sind. Solche Probleme können durch Verletzungen, Krankheit beziehungsweise Alterung entstehen. Vor allem bei der multi-modalen Fusion besteht die Möglichkeit, in diesem Fall die betreffende Charakteristik zu ignorieren und entsprechend gegebener Systemeinstellungen zu kompensieren, zum Beispiel durch die Nutzung alternativer Gewichtungen der verbleibenden Modalitäten. Die multi-algorithmische Fusion bietet den Vorteil, dass hier nur einmal Daten erfasst werden, welche dann von mindestens zwei Algorithmen verarbeitet werden.

Das Ziel dieses Abschnittes ist es, verschiedene Ansätze zu diskutieren, die sich mit der Fusion unter Verwendung einer einzelnen Modalität befasst. Aufgrund unserer Vorarbeiten konzentrieren wir uns dabei auf die dynamische Handschrift und werden uns an der Klassifikation von Ross et al. ([RNJ06], siehe auch 2.1.5 *Optimierung durch biometrische Fusion*) orientieren. Abbildung 5.9 zeigt eine Taxonomie für mögliche biometrische Fusionen basierend auf nur einer einzelnen Modalität. Diese Art der biometrischer Fusion werden wir im Folgenden auch als single-modal Fusion bezeichnen. Folgt man den jeweiligen möglichen Pfaden des Graphen in Abbildung 5.9, erhält man jeweils eine Möglichkeit, biometrische Komponenten einer Modalität miteinander zu fusionieren. Wir werden nachfolgend nicht jede einzelne Kombinationsmöglichkeit diskutieren, sondern uns auf diejenigen konzentrieren, für die wir während der jeweiligen Evaluierungen basierend auf der dynamischen Handschrift vielversprechende Ergebnisse erzielen konnten.

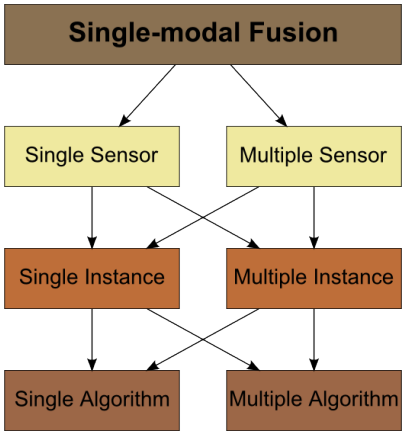


Abbildung 5.9: Schematische Darstellung der single-modalen Fusionsmöglichkeiten

5.4.1 Bestimmung von Parametern für die Fusion auf Matching Score Level

Für die Fusion auf Matching Score Level haben wir verschiedene Möglichkeiten der Parametrisierung entwickelt [SVD05]. Mit diesen werden die Matching Scores s_i durch unterschiedliche Wichtungen w_i beeinflusst. Der fusionierte Matching Score s_{fus} wird durch die Addition der gewichteten Matching Scores $s_i w_i$ bestimmt (siehe Gleichung 5.4.1).

$$\text{Fusion: } s_{fus} = s_1 w_1 + s_2 w_2 + \ldots + s_n w_n \tag{5.4.1}$$

$$\begin{aligned} \text{Matching Scores: } & s_1, s_2, \ldots, s_n \\ \text{Fusionsparameter: } & w_1, w_2, \ldots, w_n \\ \text{Bedingung: } & \sum_{i=1}^{n} w_i = 1 \end{aligned}$$

Basierend auf dieser Fusionsidee haben wir vier Strategien entwickelt, um Gewichte zu bestimmen. Diese werden im folgenden kurz beschrieben.

Binär gewichtete Fusion. Bei der binär gewichteten Fusion wird nur diejenige Fusionskomponente berücksichtigt, die während der Tests den höchsten Matching Score erreicht hat (5.4.2). Damit entspricht der fusionierte Matching Score dem Score der besten Einzelkomponente.

$$w_{binary\ i} = \begin{cases} 1, wenn\ s_i = max(s_1, s_2, \ldots, s_n) \\ 0\ sonst \end{cases} \tag{5.4.2}$$

Gleich gewichtete Fusion. Identische Gewichte werden den Fusionskomponenten bei der gleich gewichteten Fusion zugeordnet. Entsprechend der Gleichung 5.4.3 ist dieses Gewicht gleich

dem n-ten Teil der Anzahl der beteiligten Einzelkomponenten. Damit ist diese Strategie unabhängig von der Performanz der einzelnen beteiligten Komponenten.

$$w_{equal\ i} = \frac{1}{n} \tag{5.4.3}$$

Linear gewichtete Fusion. Für die linear gewichtete Fusion wird die Equal Error Rate jeder einzelnen Komponente ins Verhältnis zur Summe der EERs aller Komponenten gesetzt. Für die Fusion werden die so bestimmten Gewichte den Matching Scores in umgekehrter Reihenfolge zugeordnet, so dass die Komponente mit dem besten Score am höchsten gewichtet wird und umgekehrt. Auf diese Weise fließt die Performanz jeder einzelnen Komponente in die Wichtungsstrategie ein. Auf der anderen Seite wird der Einfluss der Komponente mit der schlechtesten Performanz auf das Fusionsergebnis geringer ausfallen.

$$w_{linear\ i} = \frac{EER_i}{\sum_{m=1}^{n} EER_m} \tag{5.4.4}$$

Quadratisch gewichtete Fusion. Da sich in initialen Tests gezeigt hat, dass mittels linear gewichteter Fusion eine Verbesserung der Verifikationsperformanz erzielt werden kann, werden die auf diese Weise ermittelten Gewichte als Grundlage für die quadratische Fusion genutzt. Durch die Multiplikation eines Gewichtes mit sich selbst und der anschließenden Division durch die Summe der quadrierten Einzelwichtungen wird die Auswirkungen der individuellen Werte auf die korrespondierenden Matching Scores erhöht. Auf diese Weise wird der Score der besten Fusionskomponente wesentlich höher bewertet als bei der linearen Wichtungsstrategie.

$$w_{quadr\ i} = \frac{{w_{linear\ i}}^2}{\sum_{m=1}^{n} {w_{linear\ m}}^2} \tag{5.4.5}$$

Basierend auf diesen Strategien können diverse Komponenten auf Matching Score Level fusioniert werden. Im Folgenden nutzen wir diese für eine Auswahl von single-modalen Fusionsmöglichkeiten von Komponenten, die auf der dynamischen Handschrift beruhen. Die hier von uns eingesetzten Strategien zur Fusion auf Matching Score Level stellen nur eine kleine Auswahl von Möglichkeiten dar, biometrische Komponenten miteinander zu kombinieren (siehe beispielsweise auch [GFA+07]).

5.4.2 Handschriftbasierte multi-algorithmic Fusion auf Matching Score Level

In Vorarbeiten (zum Beispiel in [SVD05] und [SVF11]) konnten wir zeigen, dass durch die Fusion von vier beziehungsweise zwei BioHash-Algorithmen mit jeweils unterschiedlichen Distanzmaßen

deutliche Verbesserungen bezüglich der Verifikationsperformanz erreicht werden können. Zur Bestimmung der Ähnlichkeit von Referenz- und Verifikationsdaten wurden in [SVF11] Hamming- und Canberra-Distanz und in [SVD05] zusätzlich City-Block- und Euklidischer Abstand verwendet. Dabei konnten Verbesserungen der Equal Error Rate von bis zu 12% [SVD05] beziehungsweise von bis zu 3% [SVF11] erzielt werden.

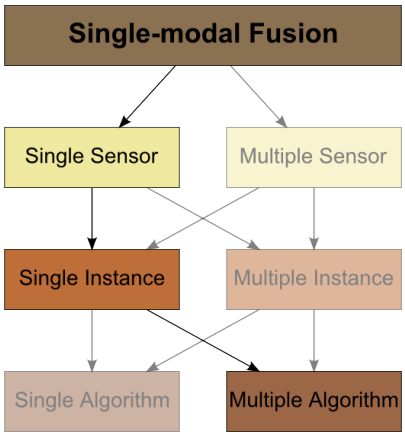


Abbildung 5.10: Schematische Darstellung der single-modalen multi-algorithmic Fusion

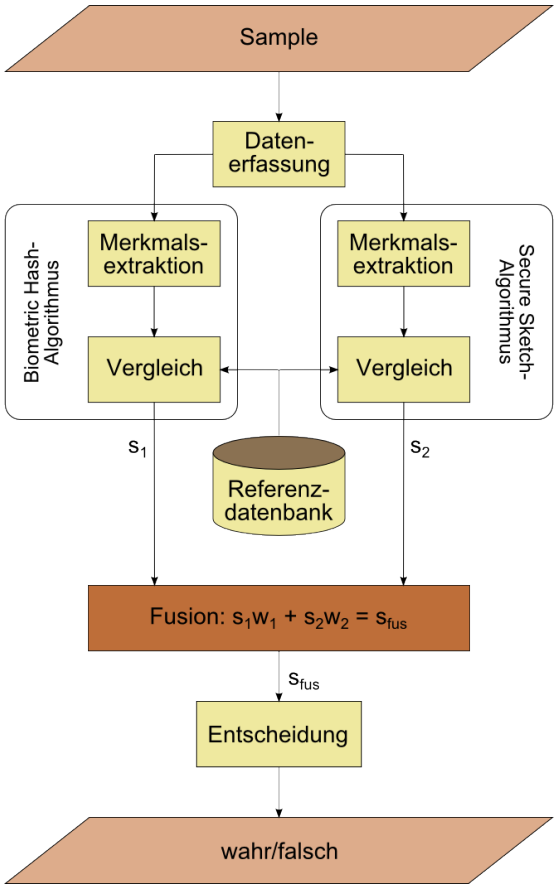


Abbildung 5.11: Multi-algorithmische Fusion auf Matching Score Level basierend auf zwei handschriftbasierten Algorithmen

Die multi-algorithmische Fusion, die wir im Rahmen dieser Arbeit durchgeführt und untersucht haben, verwendet beide Referenzalgorithmen. Wie in Abbildung 5.10 dargestellt, werden Daten mittels eines einzelnen Sensors aufgenommen. Bei den Daten handelt es sich in diesem Fall jeweils um verschiedene Samples einer Semantikklasse. Den beiden Algorithmen werden dann identische Daten zur Verfügung gestellt. Ein großer Vorteil für den Anwender besteht bei dieser Methode darin, dass die erforderlichen Daten nur einmal aufgenommen werden brauchen.

Bei der Evaluierung dieses Szenarios verwenden wir jeweils die in Abschnitt 5.4.1 *Bestimmung von Parametern für die Fusion auf Matching Score Level* vorgestellten Vorgehensweisen zur Wahl der Fusionsparameter für beide beteiligten Algorithmen in den jeweiligen Strategien. In Abbildung 5.11 wird die multi-algorithmische Fusion basierend auf der Handschrift schematisch dargestellt, wie sie von uns durchgeführt wurde. Wie in den bisher beschriebenen Optimierungsverfahren arbeiten beide Algorithmen auch in unseren Fusionsansätzen mit dem selben Set statistischer Merkmale.

Bei der Fusion auf Matching Score Level werden wir in der experimentellen Evaluierung nur die Erkennungsgenauigkeit der Verifikation mittels Equal Error Rate untersuchen. Der Grund dafür ist die Tatsache, dass mit der Bestimmung beider Matching Scores nach der Fusion keine Hash-Werte zur Verfügung stehen. Dadurch ist die Auswertung dieser zur Bestimmung von Reproduction Rate, Collision Rate und Collision Reproduction Rate nicht möglich. Dies gilt für alle in dieser Arbeit untersuchten Fusionen auf Matching Score Level.

5.4.3 Handschriftbasierte multi-instance Fusion auf Matching Score Level

Bei der von uns vorgeschlagenen handschriftbasierten multi-instance Fusion werden, entsprechend Abbildung 5.12, über einen einzelnen Sensor nacheinander je ein Sample der gewünschten Semantiken aufgenommen.

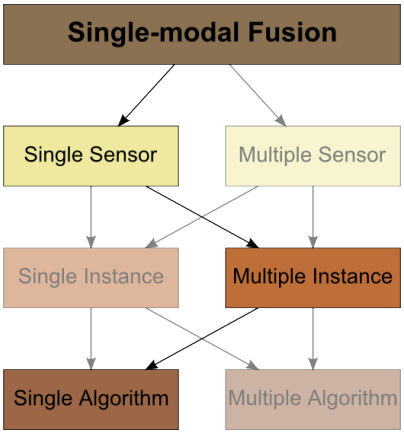


Abbildung 5.12: Schematische Darstellung der single-modalen multi-instance Fusion

Diese werden dann von jeweils einem der beiden Algorithmen im Verifikationsmodus verarbeitet. Dadurch, dass hier der Benutzer nacheinander Proben der zwei beteiligten Semantiken abgeben muss, erhöht sich die Bedienungskomplexität des Systems. Dies hält sich allerdings dadurch in

Grenzen, dass nur ein Sensor bedient werden muss. Zusätzlich zum Faktor geheimes Wissen in Bezug auf den Inhalt der geforderten Schriftproben kann bei einem entsprechendem System die Sicherheit beispielsweise durch das Wissen über die geforderte Reihenfolge der Instanzen erhöht werden. Der Einfluss der Reihenfolge der Aufnahme der benötigten Instanzen wird im Rahmen dieser Arbeit nicht untersucht.

Die Abbildung 5.13 stellt den Vorgang der von uns vorgeschlagenen single-modal multi-instance Fusion dar. Es werden die beiden unterschiedlichen Schreibinhalte nacheinander von einem Sensor erfasst und vom selben Algorithmus bis hin zum Vergleich verarbeitet. Der jeweils berechnete Matching Score geht dann gewichtet entsprechend des verwendeten Algorithmus und der aktuell eingesetzten Strategie in die Fusion ein. Der dabei berechnete fusionierte Matching Score geht dann in die schwellwertbasierte Entscheidung ein. Die Evaluierung der multi-instance Fusion umfasst im Rahmen dieser Arbeit die möglichen paarweisen Kombinationen von fünf Semantiken ohne Berücksichtigung der Reihenfolge der Aufnahme.

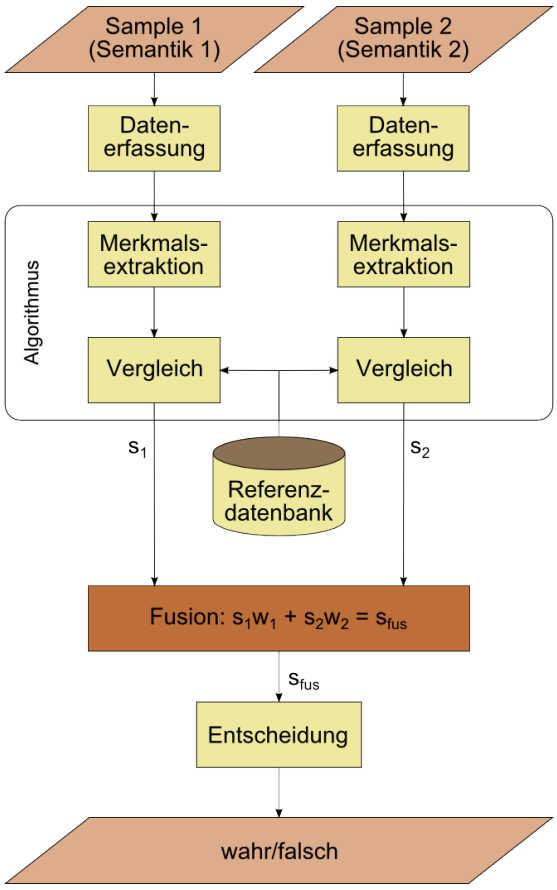


Abbildung 5.13: Multi-instance Fusion basierend auf zwei unterschiedlichen Semantiken

5.4.4 Handschriftbasierte multi-sample Fusion auf Matching Score Level

Im Gegensatz zur multi-instance Fusion bilden bei der multi-sample Fusion mehrere Samples derselben Semantik den Input (siehe Abbildung 5.14). Diese werden durch einen einzelnen Sensor aufgenommen und die Verifikation erfolgt über einen einzelnen Algorithmus. Auch diese

Form der single-instance Fusion verlangt vom Benutzer des Systems einen Mehraufwand bei der Erfassung seiner Daten, da hier ebenfalls für jede beteiligte Fusionskomponente eine separate Handschriftenprobe aufgenommen werden muss.

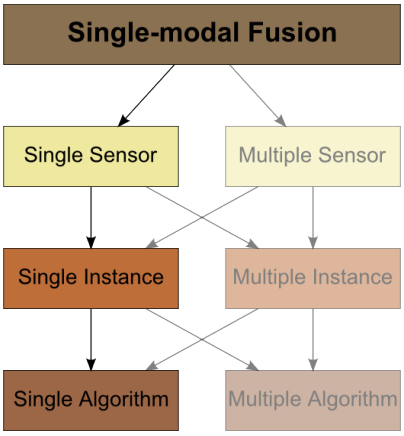


Abbildung 5.14: Schematische Darstellung der single-modalen multi-sample Fusion

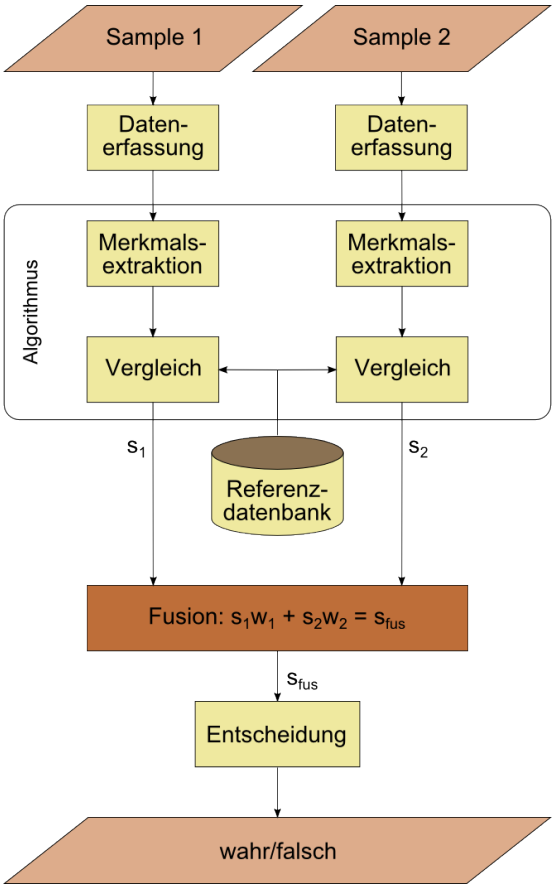


Abbildung 5.15: Multi-sample Fusion basierend auf zwei Samples derselben Semantik

Die Vorgehensweise bei der von uns eingesetzten multi-sample Fusion ist ähnlich der der multi-instance Fusion. Auch hier werden, wie in Abbildung 5.15 gezeigt, beide Samples von einem

Sensor aufgenommen. Beide Rohdaten werden einzeln vom jeweiligen Algorithmus bis inklusive des Vergleiches mit den Referenzdaten verarbeitet. Der fusionierte Matching Score, auf dem die Entscheidung beruht, berechnet sich ebenfalls aus der gewichteten Summe der einzelnen Matching Scores. Die multi-sample Fusion wird jeweils für beide Algorithmen mit fünf Semantiken evaluiert.

In Tabelle 5.2 werden die im Rahmen dieser Arbeit untersuchten Fusionsstrategien zusammenfassend gegenübergestellt. In den ersten beiden Spalten werden Strategie und Fusions-Level angegeben, während die dritte Spalte den beziehungsweise die beteiligten Algorithmen enthält. Spalte *Input* stellt die Anzahl der Semantiken und Samples dar, die jeweils an der Fusion beteiligt sind. Nachfolgend wird in Spalte *Merkmale* das zugrunde liegende Merkmalsset angegeben. Die letzten zwei Spalten informieren darüber, ob basierend auf der jeweiligen Fusion Untersuchungen der Verifikation (V) beziehungsweise Hash-Generierung möglich sind.

Tabelle 5.2: Vergleich der untersuchten Fusionsstrategien (V, H: Evaluierung des Verifikations- beziehungsweise Hash-Generierungsmodus möglich)

Fusion	Fusion Level	Algorithmus	Input	Merkmale	V	H
multi-algorithmisch	Matching Score	Biometric Hash Secure Sketch	1 Semantik 1 Sample	131 statistische	ja	nein
multi-instance	Matching Score	Biometric Hash	2 Semantiken je 1 Sample	131 statistische	ja	nein
		Secure Sketch	2 Semantiken je 1 Sample	131 statistische	ja	nein
multi-sample	Matching Score	Biometric Hash	1 Semantik 2 Samples	131 statistische	ja	nein
		Secure Sketch	1 Semantik 2 Samples	131 statistische	ja	nein

6 Aufbau und Methodologie der experimentellen Evaluierung

In diesem Kapitel werden die in der experimentellen Evaluierung eingesetzten Performanzmaße angegeben und sowohl die Herkunft als auch der Aufbau der Testdatenbank beschrieben. Darauf aufbauend erfolgt die Erläuterung der zur Untersuchung der einzelnen Teilaufgabenstellungen *A.1 Hinzufügen zusätzlicher Merkmale*, *A.2 Optimierung von Parametern*, *A.3 Analyse und Selektion von Merkmalen*, *A.4 Biometrische Fusion* und *B Alterung biometrischer Daten* (siehe Abschnitt 1.2 *Aufgabenstellung*) entworfenen Szenarien.

6.1 Evaluierungsmethodologie und Aufbau der Datenbank

Zur Bewertung der Performanz des Verifikationsmodus beider Algorithmen in Verbindung mit den in dieser Arbeit vorgeschlagenen Verfahren zur Optimierung verwenden wir die im Abschnitt 2.1.6 *Biometrische Fehlerraten* vorgestellten biometrischen Fehlerraten False Match Rate (FMR), False Non Match Rate (FNMR) und Equal Error Rate (EER). Zur Einschätzung der Veränderungen im Hash-Generierungsmodus kommen die von uns speziell für biometrische Hash-Funktionen entwickelten Performanzmaße Reproduction Rate (RR), Collision Rate (CR) und Collision Reproduction Rate (CRR) zum Einsatz (siehe auch [SVD08a], [MSV11a]), die im Abschnitt 2.1.7 *Reproduktion und Kollision im Hash-Generierungsmodus* beschrieben werden. Weiterhin werden hier Aufbau und Verwendung der Testsets in Zusammenhang mit den jeweiligen Szenarien erläutert. Diese bilden jeweils die Grundlage zur Evaluierung der beiden Algorithmen unter Betrachtung der von uns vorgeschlagenen Optimierungsansätze im Verifikations- ebenso wie im Hash-Generierungsmodus.

Um die Auswirkungen der von uns im Rahmen dieser Arbeit vorgenommenen Veränderungen an den Algorithmen zu untersuchen, haben wir die uns zu Verfügung stehenden Daten in verschiedenen Szenarien evaluiert. Als Ausgangspunkt dient dabei der Test beider Verfahren mit dem ursprünglichen Merkmal-Set. Dieses basiert auf den ersten 69 Merkmalen ([Vie06], siehe auch Tabelle A.1 im Anhang A *Statistische Merkmale*), im Folgenden wird dieses Set auch als *Set A* bezeichnet. Die Menge *Set B* beinhaltet die 62 zusätzlichen Merkmale (siehe 5.1 *Hinzufügen zusätzlicher Merkmale (A.1)* und Anhang A *Statistische Merkmale*, Tabelle A.2) die wir zur weiteren Verbesserung der Hash-Algorithmen entwickelt haben.

Set A. Menge der ersten 69 Merkmale, beschrieben in [Vie06]

Set B. Menge der neuen 62 Merkmale, beschrieben in den folgenden studentischen Arbeiten:

Set B_1. Merkmale $n_{70} - n_{89}$, beschrieben in [Guz07]

Set B_2. Merkmale $n_{90} - n_{103}$, beschrieben in [FHP08]

Set B_3. Merkmale $n_{104} - n_{131}$, beschrieben in [HHH09]

Die nachfolgenden Abschnitte beschreiben das Setup für die Evaluierungen bezüglich der unterschiedlichen Ansätze zur Optimierung und der Untersuchung der Alterung.

Alle im Rahmen dieser Arbeit durchgeführten Tests basieren auf derselben Datenbank. Diese enthält Daten von insgesamt 53 Schreibern. Die Probanden wurden gebeten, zu drei Aufnahmezeitpunkten (Sessions) im Abstand von einem Monat jeweils zehn Schriftproben in den fünf Semantiken *gegebene PIN*, *geheime PIN*, *Pseudonym*, *Symbol* und *Woher* abzugeben. Der zeitliche Abstand zwischen den einzelnen Sessions betrug jeweils circa einen Monat. Daraus ergibt sich eine Gesamtzahl von 7950 Schriftproben, die für die experimentelle Untersuchung herangezogen werden. Die fünf verwendeten Semantiken werden nachfolgend aufgelistet und kurz erläutert.

gegebene PIN. Als *gegebene PIN* wird die Ziffernfolge *77993* für alle Testpersonen vorgegeben. Dadurch werden Unterschiede zwischen den einzelnen Schriftproben unterschiedlicher Personen hauptsächlich durch die individuelle Art zu schreiben verursacht. Die Inter-Klassen-Ähnlichkeit wird durch die Verwendung eines sehr kleinen Grundzeichensatzes (Ziffern von 0-9) und zusätzlich durch die Vorgabe von nur drei Zeichen daraus stark eingeschränkt.

geheime PIN. Bei der *geheime PIN* kann die Testperson eine fünfstellige Ziffernkombination frei wählen. Auch hier steht nur die Anzahl von zehn unterschiedlichen Ziffern zur Verfügung. Allerdings ist deren Kombination untereinander nicht vorgegeben. So ergeben sich theoretisch 100.000 (10^5) mögliche fünfstellige Kombinationen. Dadurch verringert sich die Wahrscheinlichkeit, dass eine andere Testperson zufällig dieselbe Kombination wählt, beziehungsweise ein Angreifer diese rät.

Pseudonym. Während unserer Vorarbeiten haben viele Testpersonen Bedenken geäußert, Proben ihrer Unterschrift abzugeben. Zudem sind gerade Unterschriften zum großen Teil nicht anonym und bedürfen datenschutzbedingt umfangreicher Schutzmaßnahmen bei der Erfassung, Speicherung und Bearbeitung. Aus diesen Gründen sind wir dazu übergegangen, die Signatur durch das Pseudonym zu ersetzen. Für diese Semantik wird der Schreiber aufgefordert, sich einen Namen zu wählen, diesen zu üben und dann erst die eigentliche Datenerfassung seiner Proben durchzuführen.

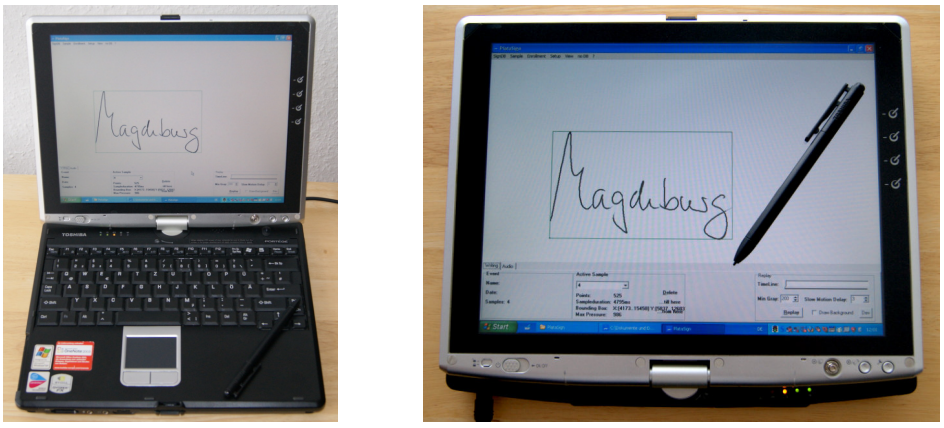
Symbol. Für die Aufnahme der Schriftproben für diese Semantik war jeweils ein Symbol oder eine einfache Zeichnung zu wählen. Dadurch gewinnt diese Semantik zu einem gewissen Grad Individualität als auch Vertraulichkeit. Letztere basiert in diesem Fall nicht nur auf der (geheimen) Wahl des Symbols, sondern auch auf der Reihenfolge der dazu gesetzten Segmente.

Woher. Die Antwort auf die Frage *Woher kommen Sie?* ist der Inahlt der Semantikklasse Woher. Hier verfügt der Schreiber über gewisse Einschränkungen, aber auch Freiheiten bei der Wahl des Inhaltes. Beobachtungen haben gezeigt, dass die Testpersonen auf diese Frage häufig mit ihrem Wohnort oder -land beziehungsweise ihrem Geburtsort oder -land antworteten. Einige Schreiber haben aber auch den Ort genannt, von dem sie vor der Datenaufnahme kamen. Durch die Eingrenzung der Angabe eines Ortes wird die Wahl für den Schreiber etwas eingeengt, wodurch sich die Wahrscheinlichkeiten erhöhen, dass entweder eine andere Testperson denselben Ort wählt oder ein Angreifer diesen eher errät.

Der zeitliche Abstand zwischen den einzelnen Sessions betrug jeweils circa einen Monat. Um einen gewissen Grad an Vergleichbarkeit beizubehalten, wurden die drei Datenaufnahmen in derselben räumlichen Umgebung durchgeführt, jeweils betreut durch dieselbe Person. Als Sensor wurde ein TabletPC (Toshiba Pórtegé M200, Abbildung 6.1) mit den in Tabelle 6.1 angegebenen technischen Eigenschaften eingesetzt.

Tabelle 6.1: Ausstattung des Sensors Toshiba Pórtegé M200 zur Erfassung der Evaluierungsdaten

Komponente	**Ausstattung**
Prozessor	Intel Pentium M 1.6 GHz
Speicher	1024 MB
Display-Größe	12,1 Zoll
grafische Auflösung	1400 x 1050 Pixel
Druck-Stufen	1024
Sampling Rate	160
Stiftwinkel	nicht unterstützt
Betriebssystem	Microsoft Windows XP Tablet Edition
Aufnahmesoftware	PlataSign



(a) (b)

Abbildung 6.1: Toshiba Pórtegé M200: Sensor zur Erfassung der Evaluierungsdaten im (a) Notebook- beziehungsweise (b) Tablet-Betrieb

Wie in Tabelle 6.1 zu sehen ist, unterstützt der Sensor keine Aufnahme der Stiftwinkel Azimut und Altitude. Die daraus ermittelten statistischen Merkmale n_{18}-n_{22}, n_{24}, n_{25} und n_{127}-n_{129} werden aber, wie in Abschnitt 5.3 *Analyse und Selektion von Merkmalen (A.3)* beschrieben, nicht ausgeschlossen. Der Grund dafür ist, dass wir überprüfen werden, ob die für die Merkmalsanalyse ausgewählten Verfahren diese Merkmale ebenfalls aussortieren werden.

6.2 Evaluierungs-Setup

In den folgenden Abschnitten gehen wir auf die einzelnen Evaluierungsszenarien ein, die wir basierend auf den Teilaufgabenstellungen in Abschnitt 1.2 *Aufgabenstellung* und den dazu entstandenen Optimierungsansätzen und zur Untersuchung der Alterung biometrischer Daten erstellt haben. Bei den Szenarien handelt es sich um die Betrachtung der Ausgangssituation, Hinzufügen neuer Merkmale, Parameteroptimierung, Merkmalsanalyse und -selektion, single-modal Fusion auf Matching Score Level und Alterung.

Mit Ausnahme der Testdaten für die Alterung werden die vorhandenen Daten jeweils so aufgeteilt, dass erstens die Informationen der Session 1 zur Erzeugung der Referenzen genutzt werden. Zweitens werden die Daten der Session 2 verwendet, um die Ansätze zur Optimierung zu verfolgen (zum Beispiel Bestimmung der Parameter, Durchführung der Merkmalsanalyse). Da es teilweise notwendig ist, für die Optimierungen Verifikationen beziehungsweise Hash-Generierungen durchzuführen, werden dafür die Referenzdaten aus Session 1 herangezogen. Die Daten der Session 3 werden drittens als Testdaten genutzt, um zusammen mit den Referenzen aus Session 1 in der experimentellen Evaluierung den Erfolg der zu untersuchenden Strategien feststellen zu können.

Für die Untersuchung der Alterung biometrischer Referenzdaten über einen kurzen Zeitraum nutzen wir die zeitliche Differenz zwischen den Aufnahme-Sessions. Dazu erstellen wir jeweils aus den Informationen einer Session sowohl Referenz- als auch Testdaten. Eine Optimierung im Alterunhsszenario findet nicht statt, da einerseits die Verwendung der Ergebnisse der Optimierungsansätze bedeuten würde, dieselben Daten sowohl für die Optimierung als auch für die Evaluierung einzusetzen. Andererseits stehen pro Session zu wenig Daten zur Verfügung, um zusätzlich Optimierungen durchzuführen.

6.2.1 Setup zu Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale

In diesem ersten Teil der Evaluierung wird für beide Algorithmen und die fünf Semantiken zuerst das Merkmal-Set Set A zur Berechnung der Performanz zu Beginn unserer Arbeiten genutzt. Diese Werte dienen als Referenz zur Einschätzung des Erfolges der einzelnen Optimierungsschritte. Aus diesem Grund werden an dieser Stelle die später zu untersuchenden Parameter Tolerance Factor und Expansion Factor gleich Null gesetzt.

Darauf aufbauend folgt der erste Ansatz zur Optimierung durch das Hinzufügen zusätzlicher Merkmale. Diese wird einerseits durch das schrittweise Hinzufügen der Merkmal-Sets Set B_1, Set B_2 und Set B_3 zum Set A durchgeführt. Nach jedem Anhängen einer zusätzlichen Teilmenge wird

jeweils die Performanz für Verifikations- und Hash-Generierungsmodus ermittelt und ausgewertet. Auf der anderen Seite wird aufgrund der unterschiedlichen Größe der vier Sets auch der Verlauf von Equal Error Rate beziehungsweise Collision Reproduction Rate abhängig von der Anzahl der einzelnen Merkmale untersucht.

Dabei werden die zur Verfügung stehenden Testdaten wie nachfolgend beschrieben verwendet (analog für beide Algorithmen in beiden Anwendungsmodi und jeweils die fünf Semantiken): Für die Referenzdaten werden die Samples der ersten Session genutzt. Zur Bestimmung der Performanzmaße Equal Error Rate, Collision Reproduction Rate, Reproduction Rate und Collision Rate werden als Testdaten die Schriftproben der dritten Session verwendet. Die Samples der mittleren Session bleiben bei diesem Teil der Evaluierung ungenutzt.

Für die erforderlichen Tests werden die zur Verfügung stehenden Daten nach dem folgenden Schema genutzt:

Daten der Session 1. Bestimmung der Referenzdaten für

- Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
- Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)

Daten der Session 2. nicht genutzt

Daten der Session 3. Ermittlung der Performanzmaße Equal Error Rate (EER), Collision Reproduction Rate (CRR), Reproduction Rate (RR) und Collision Rate (CR) für

- Biometric Hash
- Secure Sketch

6.2.2 Setup zu A.2 Optimierung von Parametern

Für den nächsten Optimierungsschritt haben wir für jeden der beiden Algorithmen einen Parameter identifiziert, der für die Quantisierung im jeweiligen Algorithmus verwendet wird. Während wir für den Biometric Hash-Algorithmus den Tolerance Factor TF wählen (siehe auch Abschnitt 4.1 *Biometric Hash-Algorithmus für die dynamische Handschrift*), nutzen wir zur Untersuchung des Secure Sketch den Parameter Expansion Factor EF (Abschnitt 4.2 *Secure Sketch-Algorithmus für die dynamische Handschrift*). Die Parameteroptimierung wird entsprechend der von uns entwickelten und in Abschnitt 5.2 *Optimierung von Parametern (A.2)* beschrieben Vorgehensweise durchgeführt (vergleiche auch [MSV11a]). Die im folgenden angegebenen Schritte erfolgen analog für beide Algorithmen und jeweils für die fünf Semantiken.

Als Referenzdaten zur Optimierung der Parameter werden die Samples der ersten Session herangezogen. Dann werden mittels der Testdaten aus der zweiten Session die Performanzmaße Equal Error Rate, Collision Reproduction Rate, Reproduction Rate und Collision Rate für unterschiedliche Parametrisierungen ermittelt. Die für jeden Durchlauf benötigten Werte für TF beziehungsweise EF werden durch schrittweise Erhöhung dieser Parameter um 0,25 bestimmt.

Diese Schrittweite hat sich in initialen Untersuchungen als ausreichend gezeigt, um geeignete Parameter zu bestimmen. Als Intervall, in dem die schrittweise Erhöhung durchgeführt wird, haben wir für TF $[0, 5]$ und für EF $[0, 13]$ gewählt, da sich eine Verringerung von EER, CRR und CR beziehungsweise eine Verbesserung von RR über diese Bereiche bei Voruntersuchungen hinaus nicht ergeben hat. Eine Ausnahme bildet hier die Semantik *Symbol* für den Secure Sketch. Hier musste das Intervall auf $[0, 30]$ erhöht werden, um eine genügend kleine CRR zur Parameterbestimmung zu ermitteln. Abschließend werden für jede Konstellation (zwei Algorithmen und fünf Semantiken) die Parameterwerte entsprechend der in Abschnitt 5.2 *Optimierung von Parametern (A.2)* definierten Szenarien (*Null*, *minEER*, *minCRR*, $RR \geq a\%$ und $CR \leq b\%$) für TF und EF bestimmt. Daraus ergeben sich für jeden der beiden Algorithmus fünf Parameter-Sets, die jeweils die Werte für die fünf Semantiken enthalten.

Die im zweiten Schritt ermittelten fünf Parameter-Sets werden dann verwendet, um basierend auf den Samples der ersten Session sowohl die Referenzdaten für die Evaluierung als auch aus den Samples der dritten Session die Testdaten zu generieren. Beide werden dann zur Bestimmung der entsprechenden Performanzmaße Equal Error Rate, Collision Reproduction Rate, Reproduction Rate und Collision Rate in beiden Anwendungsmodi ausgewertet.

Basierend auf der vorliegenden Datenbank wurde für dieses Szenario folgender Testaufbau verwendet [MSV11a]:

Daten der Session 1. Bestimmung der Referenzdaten für

- Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
- Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)

Daten der Session 2. Ermittlung der Performanzmaße Equal Error Rate (EER), Collision Reproduction Rate (CRR), Reproduction Rate (RR) und Collision Rate (CR) unter schrittweiser Erhöhung (*Schrittweite=0,25*) von Tolerance Factor TF beziehungsweise Expansion Factor EF zur Bestimmung der optimalen Parameter-Sets für

- Biometric Hash
- Secure Sketch

Daten der Session 3. Ermittlung der Performanzmaße Equal Error Rate (EER), Collision Reproduction Rate (CRR), Reproduction Rate (RR) und Collision Rate (CR) unter Verwendung der ermittelten Parameter-Sets für

- Biometric Hash
- Secure Sketch

Die für die beiden Parameter Tolerance Vector beziehungsweise Expansion Factor beobachteten besten Parameter werden bei den folgenden Evaluierungen verwendet, soweit nicht anders angegeben. Zu beachten ist, dass die auf diese Weise ermittelten Werte lediglich für dieses Testset optimiert wurden. Es wird empfohlen, die Nutzbarkeit für andere, beziehungsweise eine

Erweiterung dieses Testsets in zukünftigen Arbeiten zu überprüfen und gegebenenfalls die hier vorgeschlagene Methodik zur Bestimmung passender Parameter zu nutzen.

6.2.3 Setup zu A.3 Analyse und Selektion von Merkmalen

Die folgende Beschreibung der Vorgehensweise bezieht sich jeweils auf die beiden zu untersuchenden Algorithmen unter Verwendung der fünf Semantiken und der fünf Parameter-Sets, die entsprechend Abschnitt 6.2.2 ermittelt wurden, jeweils für Verifikations- und Hash-Generierungsmodus. Für die Analyse der Merkmale werden die Samples der ersten Session für die Referenzdaten verwendet, während die der zweiten Session als Testdaten für die Merkmalsanalyse durch die Wrapper und Filter eingesetzt werden. Die Selektion basierend auf den Wrappern erfolgt durch das schrittweise Erstellen der Merkmal-Sets, wobei nach jedem Schritt eine Verifikation durchgeführt wird, um die aktuelle Performanz zu ermitteln (siehe auch Abschnitt 5.3.2 *Verwendete Wrapper-Methoden*). Die Filter nutzen die Daten aus der zweiten Session, um Wechselwirkungen zwischen den Merkmalen beziehungsweise Nutzern abzuleiten und aus diesen Informationen ein Ranking der Merkmale zu erstellen. Dies geschieht unabhängig vom jeweils verwendeten Verfahren (siehe Abschnitt 5.3.3 *Verwendete Filter-Methoden*).

Nach der Bestimmung der jeweils besten Merkmal-Sets werden diese mittels der aus der ersten Session ermittelten Referenzdaten und der aus der dritten Session gewonnenen Testdaten evaluiert. Dabei werden drei unterschiedliche Szenarien betrachtet:

bestEER. In diesem Szenario bestimmen wir für jedes Verfahren die Anzahl an Merkmalen, die jeweils zur geringsten Equal Error Rate führen.

bestCRR. Hier gehen wir analog für die Betrachtung der Performanz der Hash-Generierung vor, indem wir jeweils die Anzahl der für die geringste Collision Reproduction Rate benötigten Merkmale ermitteln.

fix60. Für das dritte Szenario legen wir die Anzahl der zu betrachtenden Merkmale fest, indem wir aus jedem ermittelten Merkmal-Set die jeweils ersten 60 wählen. Unsere Motivation dabei ist, dass es vor allem im Hash-Modus wichtig sein kann, dass der generierte Hash immer über eine vorgegebene Anzahl von Elementen verfügt. Dies entspricht der von kryptographischen Hash-Funktionen abgeleiteten Eigenschaft der Abbildung einer beliebig großen Datenmenge auf eine mit fixer Länge (siehe beispielsweise [DK07]).

In späteren Arbeiten sollten entsprechend der anvisierten Zielanwendungen sowohl weitere Szenarien mit variabler als auch mit fester Größe des Merkmale-Sets in Abhängigkeit von den Selektionsverfahren untersucht werden. Wir betrachten in jedem der drei Szenarien die Performanzmaße für Equal Error Rate, Collision Reproduction Rate, Reproduction Rate und Collision Rate basierend auf der jeweiligen Anzahl von Merkmalen.

Die Verwendung der Datenbank für die Merkmalsselektion ähnelt der für die Parameterbestimmung. Der Unterschied liegt in der Nutzung der Daten der zweiten Session, welche in diesem

Szenario für die Analyse und Selektion der Merkmale herangezogen werden (siehe Makrushin et al. [MSV11a]):

Daten der Session 1. Bestimmung der Referenzdaten unter Verwendung der ermittelten Parameter für

- Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
- Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)

Daten der Session 2. Durchführung der Merkmalsanalyse basierend auf den in Abschnitt 5.3 *Analyse und Selektion von Merkmalen (A.3)* vorgestellten Verfahren und unter Verwendung der ermittelten Parameter-Sets für

- Biometric Hash
- Secure Sketch

Daten der Session 3. Ermittlung der Performanzmaße Equal Error Rate (EER), Collision Reproduction Rate (CRR), Reproduction Rate (RR) und Collision Rate (CR) unter Verwendung der jeweils selektierten Merkmale und der entsprechenden Parameter-Sets für

- Biometric Hash
- Secure Sketch

Die Daten der ersten Session dienen wieder der Erstellung der Referenz- und Hilfsdaten, die der dritten der Bestimmung der Performanzwerte für die Verifikation und Hash-Generierung.

6.2.4 Setup zu A.4 Biometrische Fusion

Die Nutzung der Daten für die single-modal Fusion auf Matching Score Level findet analog zur Bestimmung der optimalen Parameter und Merkmalsselektion statt. Die Durchführung der Tests verläuft auch hier für die beiden Algorithmen in beiden Anwendungsmodi und für die fünf Semantiken identisch. Auch hier werden die Samples aus Session eins zur Bestimmung der Referenzdaten verwendet. Zusammen mit den aus den Samples der zweiten Session gewonnenen Testdaten werden hier die Parameter für die Fusion auf Matching Score Level bestimmt. Dazu wird jeweils wieder die Equal Error Rate anhand der Session 2 ermittelt und anhand der in Abschnitt 5.4.1 *Bestimmung von Parametern für die Fusion auf Matching Score Level* definierten Fusionsstrategien *gleich gewichtet*, *linear gewichtet* und *quadratisch gewichtet*. Zur Evaluierung der einzelnen Fusionsstrategien und der korrespondierenden Parameter verwenden wir als Referenz die Daten aus Session eins und als Testdaten die Samples aus Session drei. Dabei werden die Fusionsverfahren entsprechend der in den Abschnitten 5.4.2 *Handschriftbasierte multi-algorithmic Fusion auf Matching Score Level*, 5.4.3 *Handschriftbasierte multi-instance Fusion auf Matching Score Level* und 5.4.4 *Handschriftbasierte multi-sample Fusion auf Matching Score Level* beschriebenen eingesetzt. Da die Fusion auf Matching Score Level erst nach dem Vergleich von Referenz- und Testdaten durchgeführt wird, ist an dieser Stelle eine Hash-Generierung nicht möglich. Aus diesem Grund wird die Evaluierung das Parameter-Set *minEER* (siehe Abschnitt

5.2 *Optimierung von Parametern (A.2)*) beschränkt.

Die Verwendung der zur Verfügung stehenden Handschriftendaten für die Strategien der single-modal Fusion sieht zusammenfassend wie folgt aus:

Daten der Session 1. Bestimmung der Referenzdaten unter Verwendung des ermittelten Parameter-Sets für *minEER* für

- Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
- Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)

Daten der Session 2. Bestimmung der Fusionsparameter für die gleich, linear und quadratisch gewichtete Fusionsstrategie unter Verwendung des ermittelten Parameter-Sets für *minEER* für

- Biometric Hash
- Secure Sketch

Daten der Session 3. Ermittlung des Performanzmaßes Equal Error Rate (EER) unter Verwendung des ermittelten Parameter-Sets für *minEER* und der Fusionsparameter für

- Biometric Hash
- Secure Sketch

Die in der Merkmalsselektion identifizierten optimalen Merkmal-Sets werden an dieser Stelle nicht berücksichtigt, es werden für die Fusion immer alle 131 Merkmale verwendet. Die Umsetzung entsprechender Tests wird für zukünftige Arbeiten empfohlen.

Zusammenfassend ist die Verwendung der Daten der drei Sessions für die oben aufgeführten Testszenarien in der Abbildung 6.2 dargestellt. Die Daten der ersten Session werden in den Szenarien zur Ermittlung der Referenzdaten (Biometric Hash: Intervallmatrix IM, Referenz-Hash-Vektor; Secure Sketch: Secure Sketch, Referenz-Merkmalsvektor) verwendet. Die Daten aus Session 2 werden zur Bestimmung der Quantisierungsparameter (Biometric Hash: Tolerance Factor, Secure Sketch: Expansion Factor), der Merkmal-Sets und der Fusionsparameter eingesetzt. Sollten dafür Verifikation beziehungsweise Hash-Generierung notwendig sein (beispielsweise Bestimmung von TF und EF, Einsatz der Wrapper), werden die auf Session 1 basierenden Referenzdaten herangezogen. Entsprechend der durchzuführenden experimentellen Evaluierung setzen wir die Daten der dritten Session zusammen mit den Referenzdaten (Session 1) jeweils im Verifikations- und Hash-Generierungsmodus ein. Dabei fließen je nach untersuchtem Szenario die ermittelten Parameter (TF/EF, Fusion) und Merkmal-Sets ein. Die Ergebnisse der szenarienabhängigen Evaluierungen können dann ausgewertet werden (siehe Kapitel 7 *Präsentation und Diskussion der experimentellen Ergebnisse*).

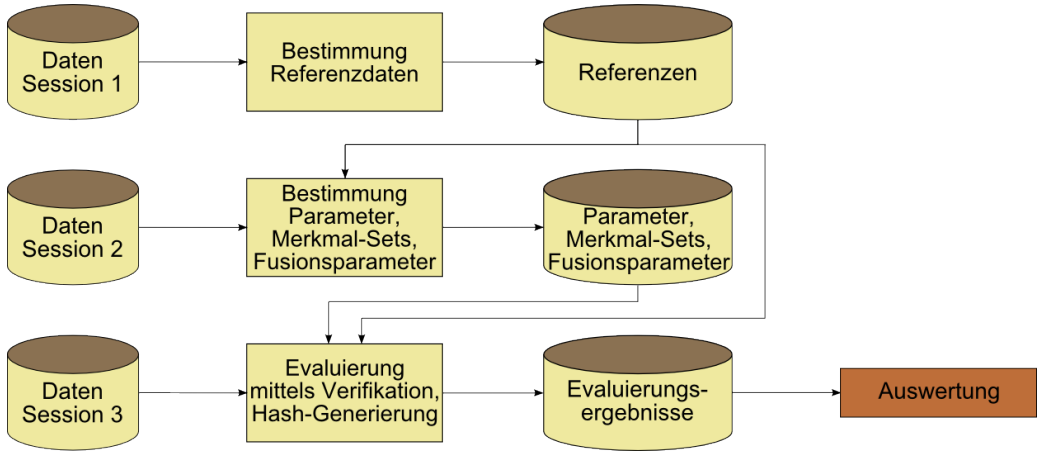


Abbildung 6.2: Nutzung der Daten der drei Sessions für die in den Abschnitten 6.2.1 bis 6.2.4 beschriebenen Aufgaben

6.2.5 Setup zu B Alterung biometrischer Daten

Die Verwendung der Daten der einzelnen Sessions unterscheidet sich bei der Untersuchung von Auswirkungen der Alterung erheblich von den in den Abschnitten 6.2.1 bis 6.2.4 beschriebenen Vorgehensweisen. Hier nutzen wir aus, dass die Daten der einzelnen Sessions zeitlich voneinander getrennt aufgenommen wurden. An dieser Stelle untersuchen wir die Kurzzeitalterung (auch Short Term Aging [SKV12]) basierend auf einem Abstand von einem Monat zwischen den einzelnen Sessions. Trotz des relativ geringen zeitlichen Abstands, ist anzunehmen, dass sich die Performanz von Verifikation beziehungsweise Hash-Generierung verändert, wenn Daten unterschiedlicher Sessions miteinander verglichen werden.

Um untersuchen zu können, ob es entsprechende Auswirkungen gibt, werden die Daten der einzelnen Sessions geteilt. Die jeweils ersten fünf Samples einer Person werden für die Ermittlung der Referenz- und Hilfsdaten genutzt, während die verbleibenden fünf Samples als Testdaten für die Berechnung der Performanzmaße herangezogen werden. Bei der Evaluierung werden die Ergebnisse von Equal Error Rate, Collision Reproduction Rate, Reproduction Rate und Collision Rate basierend auf Daten einer Session denen der zeitlich nachfolgenden Aufnahmezeitpunkte vergleichend gegenüber gestellt. Eine Verwendung der optimierten Parameter- und Merkmal-Sets findet hier nicht statt, da diese separate Testdaten voraussetzen, die in diesem Szenario nicht vorhanden sind. Basierend auf dem beschriebenen Vorgehen werden die zur Verfügung stehenden Daten wie folgt eingesetzt:

Daten der Session 1.

- Erste fünf Samples zur Bestimmung der Referenzdaten (R_1) für
  - Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
  - Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)
- Verbleibende fünf Samples als Testdaten (V_1) unter Verwendung von R_1 zur Ermittlung der Performanzmaße Equal Error Rate EER, Collision Reproduction Rate CRR,

Reproduction Rate RR und Collision Rate CR für

- Biometric Hash
- Secure Sketch

Daten der Session 2.

- Erste fünf Samples zur Bestimmung der Referenzdaten (R_2) für
  - Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
  - Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)
- Verbleibende fünf Samples als Testdaten (V_2) unter Verwendung von R_1 und R_2 zur Ermittlung der Performanzmaße Equal Error Rate EER, Collision Reproduction Rate CRR, Reproduction Rate RR und Collision Rate CR für
  - Biometric Hash
  - Secure Sketch

Daten der Session 3.

- Erste fünf Samples zur Bestimmung der Referenzdaten (R_3) für
  - Biometric Hash (Intervalmatrix IM, Referenz-BioHash)
  - Secure Sketch (Secure Sketch, Referenz-Merkmalsvektor)
- Verbleibende fünf Samples als Testdaten (V_3) unter Verwendung von R_1, R_2 und R_3 zur Ermittlung der Performanzmaße Equal Error Rate EER, Collision Reproduction Rate CRR, Reproduction Rate RR und Collision Rate CR für
  - Biometric Hash
  - Secure Sketch

7 Präsentation und Diskussion der experimentellen Ergebnisse

Im folgenden werden die Ergebnisse der experimentellen Evaluierungen basierend auf den Algorithmen Biometric Hash und Secure Sketch präsentiert. Dabei werden die verschiedenen Szenarien berücksichtigt, die in den vorhergehenden Abschnitten dieses Kapitels beschrieben wurden. Die zugrunde liegenden 131 Merkmale sind unabhängig vom jeweiligen Szenario (Kombinationen aus Semantiken und Hash- beziehungsweise Verifikationsmodus sowie Alterung) oder verwendetem Algorithmus identisch. Die Zusammenstellung der Testdaten basiert auf den im Abschnitt 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank* beschriebenen Konstellationen. Eine Ausnahme bildet das Setup für die Untersuchung der Auswirkung der Alterung auf Verifikation und Hash-Generierung, da hier eine alternative Zusammenstellung der Testdaten für die einzelnen Tests notwendig wurde (siehe Abschnitt 6.2.5 *Setup zu B Alterung biometrischer Daten*). Neben der kurzfristigen Alterung (Abschnitt 7.5) werden die folgenden Szenarien betrachtet: Situation zu Beginn der Arbeit und Optimierung durch neue Merkmale (Abschnitt 7.1), Optimierung durch Anpassung der Parameter für Biometric Hash und Secure Sketch (Abschnitt 7.2), Optimierung durch Merkmalsselektion (Abschnitt 7.3) und Optimierung durch biometrische Fusion (Abschnitt 7.4).

Da die Ergebnisse teilweise sehr umfangreich sind, haben wir uns an einigen Stellen dazu entschieden, lediglich die im Rahmen unserer Arbeit wichtigsten Ergebnisse darzustellen und zu diskutieren. Im Anhang C *Evaluierungsergebnisse* sind alle Ergebnisse sortiert nach dem jeweils zugrunde liegenden Optimierungsansatz tabellarisch angegeben. Wir werden an den entsprechenden Stellen in den einzelnen Abschnitten auf die korrespondierenden Sektionen im Anhang C verweisen.

7.1 Evaluierung: Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale

In diesem Abschnitt wird zum einen die Bestimmung der Erkennungsgenauigkeiten des Ausgangssituation ohne die in dieser Arbeit entwickelten Optimierungsansätze beschrieben. Auf der anderen Seite wird darauf aufbauend der Einfluss zusätzlicher Merkmale auf die Performanz von Verifikation und Hash-Generierung betrachtet. Für den ersten Teil der Evaluierung wurden weder optimale Parameter bestimmt, noch andere Optimierungen vorgenommen. Das bedeutet, die Parameter zur Quantisierung für den Biometric Hash-Algorithmus (Tolerance Factor TF,

Tolerance Vector TV) beziehungsweise für den Secure Sketch-Algorithmus (Expansion Factor EF) werden gleich Null gesetzt.

Zuerst wird für die Tests beider Algorithmen das Merkmal-Set Set A (Merkmale 1-69) aus [Vie06] verwendet, zu dem schrittweise das im Verlauf dieser Arbeit entstandene Merkmal-Set Set B bestehend aus Set B_1 (Merkmale 70-89), Set B_2 (Merkmale 90-103) und Set B_3 (Merkmale 104-131) hinzugefügt wird (vergleiche Abschnitte 5.1 *Hinzufügen zusätzlicher Merkmale (A.1)* und 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank*). Aus diesem Vorgehen ergeben sich jeweils die Ergebnisse für die Ausgangssituation, als auch für das Hinzufügen zusätzlicher Merkmale als erstem Optimierungsschritt. Aufgrund der großen Anzahl von Merkmalen beschränken wir uns an dieser Stelle auf die detaillierte Betrachtung der Ergebnisse für die schrittweise konkatenierten Merkmal-Sets und eine tendenzielle Sicht auf den Verlauf von Equal Error Rate und Collision Reproduction Rate bei zunehmender Menge von Merkmalen. Die zur Verfügung stehenden Daten werden entsprechend der Angaben in 6.2.1 *Setup zu Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale* vorgenommen. Dabei dienen die Daten der ersten Session zur Erstellung der Referenzen während die aus Session 3 als Testdaten zur experimentellen Evaluierung eingesetzt werden.

7.1.1 Testergebnisse für den Biometric Hash-Algorithmus

In Tabelle 7.1 sind spaltenweise die Ergebnisse für EER, CRR, RR und CR für die jeweiligen Merkmal-Sets angegeben. Die einzelnen Zeilen zeigen die Resultate für Set A aus [Vie06], das schrittweise um Set B_1, Set B_2 und Set B_3 (siehe Abschnitt 6.1 *Evaluierungsmethodologie und Aufbau der Datenbank*) erweitert wird.

Tabelle 7.1: Biometric Hash: Ergebnisse entsprechend der konkatenierten Merkmal-Sets

	gegebene PIN					geheime PIN				
version	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,22348	0,50000	0,00000	0,00000	69	0,16544	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	0,20206	0,50000	0,00000	0,00000	89	0,14167	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	0,19732	0,50000	0,00000	0,00000	103	**0,13002**	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	**0,18536**	0,50000	0,00000	0,00000	131	0,13036	0,50000	0,00000	0,00000

	Pseudonym					Symbol				
version	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,11151	0,50000	0,00000	0,00000	69	0,16311	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	0,11351	0,50000	0,00000	0,00000	89	0,12560	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	**0,10722**	0,50000	0,00000	0,00000	103	0,12604	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	0,11338	0,50000	0,00000	0,00000	131	**0,12512**	0,50000	0,00000	0,00000

	Woher				
version	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,11861	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	0,09977	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	0,10251	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	**0,09528**	0,50000	0,00000	0,00000

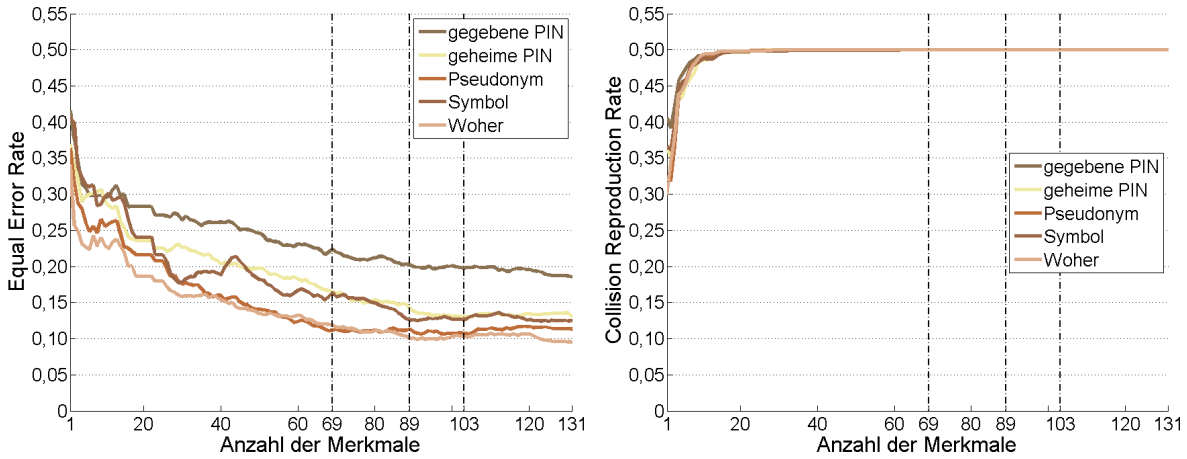
Dabei lässt sich feststellen, dass einerseits das Hinzufügen zusätzlicher Merkmale zur Ausgangsmenge Set A immer zu besseren Verifikationsergebnissen führt. Auf der anderen Seite wird das

jeweils beste Ergebnis nicht immer durch die Verwendung des größten Merkmal-Sets bestimmt. Dies ist für Set B_3 zu beobachten, bei dem sich für die Semantiken *geheime PIN* und *Pseudonym* die EER verschlechtert. Die besten Ergebnisse im Hinblick auf die Verifikationsperformanz für die fünf Semantiken sind 0,18536 (*gegebene PIN*, 131 Merkmale), 0,13002 (*geheime PIN*, 103 Merkmale), 0,10722 (*Pseudonym*, 103 Merkmale), 0,12512 (*Symbol*, 131 Merkmale) und 0,09528 (*Woher*, 131 Merkmale). Damit wird das beste Verifikationsergebnis für die Semantik *Woher* unter Verwendung aller Merkmale bestimmt. Die höchste EER wird bei diesem Vergleich durch die Semantik *gegebene PIN* generiert.

In den Abbildungen 7.1a, 7.1b, 7.2a und 7.2b werden die erreichten Werte für die Equal Error Rate beziehungsweise die Collision Reproduction Rate in Abhängigkeit von der Anzahl der eingesetzten Merkmale jeweils für die fünf Semantiken dargestellt. Die Reihenfolge der Merkmale entspricht der des Hinzufügens (vergleiche auch A *Statistische Merkmale*). Die Bereiche, die durch die einzelnen Merkmal-Sets Set A (Merkmale 1-69), Set B_1 (Merkmale 70-89), Set B_2 (Merkmale 90-103) und Set B_3 (Merkmale 104-131) gebildet werden, sind jeweils durch senkrechte gestrichelte Linien markiert.

Abbildung 7.1a zeigt einerseits, dass abgesehen von einigen Ausreißern die Equal Error Rate mit zunehmender Anzahl von Merkmalen bei allen fünf Semantiken deutlich sinkt. Andererseits ist zu erkennen, dass die Semantiken mit einem höheren Anteil an Individualität zu einer geringeren Equal Error Rate führen. Dabei kann es sich beispielsweise um individuellen geheimen Inhalt oder die Möglichkeit zur Verwendung eines umfangreicheren Zeichenvorrats handeln.

In Tabelle 7.1 ist für die Hash-Generierung deutlich zu sehen, dass das reine Hinzufügen weiterer Merkmal-Sets sowohl auf Reproduktion individueller Hashes als auch auf die Kollision keinen messbaren Einfluss hat. Die entsprechenden Ergebnisse bleiben in allen betrachteten Fällen bei CRR=0,5 basierend auf RR=0 und CR=0.



(a) Equal Error Rate

(b) Collision Reproduction Rate

Abbildung 7.1: Biometric Hash: Darstellung des Verlaufs der (a) Equal Error Rate und der (b) Collision Reproduction Rate für die fünf Semantiken

Die Betrachtung des Verlaufs der Collision Reproduction Rate über alle 131 Merkmale in Abbildung 7.1b bestätigt, dass sich die CRR noch bei alleiniger Verwendung von Set A bei allen Semantiken den Wert 0,5 erreicht. Die Anzahl der eingesetzten Merkmale für das Erreichen von CRR=0,5 (basierend auf RR=0, CR=0) beträgt dabei für die einzelnen Semantiken 63 (*gegebene PIN*), 27 (*geheime PIN*), 34 (*Pseudonym*), 63 (*Symbol*) und 30 (*Woher*). Dies motiviert die Untersuchung anderer Möglichkeiten zur Senkung der Collision Reproduction Rate mit dem Ziel, eine hohe Reproduction Rate bei geringer Collision Rate zu erreichen.

7.1.2 Testergebnisse für den Secure Sketch-Algorithmus

Bei der Betrachtung der Ergebnisse für die Equal Error Rate EER basierend auf dem Secure Sketch-Algorithmus fällt auf, dass auch hier die besten Verifikationsergebnisse für unterschiedliche Merkmalsmengen generiert werden (siehe Tabelle 7.2). Allein die *gegebene PIN* benötigt alle 131 Merkmale für das beste Ergebnis (0,20630) während für die *geheime PIN* und das *Pseudonym* 103 Merkmale notwendig sind, um die jeweils geringste Equal Error Rate (0,15902 beziehungsweise 0,15316) zu bestimmen. Lediglich die ersten 89 Merkmale werden für *Symbol* und *Woher* benötigt, um das im Vergleich beste Ergebnis zu ermitteln (0,12141 beziehungsweise 0,12048). Bei der Gegenüberstellung der besten Ergebnisse der Semantiken untereinander wird auch hier die niedrigste EER mittels Semantik *Woher* bestimmt, wohingegen auch für den Secure Sketch-Algorithmus die Semantik *gegebene PIN* am schlechtesten abschneidet.

Tabelle 7.2: Secure Sketch: Ergebnisse entsprechend der konkatenierten Merkmal-Sets

	gegebene PIN					geheime PIN				
version	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,25332	0,50000	0,00000	0,00000	69	0,19716	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	0,22683	0,50000	0,00000	0,00000	89	0,16035	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	0,21726	0,50000	0,00000	0,00000	103	**0,15902**	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	**0,20630**	0,50000	0,00000	0,00000	131	0,16013	0,50000	0,00000	0,00000

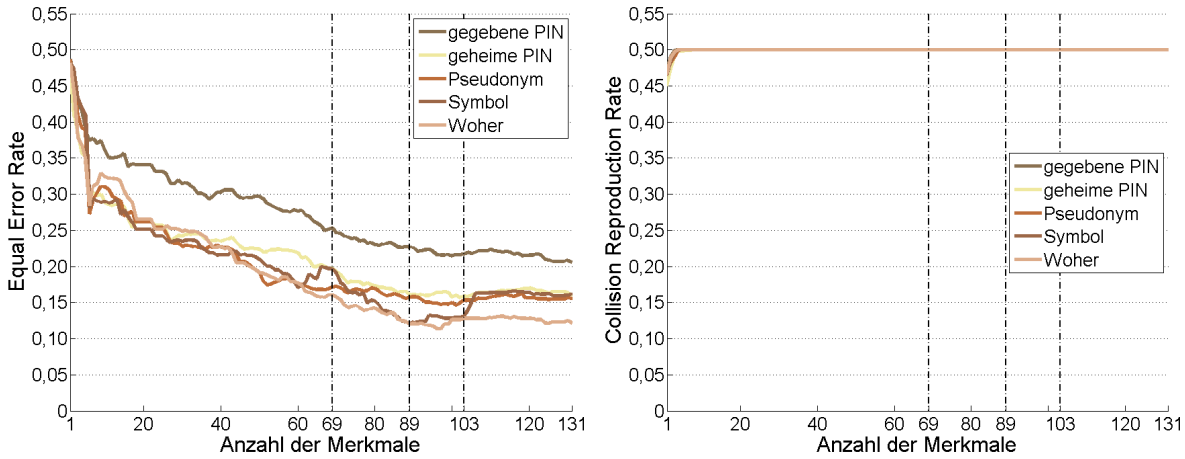
	Pseudonym					Symbol				
version	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,17112	0,50000	0,00000	0,00000	69	0,19680	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	0,15771	0,50000	0,00000	0,00000	89	**0,12141**	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	**0,15316**	0,50000	0,00000	0,00000	103	0,13048	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	0,15539	0,50000	0,00000	0,00000	131	0,16158	0,50000	0,00000	0,00000

	Woher				
version	**n**	**EER**	**CRR**	**RR**	**CR**
Set A (1-69)	69	0,16011	0,50000	0,00000	0,00000
∪ Set B_1 (1-89)	89	**0,12048**	0,50000	0,00000	0,00000
∪ Set B_2 (1-103)	103	0,12826	0,50000	0,00000	0,00000
∪ Set B_3 (1-131)	131	0,12147	0,50000	0,00000	0,00000

In Abbildung 7.2a fällt bei der Betrachtung des Verlaufs der Equal Error Rate über die Anzahl verwendeter Merkmale ebenfalls auf, dass mit größer werdender Menge die EER sinkt. Wie beim Biometric Hash-Algorithmus gibt es einige Ausreißer, aber die Tendenz ist auch hier deutlich sichtbar. Tabelle 7.2 zeigt für die Hash-Generierung, dass auch für den Secure Sketch-Algorithmus bei der Konkatenation der einzelnen Merkmal-Sets weder Hashes für die identischen Nutzer

reproduziert werden können, noch treten Kollisionen auf. Daher betragen auch in diesem Fall die Werte CRR=0,5, RR=0 und CR=0.

Betrachtet man im Vergleich dazu den Verlauf der Collision Reproduction Rate in Abbildung 7.2b ist zu erkennen, dass für die fünf Semantiken der Wert CRR=0,5 bereits mit jeweils wenigen Merkmalen erreicht ist. Dafür reichen 5 (*gegebene PIN* und *Woher*), 9 (*geheime PIN*), 6 (*Pseudonym*) und 8 (*Symbol*) Merkmale.



(a) Equal Error Rate

(b) Collision Reproduction Rate

Abbildung 7.2: Secure Sketch: Darstellung des Verlaufs der (a) Equal Error Rate und der (b) Collision Reproduction Rate für die fünf Semantiken

7.1.3 Fazit - Evaluierung: Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale

Die basierend auf den Merkmalen des Set A ermittelten Ergebnisse zeigen, dass die Performanz beider Algorithmen für ein biometrisches System zur Verifikation und zur Hash-Generierung nicht zufriedenstellend sind. Für letztere wird in jedem untersuchten Fall sowohl für die Reproduction Rate als auch die Collision Rate ein Wert von 0 ermittelt. Dies führt zu einer Collision Reproduction Rate von 0,5. Der beste Wert für den Biometric Hash-Algotithmus im Verifikationsmodus wird für die Semantik *Pseudonym* mit einer Equal Error Rate von 0,11151 bestimmt. Beim Secure Sketch-Algorithmus ergibt sich die geringste EER für die Semantik *Woher* und beträgt für Set A 0,16011. Für beide Verfahren scheint die *gegebene PIN* die ungeeignetste Semantik zu sein, da hier wesentlich schlechtere Ergebnisse berechnet werden als für die übrigen Schreibinhalte (Biometric Hash: EER=0,22348, Secure Sketch: EER=0,25332). Dieses Verhalten setzt sich auch bei den weiteren Untersuchungen fort und entspricht auch den Beobachtungen aus früheren Arbeiten. Ursache dafür ist die Verwendung eines begrenzten Vorrats an Zeichen (0-9), die dazu noch in der Konstellation *77993* in der Abfolge und Ziffernanzahl für jeden Schreiber identisch ist. Weicht man die vorgegebenen Bedingungen auf, indem jedem Probanden bei gleicher Anzahl die Wahl der Ziffern freigestellt wird (siehe Ergebnisse *geheime PIN*), verbessern sich die Ergebnisse zum Teil signifikant.

Deutlich wird bei der Betrachtung der Ergebnisse für die Equal Error Rate, dass für beide Verfahren die Verifikationsperformanz mit dem schrittweisen Konkatenieren der Merkmal-Sets in den meisten untersuchten Fällen zunimmt. Jedoch ist basierend auf unserer Methodologie erkennbar, dass dies nicht nur von den verwendeten Merkmal-Sets, sondern auch vom Algorithmus und von der Semantik abhängig ist. Ebenfalls lässt die Beobachtung des Verhaltens des Verlaufs der jeweilig ermittelten EERs den Schluss zu, dass die Performanz von den einzelnen Merkmalen abhängig ist. So kann das Hinzufügen sowohl einzelner Merkmale als auch von Gruppen von Merkmalen zu einem unerwarteten Anstieg beziehungsweise Abfall der beobachteten Equal Error Rate führen (siehe Abbildungen 7.1 und 7.2). Damit verdeutlichen die Verläufe der EER-Kurven das Potential, durch Analyse und darauf aufbauender Selektion bestimmter Merkmale zu besseren Ergebnissen zu gelangen, beziehungsweise die aktuelle Performanz nicht wesentlich zu verschlechtern.

Die Betrachtung der Ergebnisse für Equal Error Rate und Collision Reproduction Rate motiviert ebenfalls eine Anpassung der jeweiligen Parameter, die für die individuelle Einstellung der Quantisierung zuständig sind. Dadurch würde das erfolgreiche Mapping der statistischen Merkmale für die einzelnen Nutzer verbessert werden. Dabei ist darauf zu achten, dass dadurch auch die Trennschärfe zwischen den einzelnen Personen geringer wird, wodurch die Wahrscheinlichkeit von fälschlichen Annahmen beziehungsweise von Kollisionen steigt.

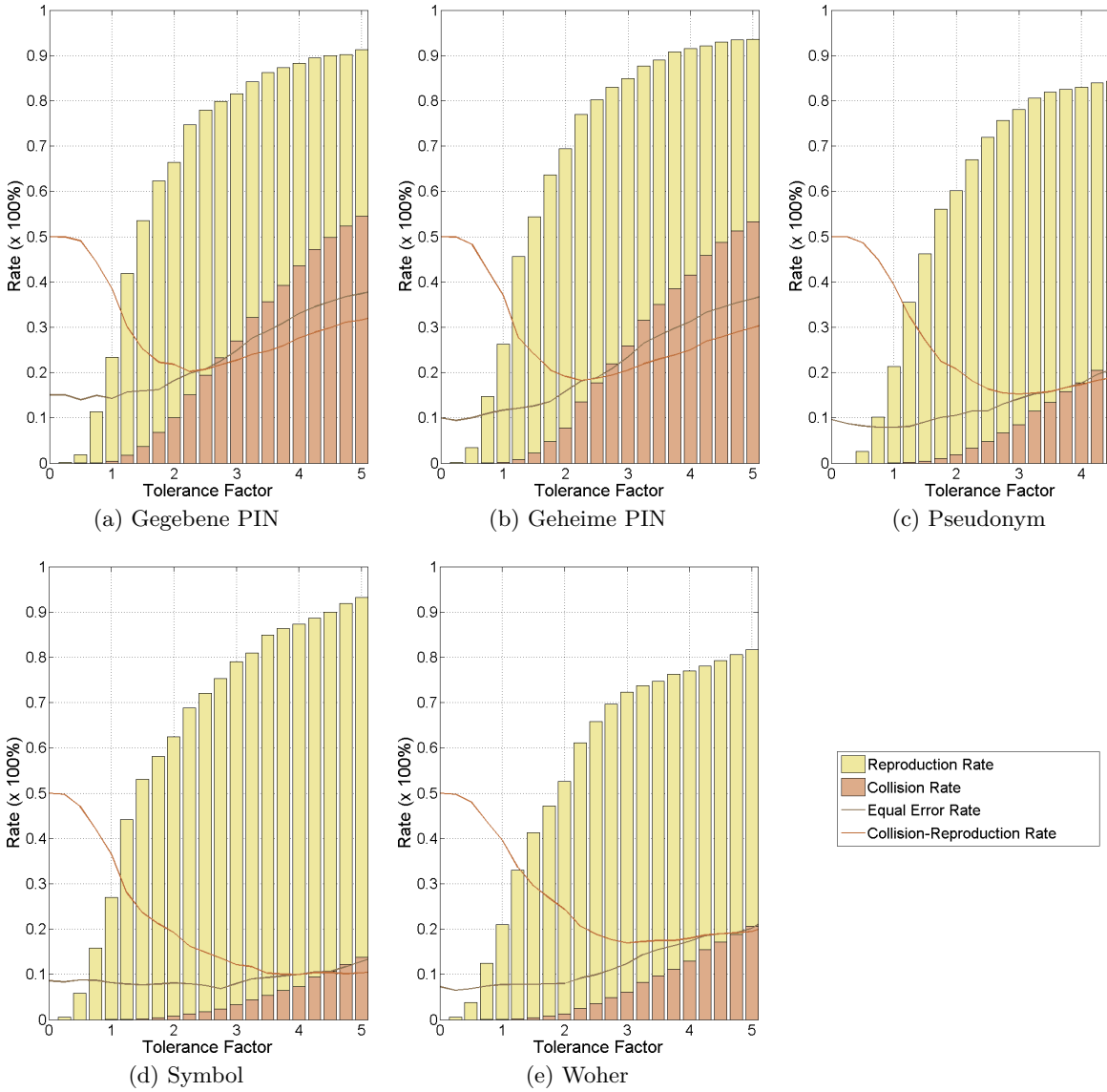
7.2 Evaluierung: A.2 Optimierung von Parametern

Für die Optimierung der Parameter haben wir den Tolerance Factor für den Biometric Hash- und den Expansion Factor für den Secure Sketch-Algorithmus gewählt. Beide ermöglichen mit einem einzelnen Wert die Anpassung der jeweiligen algorithmeninternen Quantisierung (siehe dazu Abschnitte 4.1 *Biometric Hash-Algorithmus für die dynamische Handschrift* und 4.2 *Secure Sketch-Algorithmus für die dynamische Handschrift*). Nachfolgend wird die Bestimmung von Tolerance Factor und Expansion Factor in Abhängigkeit zu den fünf Semantiken und den in Abschnitt 5.2 *Optimierung von Parametern (A.2)* beschriebenen Parameter-Sets erläutert und durchgeführt. Entsprechend der Angaben in 6.2.2 *Setup zu A.2 Optimierung von Parametern* werden die Daten aus Session 1 zur Erzeugung der Referenzdaten verwendet. Die Bestimmung der jeweils fünf Parameter-Sets pro Algorithmus erfolgt mittels der Daten aus Session 2. Daran schließt sich die experimentelle Evaluierung der Parameterwerte mit beiden Algorithmen an, welche die Samples aus der dritten Session als Testdaten nutzt.

7.2.1 Bestimmung der Parameter

Für die Optimierung der Parameter für Biometric Hash (Tolerance Factor TF) und Secure Sketch (Expansion Factor EF) werden beide Werte schrittweise erhöht. Die Schrittweite beträgt jeweils 0,25. Anschließend erfolgt für die jeweiligen Werte die Bestimmung der Fehlerraten EER, CRR, RR und CR (vergleiche Abschnitt 6.2.2). Diese Vorgehensweise verdeutlichen die Abbildungen 7.3 und 7.4 für den Biometric Hash beziehungsweise Secure Sketch. Die einzelnen Diagramme zeigen

die vier Raten in Abhängigkeit zum jeweiligen Parameter (TF in Abbildung 7.3, EF in Abbildung 7.4). Dabei ist für beide Algorithmen der Einfluss des jeweiligen Parameter auf die Verifikations- und die Hash-Generierungsperformanz zu erkennen. Ziel bei der Parameterbestimmung ist, im betrachteten Bereich einen Wert für TF beziehungsweise EF basierend auf dem Verlauf von EER, CRR, RR und CR zu finden. Für beide Verfahren werden entsprechend der Beschreibung in 5.2 *Optimierung von Parametern (A.2)* die Parameter für jeweils fünf markante Szenarien gewählt: Parameterwert gleich 0 (*Null*, keine Parameteroptimierung), minimale EER (*minEER*), minimale CRR (*minCRR*), RR größer gleich a% ($RR \geq a\%$) und CR kleiner gleich b% ($CR \leq b\%$). Die daurch für jeden der zwei Algorithmen entstehenden fünf Parameter-Sets enthalten jeweils einen Wert pro Semantik.



(a) Gegebene PIN (b) Geheime PIN (c) Pseudonym

(d) Symbol (e) Woher

Abbildung 7.3: Biometric Hash: Darstellung der Vorgehensweise bei der Bestimmung des optimalen Tolerance Factor für die fünf Semantiken

Bei der Ermittlung des Tolerance Factor werden für die fünf Semantiken das Intervall $[0, 5]$ und

als obere Grenze für die Reproduction Rate 80% ($RR \geq 80\%$) verwendet. Dieser Wert wird für alle fünf Semantiken innerhalb des untersuchten Intervalls erreicht. Für die Bestimmung des Parameters Expansion Factor des Secure Sketch-Algorithmus wird das Intervall auf $[0, 13]$ beziehungsweise $[0, 30]$ für das *Symbol* vergrößert, da hier die Raten deutlich langsamer ansteigen als beim Biometric Hash. Dieses Vorgehen ermöglicht die Bestimmung des Expansion Factor in allen zu untersuchenden Szenarien. Aus dem selben Grund wurde das Kriterium für die Reproduction Rate auf 70% gesetzt ($RR \geq 70\%$). Eine RR von mindestens 80% wird im vorgegebenen Bereich nur in zwei Fällen erreicht (*gegebene PIN*, *geheime PIN*). Die Abhängigkeit der untersuchten Maße EER, RR, CR und CRR von den Parametern TF und EF ist in den Abbildungen 7.3 und 7.4 deutlich zu sehen.

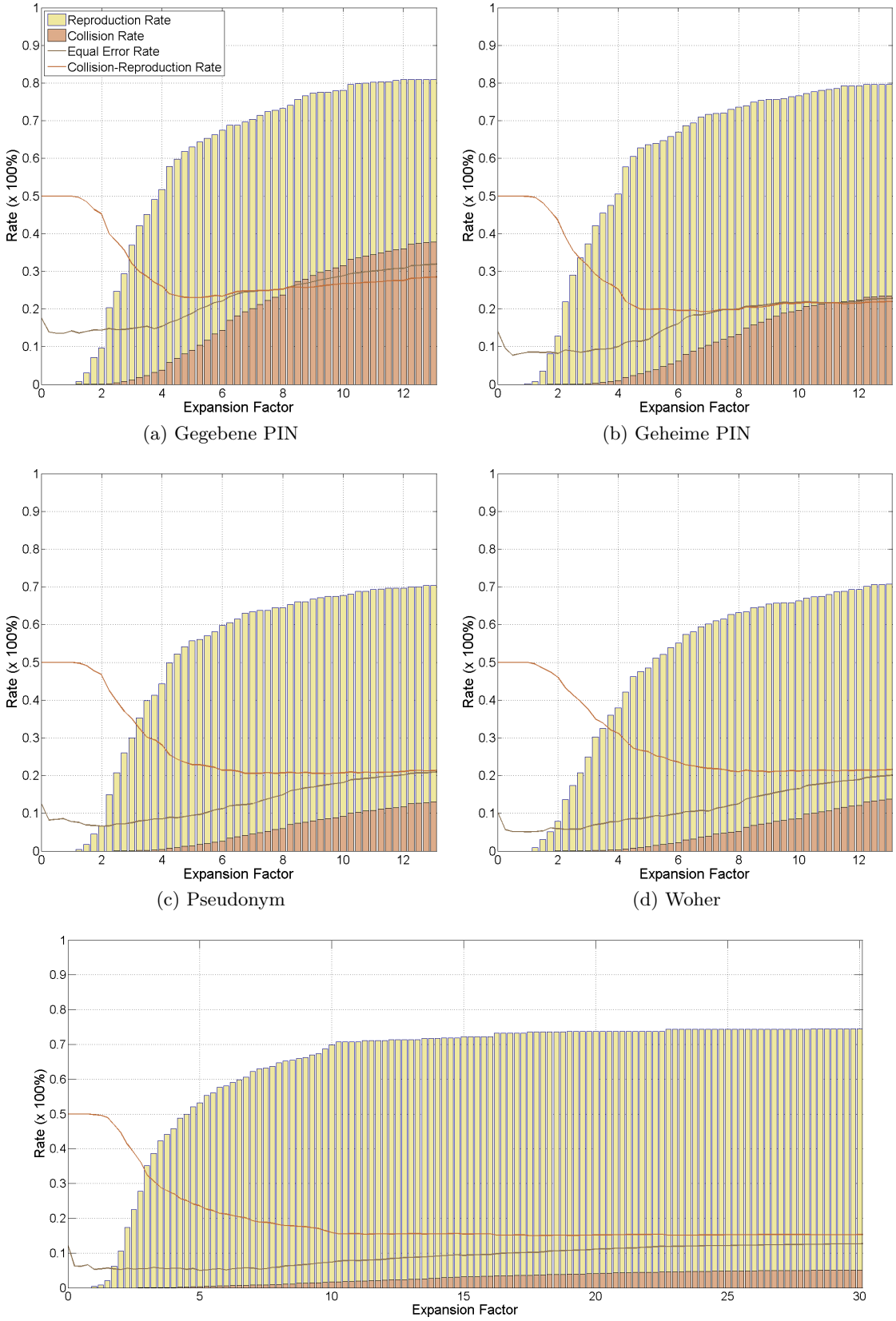
Tabelle 7.3: Biometric Hash: Optimaler Tolerance Factor für die fünf Semantiken entsprechend der Verwendung im Verifikations- beziehungsweise Hash-Generierungsmodus

Semantik	**Verifikation**	**Hash-Generierung**		
	minEER	*minCRR*	$RR \geq 80\%$	$CR \leq 5\%$
Gegebene PIN	0.50	2.25	3.00	1.50
Geheime PIN	0.25	2.25	2.50	1.75
Pseudonym	0.75	3.00	3.25	2.50
Symbol	2.75	4.00	3.25	3.25
Woher	0.25	3.00	4.75	2.75

Basierend auf den Szenarien werden abhängig von der Semantik die entsprechenden Werte für die Parameter von Biometric Hash- und Secure Sketch-Algorithmus gewählt. Diese werden für die Evaluierungen genutzt und sind in den Tabellen 7.3 und 7.4 zusammenfassend aufgelistet. Zu sehen ist, dass die Parameter in Abhängigkeit von der jeweiligen Semantik als auch der einzelnen Parameter-Sets zum Teil signifikant abweichen. Besonders deutlich stellt sich dies für den Secure Sketch dar. Hier variiert EF beispielsweise beim *Symbol* sehr stark von 5 ($minEER$) bis 28 ($CR \leq 5\%$).

Tabelle 7.4: Secure Sketch: Optimaler Expansion Factor für die fünf Semantiken entsprechend der Verwendung im Verifikations- beziehungsweise Hash-Generierungsmodus

Semantik	**Verifikation**	**Hash-Generierung**		
	minEER	*minCRR*	$RR \geq 70\%$	$CR \leq 5\%$
Gegebene PIN	0.75	5.25	10.75	4.00
Geheime PIN	0.50	6.75	6.75	5.50
Pseudonym	2.00	9.50	12.75	7.25
Symbol	5.00	17.50	10.25	28.00
Woher	1.00	9.00	12.25	7.75



(a) Gegebene PIN

(b) Geheime PIN

(c) Pseudonym

(d) Woher

(e) Symbol

Abbildung 7.4: Secure Sketch: Darstellung der Vorgehensweise bei der Bestimmung des optimalen Expansion Factor für die fünf Semantiken

7.2.2 Testergebnisse für den Biometric Hash-Algorithmus

Die folgenden zwei Tabellen zeigen die Ergebnisse, die basierend auf der Optimierung der Parameter mittels Biometric Hash (Tabelle 7.5) beziehungsweise Secure Sketch (Tabelle 7.6) erzielt werden. Betrachtet werden dabei Verifikations- als auch Hash-Generierungsperformanz für die fünf Semantiken und die fünf Parameter-Sets. Dabei werden die Ergebnisse der Parameter-Sets zeilenweise, die Semantiken spaltenweise angegeben. Die Ausgangssituation entsprechend den Ergebnissen aus Abschnitt 7.1 *Evaluierung: Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale* wird dabei jeweils durch das Parameter-Set *Null* in den Tabellen dargestellt.

Für den Biometric Hash zeigt Tabelle 7.5 die Ergebnisse basierend auf allen 131 Merkmalen. Für die EER stellen wir fest, dass für alle Semantiken bis auf die *gegebene PIN* die besten Ergebnisse mittels Parameter-Set *minEER* berechnet werden. Die geringste EER wird dabei mit 0,07366 für die Semantik *Symbol* bestimmt. Damit liegt hier die EER circa 41% unter der ohne Parameteroptimierung (*Null*) bestimmten. Mit 0,09048 wird die beste CRR ebenfalls für das *Symbol* mit dem Parameter-Set *minCRR* gemessen, was einer Verbesserung um knapp 82% gegenüber der nicht parametrisierten Variante entspricht, deren Wert für alle fünf Semantiken 0,5 beträgt. Dabei wird eine Reproduction Rate von 0,89623 bei einer Collision Rate von 0,07718 erreicht. Weiterhin lässt sich feststellen, dass die für den Tolerance Factor ermittelten Parameter-Sets nicht in jedem Fall dieselbe Wirkung zeigen, wie für das zu ihrer Bestimmung verwendete Testset. So werden die besten Ergebnisse für EER und CRR nicht immer mittels der Parametrisierungen *minEER* beziehungsweise *minCRR* bestimmt. Für die Equal Error Rate funktioniert dies aber in vier von fünf Fällen. Ausnahme ist hier die *gegebene PIN* basierend auf einem Tolerance Factor von 0.

7.2.3 Testergebnisse für den Secure Sketch-Algorithmus

Die Tabelle 7.6 stellt die entsprechenden Ergebnisse basierend auf dem Secure Sketch-Algorithmus für die 131 Merkmale dar und welche demzufolge mit dem Parameter Expansion Factor berechnet wurden. Hier werden die im Vergleich besten Ergebnisse für die EER jeweils basierend auf dem Parameterwert *minEER* berechnet. Dabei kann die kleinste EER für die Semantik *Woher* ermittelt werden, sie beträgt 0,06817. Am schlechtesten schneidet die *gegebene PIN* mit einer EER von 0,22926 ab. Im direkten Vergleich zu den Werten, welche ohne Parametrisierung bestimmt wurden, bietet die Semantik *Symbol* die höchste Verbesserung. Hier verringert sich die EER von 0,16013 auf 0,07259, das entspricht einer Verbesserung von 55%.

Im Hash-Generierungsmodus werden für vier Semantiken die besten Werte für die CRR mittels Parameter-Set *minCRR* berechnet. Die Ausnahme bildet das *Symbol* mit einer CRR von 0,17311, ermittelt basierend auf $CR \leq 5\%$. Dieser Wert ist auch die kleinste Collision Reproduction Rate bei der Gegenüberstellung der fünf Semantiken. Die ebenfalls damit erreichte höchste Verbesserung verglichen mit der Parameterisierung *Null* beträgt circa 65%. Dem liegt eine Reproduction Rate von 0,70566 bei einer Collision Rate von 0,05189 zugrunde.

Tabelle 7.5: Biometric Hash: Ergebnisse aus Verifikations- beziehungsweise Hash-Generierungsmodus basierend auf den fünf Parameter-Sets für alle 131 Merkmale

	gegebene PIN				geheime PIN			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**	**EER**	**CRR**	**RR**	**CR**
Null	**0,18536**	0,50000	0,00000	0,00000	0,13036	0,50000	0,00000	0,00000
minEER	0,20082	0,49253	0,01509	0,00015	**0,12790**	0,49906	0,00189	0,00000
minCRR	0,22124	**0,22206**	0,73019	0,17431	0,20295	0,20758	0,72830	0,14347
$RR \geq 80\%$	0,27398	0,25212	0,79623	0,30047	0,20620	**0,20475**	0,77736	0,18687
$CR \leq 5\%$	0,19783	0,28759	0,46604	0,04122	0,16168	0,23169	0,58868	0,05207

	Pseudonym				Symbol			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**	**EER**	**CRR**	**RR**	**CR**
Null	0,11338	0,50000	0,00000	0,00000	0,12512	0,50000	0,00000	0,00000
minEER	**0,08943**	0,45004	0,10000	0,00007	**0,07366**	0,13610	0,75472	0,02692
minCRR	0,15291	**0,17099**	0,73962	0,08160	0,09259	**0,09048**	0,89623	0,07718
$RR \geq 80\%$	0,16874	0,17698	0,75849	0,11245	0,08866	0,10825	0,82830	0,04481
$CR \leq 5\%$	0,12517	0,18101	0,68302	0,04503	0,08866	0,10825	0,82830	0,04481

	Woher			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**
Null	0,09528	0,50000	0,00000	0,00000
minEER	**0,08559**	0,50000	0,00000	0,00000
minCRR	0,12157	**0,15688**	0,74528	0,05904
$RR \geq 80\%$	0,18363	0,16950	0,85283	0,19184
$CR \leq 5\%$	0,10910	0,16272	0,72264	0,04808

Tabelle 7.6: Secure Sketch: Ergebnisse aus Verifikations- beziehungsweise Hash-Generierungsmodus basierend auf den fünf Parametersets für alle 131 Merkmale

	gegebene PIN				geheime PIN			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**	**EER**	**CRR**	**RR**	**CR**
Null	0,20630	0,50000	0,00000	0,00000	0,16013	0,50000	0,00000	0,00000
minEER	**0,16211**	0,50000	0,00000	0,00000	**0,10927**	0,50000	0,00000	0,00000
minCRR	0,22595	**0,25789**	0,60566	0,12144	0,19485	**0,19726**	0,70943	0,10396
$RR \geq 70\%$	0,32181	0,30517	0,74151	0,35185	0,19485	**0,19726**	0,70943	0,10396
$CR \leq 5\%$	0,18792	0,29915	0,44717	0,04546	0,15617	0,20905	0,63396	0,05207

	Pseudonym				Symbol			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**	**EER**	**CRR**	**RR**	**CR**
Null	0,15539	0,50000	0,00000	0,00000	0,16158	0,50000	0,00000	0,00000
minEER	**0,08757**	0,47170	0,05660	0,00000	**0,07259**	0,26392	0,47547	0,00330
minCRR	0,18061	**0,20708**	0,67547	0,08962	0,10021	0,17393	0,69057	0,03843
$RR \geq 70\%$	0,20777	0,21214	0,70943	0,13371	0,08165	0,17866	0,66226	0,01959
$CR \leq 5\%$	0,13610	0,21179	0,62642	0,05000	0,12516	**0,17311**	0,70566	0,05189

	Woher			
Parameter-Set	**EER**	**CRR**	**RR**	**CR**
Null	0,12147	0,50000	0,00000	0,00000
minEER	**0,06817**	0,50000	0,00000	0,00000
minCRR	0,15270	**0,19282**	0,68868	0,07431
$RR \geq 70\%$	0,19986	0,21190	0,71132	0,13512
$CR \leq 5\%$	0,12367	0,20468	0,63962	0,04898

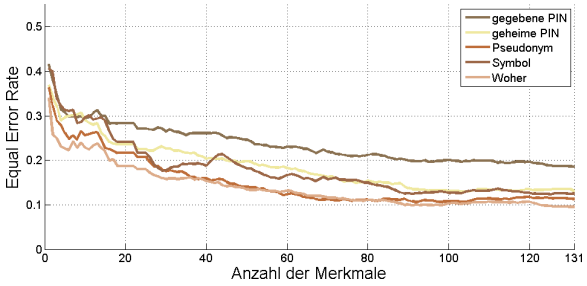
7.2.4 Fazit - Evaluierung: A.2 Optimierung von Parametern

Sowohl beim Biometric Hash- als auch beim Secure Sketch-Algorithmus kam es durch die Anpassung des Parameters Tolerance Factor beziehungsweise Expansion Factor zu einer teilweise erheblichen Verbesserung der Ergebnisse für Verifikation und Hash-Generierung. Im Vergleich beider Verfahren basierend auf allen 131 Merkmalen bezüglich Verifikationsperformanz schneidet der Secure Sketch für alle fünf Semantiken am besten ab (siehe Tabellen 7.5 und 7.6). Die geringste erreichte Equal Error Rate beträgt hier 0,06817 für die Semantik *Woher* bei der Parametrisierung *minEER*. Ohne Parameteroptimierung liegt die EER für die gleiche Semantik bei 0,12147. Das entspricht einer Verbesserung um circa 44%. Für dieselbe Semantik bestimmt der Biometric Hash eine EER von 0,08559 (*minEER*) beziehungsweise 0,09528 (*Null*).

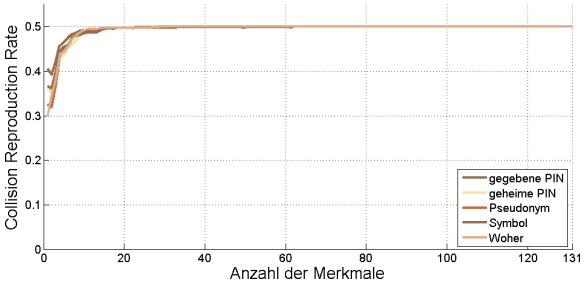
Anders sieht es im Hash-Generierungsmodus für beide Algorithmen aus. Hier wird bei Verwendung von 131 Merkmalen für vier von fünf Semantiken jeweils die beste Collision Reproduction Rate mittels Biometric Hash ermittelt. Dabei beträgt die beste CRR 0,09048 für das *Symbol* basierend auf Parameter-Set *minCRR*. Demgegenüber ist die beste CRR, welche durch den Secure Sketch für das *Symbol* ermittelt wird mit 0,17311 ($CR \leq 5\%$) fast doppelt so hoch. Die Ausnahme bildet die *geheime PIN*, für die der Secure Sketch eine CRR von 0,19726 ($minCRR/RR \geq 70\%$) berechnet und der Biometric Hash mit 0,20475 etwas schlechter abschneidet ($RR \geq 80\%$).

Wenn wir den Verlauf von Equal Error Rate beziehungsweise Collision Reproduction Rate in den Abbildungen 7.5 (Biometric Hash) und 7.6 (Secure Sketch) betrachten, wird deutlich, dass die Veränderung der Parameter beider Algorithmen zu teilweise signifikanten Verbesserungen führen. Im Vergleich zu den Ergebnissen ohne Parametrisierung (7.5a, 7.5b, 7.6a und 7.6b) kann durch die Anpassung dieser Werte zum Teil sowohl die Equal Error Rate als auch die Collision Reproduction Rate verringert werden. Besonders fällt das bei der Collision Reproduction Rate auf. Hier verändert sich der Verlauf der Kurven deutlich, vor allem für die Parameter-Sets, die auf der Verbesserung der CRR basieren (*minCRR*, $RR \geq 80\%/RR \geq 70\%$, $CR \leq 5\%$). Diese bleibt hier über den gesamten Verlauf weit unterhalb der 0,5, welche in der Ausgangssituation (7.5b/ 7.6b) schon mit sehr wenig Merkmalen erreicht wird.

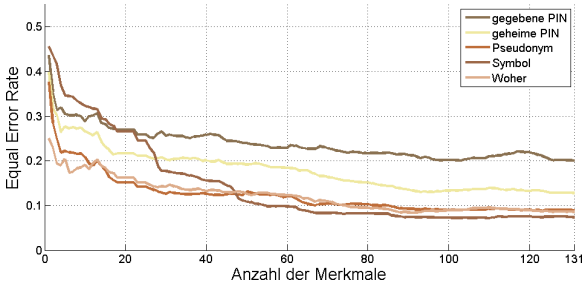
Betrachten wir den Erfolg der Parameter-Sets stellen wir fest, dass bei der Verwendung von 131 Merkmalen die jeweils beste Equal Error Rate in neun von zehn Fällen über die Parametrisierung *minEER* bestimmt wurde. Dabei entfallen vier auf den Biometric Hash und fünf auf den Secure Sketch. Auf der anderen Seite basieren die besten Ergebnisse für die Collision Reproduction Rate auf den Parametern zur Optimierung der Hash-Generierung (*minCRR*, $RR \geq 80\%/RR \geq 70\%$ und $CR \leq 5\%$). Dabei ist das Parameter-Set *minCRR* mit acht von zehn am erfolgreichsten. Auf die beiden Algorithmen entfallen unter dessen Verwendung jeweils vier beste CRRs.



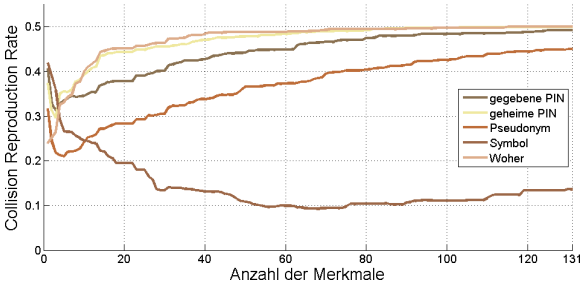
(a) EER - *Null*



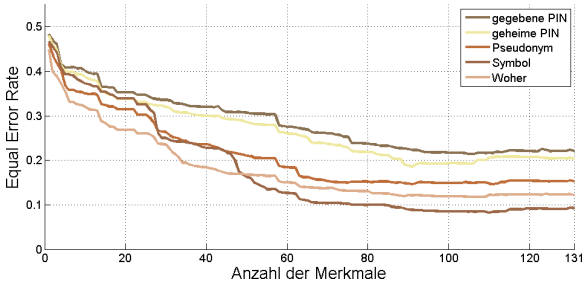
(b) CRR - *Null*



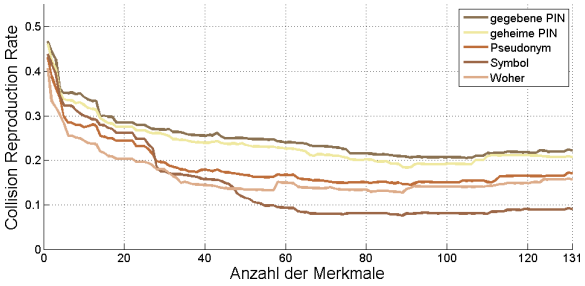
(c) EER - *minEER*



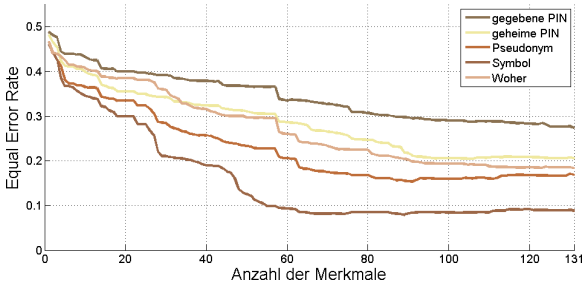
(d) CRR - *minEER*



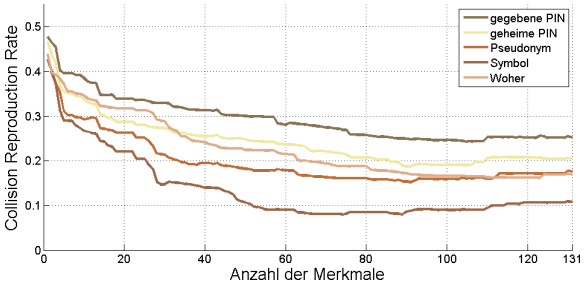
(e) EER - *minCRR*



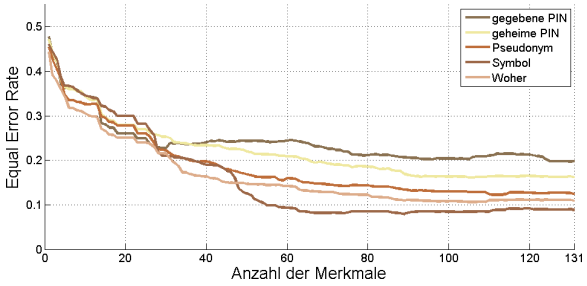
(f) CRR - *minCRR*



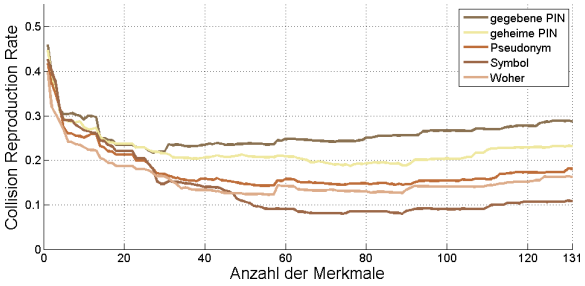
(g) EER - $RR \geq 80\%$



(h) CRR - $RR \geq 80\%$

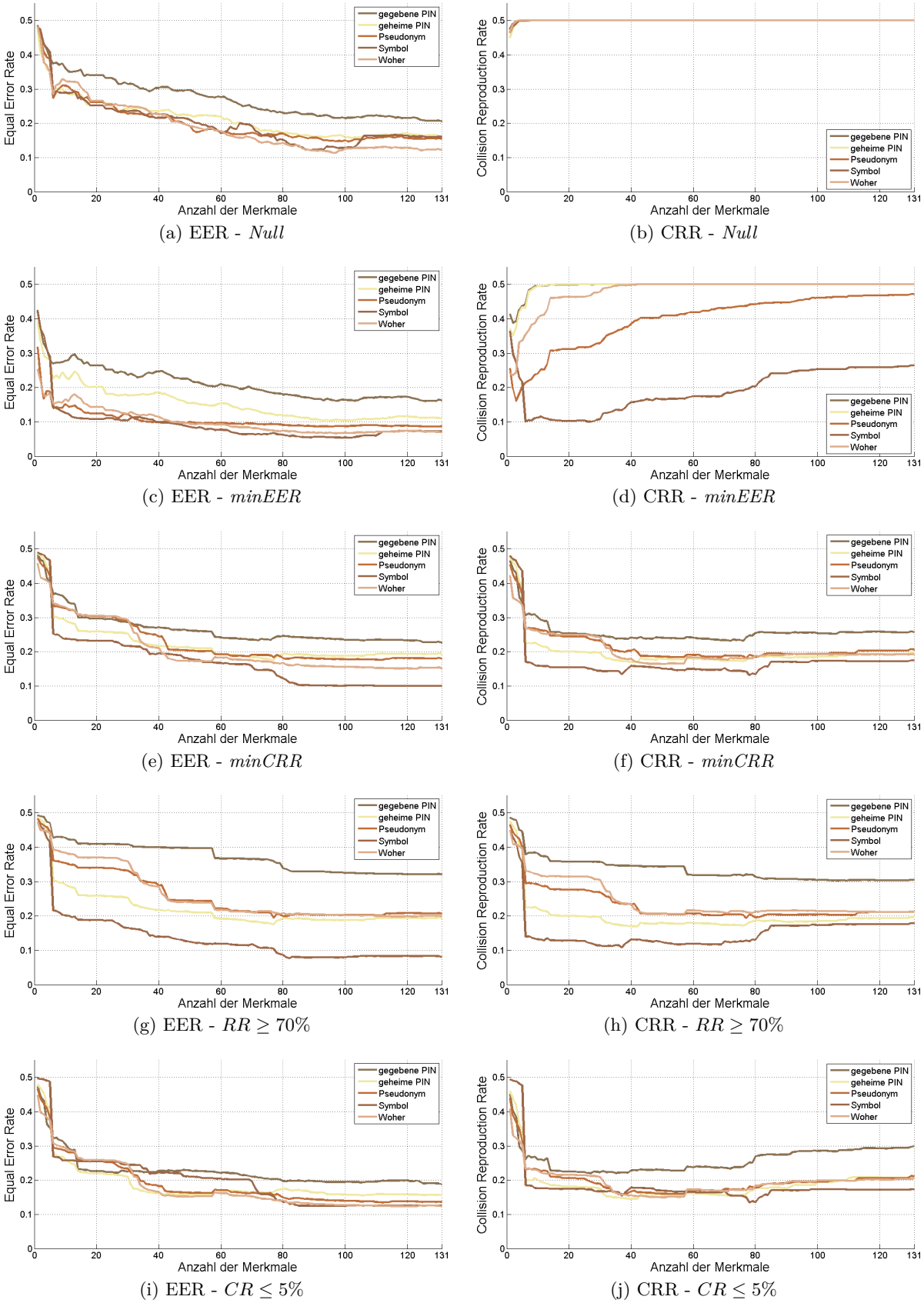


(i) EER - $CR \leq 5\%$



(j) CRR - $CR \leq 5\%$

Abbildung 7.5: Biometric Hash: Darstellung des Verlaufs der Equal Error Rate und der Collision Reproduction Rate für die fünf Parameter-Sets und die fünf Semantiken



(a) EER - *Null*
(b) CRR - *Null*
(c) EER - *minEER*
(d) CRR - *minEER*
(e) EER - *minCRR*
(f) CRR - *minCRR*
(g) EER - $RR \geq 70\%$
(h) CRR - $RR \geq 70\%$
(i) EER - $CR \leq 5\%$
(j) CRR - $CR \leq 5\%$

Abbildung 7.6: Secure Sketch: Darstellung des Verlaufs der Equal Error Rate und der Collision Reproduction Rate für die fünf Parameter-Sets und die fünf Semantiken

7.3 Evaluierung: A.3 Analyse und Selektion von Merkmalen

In diesem Abschnitt werden die Resultate der Merkmalsanalyse und -selektion (siehe Abschnitt 5.3 *Analyse und Selektion von Merkmalen (A.3)*) entsprechend vorgestellt und diskutiert. Dabei gehen wir zuerst auf die Evaluierung der selektierten Merkmal-Sets im Kontext der einzelnen Semantiken und Parameter-Sets bezüglich der beiden Algorithmen Biometric Hash und Secure Sketch entsprechend der Methodologie in Abschnitt 6.2.3 *Setup zu A.3 Analyse und Selektion von Merkmalen* ein. Hier werden die Daten aus Session 1 und Session 3 ebenfalls als Referenz- beziehungsweise Testdaten für die experimentelle Evaluierung verwendet während die der Session zwei zur Durchführung der Analyse und der darauf aufbauenden Selektion der Merkmale genutzt werden. Daran schließt sich eine allgemeine Betrachtung der Ergebnisse in Bezug auf die Merkmalsanalyse durch Filter und Wrapper mit den jeweiligen untersuchten Verfahren an. Wir haben eine Betrachtung dazu durchgeführt, wie erfolgreich die von uns gewählten Verfahren zur Merkmalsanalyse für die von uns experimentell untersuchten Testszenarien sind. Ein Aspekt ist, wie stufen die beiden Kategorien von Analyseverfahren die einzelnen Merkmale ein. Weiter vergleichen wir, wie häufig das beste Ergebnis für Verifikation und Hash-Generierung basierend auf den von den jeweiligen Wrappern und Filtern ermittelten Merkmal-Sets bestimmt wurde.

Die Tabellen, in denen wir die Testergebnisse in den folgenden Abschnitten vorstellen (Tabellen 7.7, 7.10 und 7.12), sind wie folgt aufgebaut: Jede Tabelle zeigt die Ergebnisse für die Szenarien optimale Anzahl von Merkmalen für die minimale Equal Error Rate (*(a) bestEER*) beziehungsweise die minimale Collision Reproduction Rate (*(b) bestCRR*). Die erste Spalte (*selection*) enthält jeweils die betrachteten Selektionsverfahren. Dabei werden zum Vergleich jeweils in Zeile *raw* die Ausgangswerte für das gesamte Set von 131 Merkmale angegeben. Die folgenden Spalten geben im Wechsel jeweils die Anzahl verwendeter Merkmale (n) und die ermittelten Maße Equal Error Rate (*EER*), Collision Reproduction Rate (*CRR*), Reproduction Rate (*RR*) beziehungsweise Collision Rate (*CR*) für die Szenarien an. Der jeweils beste Wert eines Szenarios für die Filter und Wrapper wird pro Semantik fett gedruckt dargestellt, während das beste Ergebnis aller Verfahren zusätzlich kursiv angezeigt wird. Wir beschränken uns dabei aus Platzgründen auf eine Auswahl von Ergebnistabellen. Die Tabellen mit allen untersuchten Konstellationen sind im Anhang C *Evaluierungsergebnisse* zu finden. Die Tabellen 7.9, 7.11 und 7.14 für das Szenario *fix60* mit festgelegter Anzahl der jeweils ersten 60 Merkmale sind entsprechend aufgebaut.

7.3.1 Testergebnisse für den Biometric Hash-Algorithmus

In Tabelle 7.7 werden die Resultate für Parameter-Set *Null* des Tolerance Factor dargestellt. Bei der Betrachtung der Ergebnisse der EER als Maß für die Verifikationsperformanz ist für alle untersuchten Fälle unter *(a) bestEER* eine Verbesserung im Vergleich zur ursprünglichen Menge an Merkmalen zu erkennen. Für die Filter werden die besten Ergebnisse von unterschiedlichen Verfahren erreicht und zwar für *gegebene PIN*, *geheime PIN* und *Symbol* durch *anova-2class*, für *Pseudonym* durch *correlation* und für *Woher* durch *entropy*. Für jede der fünf Semantiken wird für die Verifikation das beste Ergebnis durch den Wrapper *simple* ermittelt.

Tabelle 7.7: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER gegebene PIN					(b) bestCRR gegebene PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,18536	0,50000	0,00000	0,00000	131	*0,18536*	0,50000	0,00000	0,00000
anova	103	0,18389	0,50000	0,00000	0,00000	3	**0,35484**	0,35229	0,50000	0,20457
anova-2class	64	**0,17913**	0,50000	0,00000	0,00000	1	0,47985	0,46667	0,86038	0,79372
correlation	117	0,18469	0,50000	0,00000	0,00000	1	0,37064	**0,34722**	0,56226	0,25671
entropy	112	0,18270	0,50000	0,00000	0,00000	1	0,47985	0,46667	0,86038	0,79372
joint-entropy	131	0,18536	0,50000	0,00000	0,00000	2	0,36249	0,41147	0,24528	0,06822
simple	66	***0,17042***	0,50000	0,00000	0,00000	1	0,33494	0,31682	0,62830	0,26194
sfs	64	0,18480	***0,49811***	0,00377	0,00000	2	**0,32209**	***0,31626***	0,50566	0,13817
sbs	104	0,18471	0,50000	0,00000	0,00000	2	0,36816	0,35386	0,54906	0,25679
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,13036	0,50000	0,00000	0,00000	131	*0,13036*	0,50000	0,00000	0,00000
anova	101	0,12772	0,50000	0,00000	0,00000	2	0,35723	0,36745	0,33585	0,07076
anova-2class	114	**0,12671**	0,50000	0,00000	0,00000	6	0,41826	0,45459	0,16038	0,06956
correlation	118	0,13036	0,50000	0,00000	0,00000	1	**0,33463**	**0,31818**	0,63208	0,26843
entropy	116	0,12788	0,50000	0,00000	0,00000	2	0,47209	0,47219	0,51132	0,45570
joint-entropy	131	0,13036	0,50000	0,00000	0,00000	1	0,40626	0,37157	0,44340	0,18654
simple	58	***0,11927***	0,50000	0,00000	0,00000	1	**0,26080**	0,25639	0,75283	0,26560
sfs	104	0,12412	***0,49906***	0,00189	0,00000	2	0,26916	***0,25345***	0,70943	0,21633
sbs	105	0,12047	0,50000	0,00000	0,00000	1	0,38521	0,35718	0,52075	0,23512
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,11338	0,50000	0,00000	0,00000	131	*0,11338*	0,50000	0,00000	0,00000
anova	50	0,10778	0,50000	0,00000	0,00000	1	**0,34585**	**0,29900**	0,54906	0,14706
anova-2class	117	0,11203	0,50000	0,00000	0,00000	7	0,41152	0,46800	0,08679	0,02279
correlation	98	**0,10202**	0,50000	0,00000	0,00000	1	0,39623	0,37547	0,52453	0,27547
entropy	117	0,11301	0,50000	0,00000	0,00000	1	0,40768	0,38737	0,72264	0,49739
joint-entropy	105	0,10855	0,50000	0,00000	0,00000	2	0,35199	0,36861	0,33396	0,07119
simple	56	***0,08318***	0,50000	0,00000	0,00000	1	0,29854	0,27464	0,66604	0,21531
sfs	112	0,10910	0,50000	0,00000	0,00000	2	**0,28097**	***0,26321***	0,54906	0,07547
sbs	118	0,11338	0,50000	0,00000	0,00000	1	0,33224	0,29040	0,58491	0,16571
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12512	0,50000	0,00000	0,00000	131	*0,12512*	0,50000	0,00000	0,00000
anova	101	0,11963	0,50000	0,00000	0,00000	2	0,40822	0,37417	0,81132	0,55965
anova-2class	67	**0,09934**	0,50000	0,00000	0,00000	7	0,40773	0,44278	0,20566	0,09122
correlation	85	0,11363	0,50000	0,00000	0,00000	1	0,38765	0,37587	0,57170	0,32344
entropy	115	0,12350	0,50000	0,00000	0,00000	4	0,40741	0,39868	0,48491	0,28226
joint-entropy	126	0,12436	0,50000	0,00000	0,00000	1	**0,35174**	**0,32186**	0,57736	0,22108
simple	22	***0,08850***	***0,49717***	0,00566	0,00000	1	**0,26978**	***0,23755***	0,69245	0,16756
sfs	91	0,10937	0,49906	0,00189	0,00000	1	**0,26978**	***0,23755***	0,69245	0,16756
sbs	105	0,11774	0,50000	0,00000	0,00000	2	0,38542	0,36379	0,52075	0,24833
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,09528	0,50000	0,00000	0,00000	131	*0,09528*	0,50000	0,00000	0,00000
anova	116	0,09447	0,50000	0,00000	0,00000	1	0,34759	**0,29399**	0,53019	0,11818
anova-2class	81	0,09008	0,50000	0,00000	0,00000	5	0,42683	0,46212	0,16226	0,08650
correlation	118	0,09528	0,50000	0,00000	0,00000	1	0,34430	0,33064	0,71321	0,37449
entropy	69	**0,08539**	0,50000	0,00000	0,00000	1	0,39509	0,39329	0,59811	0,38469
joint-entropy	131	0,09528	0,50000	0,00000	0,00000	2	**0,30417**	0,31762	0,43208	0,06731
simple	52	***0,08505***	0,50000	0,00000	0,00000	1	**0,26356**	***0,23257***	0,70189	0,16702
sfs	118	0,09528	0,50000	0,00000	0,00000	1	**0,26356**	***0,23257***	0,70189	0,16702
sbs	114	0,09425	0,50000	0,00000	0,00000	2	0,37025	0,35630	0,57170	0,28429

Dabei weist die Semantik *Pseudonym* mit 0,08318 die geringste Equal Error Rate auf. Verglichen mit der entsprechenden EER von 0,11998 für das *Pseudonym* basierend auf allen Merkmalen bedeutet dies eine Verbesserung von circa 27%. Eine wesentliche Beobachtung stellt das Ergebnis für den Wrapper *simple* und die Semantik *Symbol* dar. Hier reichen lediglich 22 Mermale, um eine EER von 0,08850 zu bestimmen. Das entspricht einer Verringerung der mittels aller 131 Merkmale berechneten EER von 0,12512 um circa 29%. Das zeigt, dass die bei der Selektion nicht berücksichtigten 109 Merkmale einen negativen Einfluss in dieser Konstellation ausüben. Die 22 Merkmale sind in Tabelle 7.8 in der Reihenfolge ihrer Selektion durch den *simple* Wrapper aufgelistet. Bei Auswahl dieser Merkmale kann bei einer Umsetzung des Biometric Hash-Algorithmus Rechenzeit eingespart werden, wobei zu bedenken ist, dass zum Teil das nicht selektierte Merkmal n_{31} (PathLength) als Grundlage benötigt wird. Die Performanz der Hash-Generierung weist in diesem Szenario schlechte Werte auf. Die Collision Reproduction Rate beträgt in den meisten Fällen 0,5. Nur für drei Ergebnisse ist die CRR geringfügig kleiner. Dies rührt von der Reproduction Rate her, die jeweils circa 0,002 beträgt und daher vernachlässigbar ist.

Tabelle 7.8: Biometric Hash: Übersicht über die durch den Wrapper *simple* selektierten 22 Merkmale, für die mittels Parameter-Set *Null* für den TF und Semantik *Symbol* eine EER von 0,08850 bestimmt wird, in der Reihenfolge ihrer Selektion

n_i	Name	Beschreibung
67	PathRatio3	PathLength(ConvexHull(Segments)) vs PathLength(BoundingBox) * 1000
86	StartEndRatio	Distance (Start Point; End Point) * 1000 DIV PathLength
23	AvgPressure	Average Writing Pressure relative to MaxPressure * 1000
64	AreaRatio3	Area(ConvexHull(Segments)) vs Area(BoundingBox) * 1000
29	RatioTPenUpPEnDown	Ratio of TPenUp by TTotal * 1000
26	Vx_TN	Normalised Average velocity in x direction in pixels / VxMax * 1000
87	XMaxXMinRatio	Distance (Maximum X Point; Minimum X Point) * 1000 DIV PathLength
56	PenDPress	Average PenDown Pressure normalized to 1 * 1000
88	YMaxYMinRatio	Distance (Maximum Y Point; Minimum Y Point) * 1000 DIV PathLength
49	AreaX4	Numeric Integration of X values for 4th one-fifth time period
95	dotcnt_min	minima Number of neighbour Points
27	Vy_TN	Normalised Average velocity in y direction in pixels / VyMax * 1000
47	AreaX2	Numeric Integration of X values for 2nd one-fifth time period
51	AreaY1	Numeric Integration of Y values for 1st one-fifth time period
48	AreaX3	Numeric Integration of X values for 3rd one-fifth time period
89	StartCentroidEndRatio	Distance (Start Point; Centroid) * 1000 DIV Distance (End Point; Centroid)
16	CentroidY_SN	Centroid of vertical pen position normalised to bounding box height * 1000
131	PressureDeviation	standard deviation of the Pressure
28	TPenUp	Absolute cummulated Pen-up time in ms
50	AreaX5	Numeric Integration of X values for 4th one-fifth time period
18	CentroidAzimuth_SN	Horizontal azimuth of centroid from origin normalised to PI/2 * 1000
55	AreaY5	Numeric Integration of Y values for 4th one-fifth time period

Das ändert sich, wenn wir die CRR-Werte in Szenario *bestCRR* (siehe (b) in Tabelle 7.7) betrachten. Hier ergeben sich Raten, die deutlich geringer sind als 0,5. Trotzdem fallen sie für die Erzeugung von Hashes noch zu hoch aus. Als Beispiel sei an dieser Stelle die kleinste Collision Reproduction Rate in den in Tabelle 7.7 betrachteten Konstellationen hervorgehoben. Diese wird für die Semantik *Woher* jeweils durch die Wrapper *simple* und *sfs* mit 0,23257 berechnet. Sie resultiert aus einer Reproduction Rate von 0,70189 und einer Collision Rate von 0,16702. Trotz der deutlich verbesserten Collision Reproduction Rate sind die in diesem Szenario ermittelten Werte für den Einsatz als biometrische Hashes ebenfalls nicht geeignet. Hinzu kommt, dass die

Anzahl der Merkmale in einem Intervall von $[1, 7]$ liegen. Dies bestätigt die bereits in Abbildung 7.1b gezeigte Tendenz, dass für die jeweils kleinste CRR nur eine geringe Anzahl von Merkmalen benötigt wird. Die sehr geringe Anzahl von Merkmalen führt auch dazu, dass die Equal Error Rate in keinem der für dieses Szenario untersuchten Fälle im Vergleich zur Ausgangssituation mit 131 Merkmalen verbessert wird. So ist beispielsweise die durch den Wrapper *sfs* für die *geheime PIN* erreichte EER von 0,26080 einerseits die im Szenario *bestCRR* kleinste. Auf der anderen Seite ist sie doppelt so groß, wie die EER, die für diese Semantik mit allen 131 Merkmalen bestimmt wurde.

Tabelle 7.9 zeigt die Ergebnisse für das Szenario, dass pauschal die ersten 60 Merkmale aus dem jeweilig ermittelten Ranking der Verfahren zur Analyse der Merkmale ausgewählt werden. An dieser Stelle kommt es bei den fünf Semantiken wieder zu einer Verbesserung gegenüber der ursprünglichen Equal Error Rate bei 131 Merkmalen.

Tabelle 7.9: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,23081	0,49906	0,00189	0,00000	60	0,18235	0,50000	0,00000	0,00000
anova	60	0,22176	0,50000	0,00000	0,00000	60	**0,14917**	0,50000	0,00000	0,00000
anova-2class	60	**0,18438**	0,50000	0,00000	0,00000	60	0,15176	0,50000	0,00000	0,00000
correlation	60	0,23724	0,50000	0,00000	0,00000	60	0,17219	0,50000	0,00000	0,00000
entropy	60	0,20225	0,50000	0,00000	0,00000	60	0,14931	0,50000	0,00000	0,00000
joint-entropy	60	0,22600	0,50000	0,00000	0,00000	60	0,17486	0,50000	0,00000	0,00000
wrapper	60	***0,17441***	0,50000	0,00000	0,00000	60	***0,12463***	0,50000	0,00000	0,00000
sfs	60	0,19542	***0,49811***	0,00377	0,00000	60	0,16906	***0,49717***	0,00566	0,00000
sbs	60	0,20451	0,50000	0,00000	0,00000	60	0,15035	0,50000	0,00000	0,00000

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,12331	0,50000	0,00000	0,00000	60	0,16666	0,49906	0,00189	0,00000
anova	60	**0,12296**	0,50000	0,00000	0,00000	60	0,14501	0,50000	0,00000	0,00000
anova-2class	60	0,12861	0,50000	0,00000	0,00000	60	**0,10459**	0,50000	0,00000	0,00000
correlation	60	0,12414	0,50000	0,00000	0,00000	60	0,14274	0,50000	0,00000	0,00000
entropy	60	0,12362	0,50000	0,00000	0,00000	60	0,15935	0,50000	0,00000	0,00000
joint-entropy	60	0,14073	0,50000	0,00000	0,00000	60	0,20860	0,49906	0,00189	0,00000
wrapper	60	***0,08839***	0,50000	0,00000	0,00000	60	***0,10216***	0,50000	0,00000	0,00000
sfs	60	0,13048	***0,49528***	0,00943	0,00000	60	0,13898	***0,49434***	0,01132	0,00000
sbs	60	0,15840	0,50000	0,00000	0,00000	60	0,15496	0,50000	0,00000	0,00000

	Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,13232	0,50000	0,00000	0,00000
anova	60	0,11616	0,50000	0,00000	0,00000
anova-2class	60	0,09997	0,50000	0,00000	0,00000
correlation	60	0,12808	0,50000	0,00000	0,00000
entropy	60	**0,09378**	0,50000	0,00000	0,00000
joint-entropy	60	0,14013	0,50000	0,00000	0,00000
wrapper	60	***0,08678***	0,50000	0,00000	0,00000
sfs	60	0,12642	***0,49623***	0,00755	0,00000
sbs	60	0,12412	0,50000	0,00000	0,00000

Im Vergleich zur optimalen Anzahl von Merkmalen für die EER in Tabelle 7.7 (a) jedoch nicht in jedem Fall. Die besten Ergebnisse basieren wiederum auf dem *simple* Wrapper. Mit einer EER von 0,08839 wird das beste Ergebnis auch hier für das *Pseudonym* bestimmt. Dieser liegt nur circa 0,25% über der mit 54 Merkmalen bestimmten EER von 0,08318 für Szenario *minEER*. Da die Anzahl der Merkmale meist außerhalb des Bereiches liegen, in dem sich für den Biometric Hash eine Collision Reproduction Rate von kleiner als 0,5 ergibt (siehe auch Abbildung 7.1b), beträgt die CRR in den meisten Fällen 0,5 beziehungsweise geringfügig weniger.

Die Ergebnisse für den Biometric Hash-Algorithmus und die Parametrisierung *minEER* sind in Tabelle 7.10 dargestellt. Für das Szenario *(a) bestEER* ist zu erkennen, dass für jede Semantik und jedes untersuchte Selektionsverfahren eine bessere Equal Error Rate als für den Ausgangszustand *raw* bestimmt wurde. Eine Ausnahme bildet *anova* mit demselben Ergebnis wie *raw*, welches allerdings durch ein Merkmal-Set ermittelt wird, das 13 Merkmale weniger enthält. Die geringste Equal Error Rate für die Merkmalsselektion basierend auf dem Biometric Hash-Algorithmus wird mit 0,04339 für die Semantik *Symbol* mittels Wrapper *sfs* erreicht. Im Vergleich zum Ausgangswert (*raw*) von 0,07366 bedeutet das eine Verbesserung um circa 41,3%. Die Equal Error Rate, welche ohne optimierte Parametrisierung und Merkmalsselektion ermittelt wurde (EER=0,12512), wird hier um circa 65,3% verbessert. Im Gegensatz zur Parametrisierung *Null* wird hier nicht jedes beste Ergebnis für EER pro Semantik durch einen einzigen Merkmalsselektor erreicht. Es werden die geringsten EERs durch *anova-2class* (0,17395, *gegebene PIN*), *sbs* (0,12187, *geheime PIN*), *sfs* (0,08515, *Pseudonym* und 0,04339, *Symbol*) sowie *simple* (0,07488, *Woher*) bestimmt. Die Performanz für das Symbol stellt sich dabei für die Nutzung in einem biometrischen System für dynamische Handschriftverifikations als durchaus einsetzbar dar. Die Anzahl der verwendeten Merkmale liegt in diesem Fall bei 79.

Die Performanz der Hash-Generierung für die Parametrisierung minEER liegt für fast alle Semantiken im Bereich zwischen 0,41 und 0,5 für die CRR. Die Ausnahme bildet das *Symbol* mit einer Collision Reproduction Rate von 0,04338 (*sfs*), berechnet, basierend auf einer Reproduction Rate von 0,95660 und einer Collision Rate 0,04336. Das bedeutet, dass für circa 95,6% der Nutzer die individuellen Hashes reproduziert werden können, während bei 4,3% der Versuche Kollisionen auftreten. Für die anderen vier Semantiken stellt sich dieses Verhältnis im besten Fall (*Pseudonym*, *sfs*, CRR=0,41513) mit circa 17% zu 0% dar. Das bedeutet, es treten zwar keine Kollisionen auf, aber nur eine geringe Anzahl von Hashes kann identisch reproduziert werden.

Im Szenario *(b) bestCRR* in Tabelle 7.10, ist wieder eine deutliche Verschlechterung der ermittelten EERs zu erkennen. Mit Ausnahme der Semantik *Symbol* sind alle EERs größer als bei Verwendung aller 131 Merkmale (*raw*). Für das *Symbol* beträgt die EER 0,04609 bei Verwendung von 75 Merkmalen. Bei der Hash-Generierung kann für das *Symbol* mit 46 Merkmalen eine weitere Verbesserung erzielt werden. Die CRR, die gleichzeitig die beste für den Biometric Hash-Algorithmus im Kontext der Merkmalsselektion ist, beträgt an dieser Stelle 0,03672, basierend auf einer RR von 0,99434 und einer CR von 0,06778. Das bedeutet eine Reproduktion der individuellen Hashes von nahezu 100% bei einer Kollisionswahrscheinlichkeit von circa 6,8%.

Tabelle 7.10: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER gegebene PIN					(b) bestCRR gegebene PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,20082	0,49253	0,01509	0,00015	131	*0,20082*	0,49253	0,01509	0,00015
anova	118	0,20082	0,49253	0,01509	0,00015	6	0,29120	0,27778	0,61509	0,17065
anova-2class	63	***0,17395***	0,47395	0,05283	0,00073	9	0,33073	0,32201	0,70377	0,34779
correlation	29	0,20041	**0,44294**	0,12264	0,00853	4	**0,27336**	**0,26693**	0,66981	0,20366
entropy	65	0,17922	0,47854	0,04340	0,00047	13	0,33044	0,33091	0,59245	0,25428
joint-entropy	103	0,19956	0,48507	0,03019	0,00033	8	0,28924	0,29572	0,57170	0,16313
simple	25	0,18745	***0,42975***	0,14528	0,00479	3	0,27172	0,26669	0,70566	0,23904
sfs	88	**0,18514**	0,45675	0,08679	0,00029	7	**0,23043**	***0,21892***	0,72075	0,15860
sbs	118	0,20082	0,49253	0,01509	0,00015	7	0,26928	0,26319	0,72453	0,25091
selection	**geheime PIN**					**geheime PIN**				
	n	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12790	0,49906	0,00189	0,00000	131	*0,12790*	0,49906	0,00189	0,00000
anova	115	**0,12613**	0,49906	0,00189	0,00000	4	0,27756	0,30998	0,45472	0,07467
anova-2class	118	0,12790	0,49906	0,00189	0,00000	8	**0,27380**	**0,27005**	0,71321	0,25330
correlation	118	0,12790	0,49906	0,00189	0,00000	3	0,30402	0,28643	0,57925	0,15210
entropy	87	0,12789	**0,49813**	0,00377	0,00004	11	0,29899	0,30261	0,56792	0,17315
joint-entropy	131	0,12790	0,49906	0,00189	0,00000	3	0,29930	0,30778	0,50943	0,12500
simple	115	0,12613	0,49906	0,00189	0,00000	3	0,22961	0,21234	0,75472	0,17939
sfs	116	0,12679	0,49811	0,00377	0,00000	5	**0,22514**	***0,20463***	0,70000	0,10925
sbs	100	***0,12187***	***0,49340***	0,01321	0,00000	4	0,31121	0,28786	0,56415	0,13988
selection	**Pseudonym**					**Pseudonym**				
	n	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08943	0,45004	0,10000	0,00007	131	*0,08943*	0,45004	0,10000	0,00007
anova	118	0,08943	0,45004	0,10000	0,00007	11	**0,18880**	**0,22003**	0,62453	0,06459
anova-2class	111	0,08818	0,44534	0,10943	0,00011	17	0,20076	0,22235	0,62453	0,06923
correlation	104	**0,08422**	0,44439	0,11132	0,00011	6	0,23060	0,23015	0,66038	0,12068
entropy	111	0,08892	**0,44343**	0,11321	0,00007	12	0,22532	0,22153	0,72264	0,16571
joint-entropy	124	0,08869	0,44909	0,10189	0,00007	14	0,21824	0,22685	0,66038	0,11408
simple	99	0,08844	0,43687	0,12642	0,00015	5	0,18947	0,16881	0,77925	0,11687
sfs	106	***0,08515***	***0,41513***	0,16981	0,00007	9	0,13690	0,13523	0,78868	0,05914
sbs	112	0,08582	0,42647	0,14717	0,00011	9	**0,11919**	***0,11560***	0,87736	0,10856
selection	**Symbol**					**Symbol**				
	n	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,07366	0,13610	0,75472	0,02692	131	0,07366	0,13610	0,75472	0,02692
anova	98	0,07229	**0,11854**	0,79434	0,03142	35	0,09572	**0,07992**	0,93962	0,09946
anova-2class	113	0,07151	0,13458	0,75849	0,02765	77	0,10521	0,10983	0,85283	0,07250
correlation	118	0,07366	0,13610	0,75472	0,02692	60	0,10276	0,10185	0,85283	0,05653
entropy	118	0,07366	0,13610	0,75472	0,02692	71	0,11724	0,11168	0,85472	0,07808
joint-entropy	111	**0,07150**	0,11978	0,79057	0,03012	86	**0,08145**	0,10952	0,82075	0,03980
simple	76	0,07048	0,10684	0,81887	0,03255	35	0,08150	0,08082	0,91509	0,07674
sfs	79	***0,04339***	***0,04338***	0,95660	0,04336	75	***0,04609***	0,04122	0,96415	0,04659
sbs	83	0,05039	0,04933	0,94906	0,04771	46	0,06382	***0,03672***	0,99434	0,06778
selection	**Woher**					**Woher**				
	n	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08559	0,50000	0,00000	0,00000	131	*0,08559*	0,50000	0,00000	0,00000
anova	112	0,08419	0,49906	0,00189	0,00000	1	0,25035	0,22529	0,72453	0,17511
anova-2class	61	0,07236	0,49811	0,00377	0,00000	8	0,24160	0,24508	0,60566	0,09583
correlation	118	0,08559	0,50000	0,00000	0,00000	2	0,34888	0,33423	0,55094	0,21941
entropy	68	**0,06257**	**0,49811**	0,00377	0,00000	9	**0,21633**	**0,22137**	0,62642	0,06916
joint-entropy	131	0,08559	0,50000	0,00000	0,00000	2	0,26480	0,23902	0,65849	0,13654
simple	40	***0,07488***	***0,48213***	0,03585	0,00011	2	0,18211	0,17509	0,77358	0,12377
sfs	117	0,08541	0,50000	0,00000	0,00000	2	**0,17358**	***0,15292***	0,80000	0,10584
sbs	118	0,08559	0,50000	0,00000	0,00000	3	0,21177	0,20867	0,78302	0,20036

Tabelle 7.11: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,22926	0,44911	0,10377	0,00200	60	0,18448	0,48396	0,03208	0,00000
anova	60	0,21671	0,47600	0,04906	0,00105	60	0,15359	0,49245	0,01509	0,00000
anova-2class	60	**0,17936**	0,46925	0,06226	0,00076	60	**0,14133**	0,49530	0,00943	0,00004
correlation	60	0,21724	0,47242	0,05660	0,00145	60	0,17521	0,49247	0,01509	0,00004
entropy	60	0,18733	0,47493	0,05094	0,00080	60	0,14514	0,49343	0,01321	0,00007
joint-entropy	60	0,23361	**0,45633**	0,08868	0,00134	60	0,16982	**0,48962**	0,02076	0,00000
simple	60	***0,20598***	0,47763	0,04528	0,00054	60	0,13891	0,48774	0,02453	0,00000
sfs	60	0,21082	***0,39604***	0,21321	0,00530	60	0,15626	***0,44915***	0,10189	0,00018
sbs	60	0,23749	0,40902	0,18491	0,00294	60	***0,13519***	0,48019	0,03962	0,00000

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,12023	0,37277	0,25660	0,00214	60	0,09800	0,09967	0,86415	0,06350
anova	60	0,10563	0,39372	0,21321	0,00065	60	**0,08304**	**0,09570**	0,85094	0,04234
anova-2class	60	0,10558	**0,37944**	0,24151	0,00040	60	0,12007	0,11892	0,86792	0,10577
correlation	60	0,12013	0,40158	0,19811	0,00127	60	0,10276	0,10185	0,85283	0,05653
entropy	60	**0,10018**	0,38229	0,23585	0,00044	60	0,12100	0,11707	0,86038	0,09452
joint-entropy	60	0,12854	0,38440	0,23396	0,00276	60	0,12017	0,11676	0,85094	0,08447
simple	60	***0,09238***	0,39082	0,21887	0,00051	60	0,07358	0,09329	0,85094	0,03752
sfs	60	0,10799	***0,28329***	0,43585	0,00243	60	***0,04843***	***0,04156***	0,96604	0,04917
sbs	60	0,10735	0,30210	0,40000	0,00421	60	0,05975	0,04414	0,97358	0,06187

	Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,12345	0,48868	0,02264	0,00000
anova	60	0,10202	**0,49528**	0,00943	0,00000
anova-2class	60	0,07454	0,49717	0,00566	0,00000
correlation	60	0,11161	0,49717	0,00566	0,00000
entropy	60	**0,06684**	0,49717	0,00566	0,00000
joint-entropy	60	0,13491	0,49623	0,00755	0,00000
simple	60	***0,08287***	0,49340	0,01321	0,00000
sfs	60	0,10239	***0,45849***	0,08302	0,00000
sbs	60	0,11028	0,46511	0,06981	0,00004

Hier fällt auf, dass die Anzahl der selektierten Merkmale beim *Symbol* deutlich höher sind als bei den anderen Semantiken. Im Vergleich einerseits mit der Collision Reproduction Rate für das Symbol basierend auf allen 131 Merkmalen und der Parametrisierung *minEER* (CRR=0,13610) ergibt sich eine Verbesserung um circa 73%. Legen wir auf der anderen Seite die CRR von 0,5 für 131 Merkmale und ohne optimiertes Parameterset (*Null*) zugrunde, ergibt sich eine Verbesserung um etwa 92,7%.

Setzen wir die Anzahl der Merkmale auf die ersten 60 fest (siehe Tabelle 7.11 (c)), verschlechtern sich alle EERs, die auf den ersten 60 Merkmalen der Ausgangssituation beruhen. Die basierend auf den selektierten Merkmal-Sets bestimmten EERs verschlechtern sich im Vergleich mit den Werten im Szenario *bestEER*, liegen aber deutlich unter den Ergebnissen für die Equal Error Rate im Szenario *bestCRR*. Die beste Equal Error Rate wird auch hier für das *Symbol* und Wrapper *sfs* bestimmt und beträgt mit 0,04843 nur circa 0,005 weniger als die beste EER in

bestEER. Zur Berechnung diese Wertes werden 19 Merkmale weniger verwendet. Ähnlich sieht es für die Hash-Generierung aus. Auch hier liefert das *Symbol* basierend auf dem *sfs* Wrapper die kleinste Collision Reproduction Rate, die mit 0,04156 geringer als bei *minEER* und höher als bei *minCRR* ausfällt. Die zugrunde liegenden Werte für Reproduction Rate und Collision Rate sind 0,96604 beziehungsweise 0,04917.

7.3.2 Testergebnisse für den Secure Sketch-Algorithmus

In Tabelle 7.12 sind die Ergebnisse für die Merkmalsselektion für den Secure Sketch-Algorithmus und Parameter-Set *minEER* dargestellt. Für diese Konstellation werden sowohl die beste Equal Error Rate als auch die beste Collision Reproduction Rate ermittelt. Auch in diesem Fall ist die entsprechende Semantik das *Symbol*. Basierend auf einem Set von lediglich 26 Merkmalen, bestimmt durch den Wrapper *sbs*, wird hier im Szenario *bestEER* mit 0,03289 die kleinste EER gemessen. Legt man für einen Vergleich die mittels 131 Merkmalen im selben Szenario bestimmte EER von 0,07259 zugrunde, ergibt sich eine Verbesserung von circa 54,7%. Eine Verringerung um etwa 79,7% ergibt sich, wenn wir als Vergleichswert die Equal Error Rate verwenden, die für alle 131 Merkmale ohne optimierte Parametrisierung für das *Symbol* durch den Secure Sketch-Algorithmus ermittelt wird. Dabei wird die EER von 0,16158 auf die genannten 0,03286 gesenkt. Für das Szenario *minEER* wird auch eine sehr geringe Collision Reproduction Rate von 0,03142 ermittelt. Diese basiert auf einer Reproduction Rate von 0,96604 und einer Collision Rate von 0,02888.

Die beste CRR während der Evaluierung der Merkmalsselektion wird basierend auf einem Set von 22 Merkmalen und dem Szenario *bestCRR* berechnet und beträgt 0,02848. Die Reproduction Rate und die Collision Rate, die die Grundlage für dieses Resultat bilden, betragen 0,98491 beziehungsweise 0,04187. Dies entspricht einer Wahrscheinlichkeit von circa 98,5%, dass eine Person basierend auf unserem Testset und unserer Methodologie ihren eigenen biometrischen Hash reproduzieren kann. Eine Kollision mit dem Hash einer beliebigen anderen Person ist dabei zu etwa 4,2% wahrscheinlich. Auf der einen Seite beträgt die Verbesserung der CRR circa 89,2%, ausgehend von der CRR von 0,26392 für das Parameter-Set *minEER*, das *Symbol* und das komplette Set der 131 Merkmale. Wählen wir als Vergleichswert die Collision Reproduction Rate von 0,5 zu Beginn der Optimierung, können wir eine Verringerung um 94,3% vermerken.

Die zur Bestimmung der besten Collision Reproduction Rate verwendeten 22 Merkmale sind in Tabelle 7.13 entsprechend der Reihenfolge ihrer Selektion durch den Wrapper *sbs* aufgelistet. Im Vergleich zu den 22 Merkmalen, die in Tabelle 7.8 dargestellt sind, fällt auf, dass 10 Merkmale in beiden Listen identisch sind. Diese Tatsache motiviert die im Folgenden beschriebenen Ansätze sowohl zur Erstellung eines Rankings entsprechend der Auswahl durch die Wrapper und Filter, als auch der Auswertung der Häufigkeit, mit denen die einzelnen Verfahren das Merkmal-Set bestimmen, welches während der experimentellen Evaluierung das beste Ergebnis generiert.

Tabelle 7.12: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,16211	0,50000	0,00000	0,00000	131	*0,16211*	0,50000	0,00000	0,00000
anova	116	0,15996	0,50000	0,00000	0,00000	1	0,34135	**0,32139**	0,74151	0,38429
anova-2class	41	***0,15312***	***0,49528***	0,00943	0,00000	6	0,33016	0,32903	0,54717	0,20522
correlation	118	0,16211	0,50000	0,00000	0,00000	2	**0,32633**	0,33387	0,57736	0,24510
entropy	65	0,15592	0,50000	0,00000	0,00000	2	0,41604	0,37440	0,87358	0,62239
joint-entropy	131	0,16211	0,50000	0,00000	0,00000	2	0,36197	0,35981	0,40377	0,12340
simple	45	0,15989	**0,49906**	0,00189	0,00000	2	0,26976	0,24243	0,67547	0,16034
sfs	67	0,15508	**0,49906**	0,00189	0,00000	3	**0,26736**	***0,24205***	0,63774	0,12184
sbs	105	**0,15474**	0,50000	0,00000	0,00000	4	0,32971	0,31404	0,57170	0,19978
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,10927	0,50000	0,00000	0,00000	131	*0,10927*	0,50000	0,00000	0,00000
anova	109	0,10587	0,50000	0,00000	0,00000	2	0,35898	0,35731	0,35472	0,06934
anova-2class	68	***0,09452***	0,50000	0,00000	0,00000	1	0,33302	0,25296	0,98679	0,49271
correlation	118	0,10927	0,50000	0,00000	0,00000	1	0,30827	0,29238	0,74906	0,33382
entropy	115	0,10640	0,50000	0,00000	0,00000	2	**0,26835**	**0,22580**	0,86604	0,31763
joint-entropy	101	0,10388	0,50000	0,00000	0,00000	2	0,33250	0,36785	0,33962	0,07533
simple	84	0,10565	0,50000	0,00000	0,00000	2	0,23945	0,21733	0,69057	0,12522
sfs	87	**0,09945**	0,50000	0,00000	0,00000	3	**0,16674**	***0,16566***	0,83208	0,16339
sbs	105	0,10524	0,50000	0,00000	0,00000	2	0,29596	0,28282	0,56226	0,12790
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08757	0,47170	0,05660	0,00000	131	*0,08757*	0,47170	0,05660	0,00000
anova	118	0,08757	0,47170	0,05660	0,00000	5	0,17106	**0,15071**	0,79623	0,09764
anova-2class	33	***0,08424***	0,34071	0,32075	0,00218	6	0,21186	0,19537	0,71698	0,10773
correlation	95	0,08512	0,46228	0,07547	0,00004	7	0,17693	0,23748	0,57170	0,04666
entropy	118	0,08757	0,47170	0,05660	0,00000	3	0,19172	0,18552	0,82453	0,19557
joint-entropy	126	0,08679	0,47170	0,05660	0,00000	12	**0,15695**	0,19866	0,65283	0,05015
simple	118	0,08757	0,47170	0,05660	0,00000	2	0,16261	0,14673	0,82264	0,11611
sfs	60	**0,08432**	0,36515	0,26981	0,00011	5	0,13566	0,13409	0,78113	0,04931
sbs	64	0,08577	0,34067	0,31887	0,00022	9	**0,10743**	***0,10250***	0,90377	0,10878
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,07259	0,26392	0,47547	0,00330	131	0,07259	0,26392	0,47547	0,00330
anova	98	0,06132	0,23456	0,53585	0,00497	24	**0,07377**	**0,07132**	0,91698	0,05962
anova-2class	63	0,06212	**0,18790**	0,63208	0,00787	7	0,11395	0,11292	0,88491	0,11074
correlation	115	0,06857	0,26116	0,48113	0,00345	23	0,12350	0,11451	0,84906	0,07808
entropy	118	0,07259	0,26392	0,47547	0,00330	8	0,12035	0,11348	0,86415	0,09111
joint-entropy	107	**0,05755**	0,25490	0,49434	0,00414	16	0,14228	0,12779	0,89245	0,14804
simple	63	0,05430	0,20171	0,60189	0,00530	15	0,07907	0,07640	0,91698	0,06978
sfs	76	0,04845	0,15492	0,69811	0,00795	22	0,05273	0,06054	0,90189	0,02297
sbs	26	***0,03286***	***0,03142***	0,96604	0,02888	22	***0,04078***	***0,02848***	0,98491	0,04187
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,06817	0,50000	0,00000	0,00000	131	*0,06817*	0,50000	0,00000	0,00000
anova	118	0,06817	0,50000	0,00000	0,00000	1	**0,22336**	**0,20555**	0,76226	0,17337
anova-2class	80	0,06098	**0,49906**	0,00189	0,00000	5	0,23786	0,22941	0,65094	0,10976
correlation	118	0,06817	0,50000	0,00000	0,00000	1	0,31189	0,28336	0,79245	0,35918
entropy	70	***0,04823***	0,50000	0,00000	0,00000	1	0,25254	0,21578	0,85849	0,29006
joint-entropy	131	0,06817	0,50000	0,00000	0,00000	2	0,26954	0,28151	0,51509	0,07812
simple	61	**0,06321**	0,49906	0,00189	0,00000	1	0,21356	0,18044	0,87736	0,23824
sfs	69	0,06602	***0,49340***	0,01321	0,00000	2	**0,18757**	***0,16736***	0,77925	0,11397
sbs	103	0,06651	0,49811	0,00377	0,00000	2	0,24033	0,22074	0,81698	0,25845

Tabelle 7.13: Secure Sketch: Übersicht über die durch den Wrapper *sbs* selektierten 22 Merkmale, für die mittels Parameter-Set *minEER* für den EF und Semantik *Symbol* eine CRR von 0,02848 bestimmt wird, in der Reihenfolge ihrer Selektion

n_i	Name	Beschreibung
6	SegmentCount	Number of consecutive pen-down segments
55	AreaY5	Numeric Integration of Y values for 4th one-fifth time period
54	AreaY4	Numeric Integration of Y values for 4th one-fifth time period
86	StartEndRatio	Distance (Start Point; End Point) * 1000 DIV PathLength
26	Vx_TN	Normalised Average velocity in x direction in pixels / VxMax * 1000
80	NoIntersections_X4	Number of intersections of the vertical line X4 with the sample
51	AreaY1	Numeric Integration of Y values for 1st one-fifth time period
63	AreaRatio2	Area(ConvexHull(Segments)) vs Area(ConvexHull(Sample)) * 1000
3	AspectRatio	Image Width * 1000 DIV Height
50	AreaX5	Numeric Integration of X values for 4th one-fifth time period
87	XMaxXMinRatio	Distance (Maximum X Point; Minimum X Point) * 1000 DIV PathLength
97	dotcnt_avg	average Number of neighbour Points
46	AreaX1	Numeric Integration of X values for 1st one-fifth time period
89	StartCentroidEndRatio	Distance (Start Point; Centroid) * 1000 DIV Distance (End Point; Centroid)
123	Cluster3XYAvgPrs	average pressure of xy cluster 3
47	AreaX2	Numeric Integration of X values for 2nd one-fifth time period
4	VxAbsolute	Average velocity in x direction in 1000 * pixels / ms
118	Cluster6X	normalized X-coordinate of cluster 5
66	PathRatio2	PathLength(ConvexHull(Segments)) vs PathLength(ConvexHull(Sample)) * 1000
114	Cluster4X	normalized X-coordinate of cluster 3
18	CentroidAzimuth_SN	Horizontal azimuth of centroid from origin normalised to PI/2 * 1000
131	PressureDeviation	standard deviation of the Pressure

Betrachten wir die Secure Sketch-Ergebnisse basierend auf den jeweils ersten 60 Merkmalen der selektierten Sets in Tabelle 7.14, ist erkennbar, dass hier ebenfalls Verbesserungen im Vergleich zum kompletten Merkmal-Set erreicht werden. In diesem Fall wird das beste Ergebnis für die Verifikation mit einer Equal Error Rate von 0,05074 für die Semantik *Woher* bestimmt. Dieses resultiert aus dem von Filter *entropy* erzeugten Merkmal-Set und liegt damit geringfügig über der durch *minEER* erreichten EER von 0,04823 basierend auf 70 Merkmalen (siehe Tabelle 7.12). Die kleinste Collision Reproduction Rate wird auch für diese Konstellation für das *Symbol* berechnet. Im Vergleich zur besten CRR für den Secure Sketch-Algorithmus (siehe Tabelle 7.12, *Symbol*, (b) *bestCRR*, CRR=0,02848) ist dieser mit 0,06475 mehr als doppelt so hoch. Grund dafür ist die höhere Anzahl der Merkmale. Während die ersten 22 identisch sind, kommen hier noch 38 hinzu.

Um die Eignung der von uns für die Merkmalsanalyse und -selektion verwendeten Verfahren für die handschriftbasierte Biometrie eingehender zu prüfen, haben wir aufbauend auf den Ergebnissen weitere Untersuchungen durchgeführt. In den folgenden Abschnitten gehen wir auf die Rankings der Merkmale ein, die durch die Wrapper und Filter bei der Analyse ermittelt wurden. Dabei beschränken wir uns allein auf die Unterscheidung zwischen Wrapper und Filter und gehen nicht auf die einzelnen Verfahren ein. Die Rankings basieren auf der durch die Merkmalsanalyse ermittelten Reihenfolge der Merkmale. Diese werden für die Wrapper beziehungsweise Filter summiert und auf eine Rangliste im Bereich [1,131] abgebildet. Die Diagramme in den folgenden Abbildungen 7.7 bis 7.17 zeigen die Resultate dieses Vorgehens. Die einzelnen Diagramme stellen die Ranking-Werte pro Merkmal für die fünf Semantiken als gestapelte Balken dar. Da es sich dabei um eine aufsteigendes Ranking handelt, ist zu beachten, dass je geringer der Höhe eines Balkens oder Abschnittes ist, umso besser eignet sich das Merkmal für die betreffende

Tabelle 7.14: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,20977	0,50000	0,00000	0,00000	60	0,15475	0,50000	0,00000	0,00000
anova	60	0,20325	0,50000	0,00000	0,00000	60	0,14728	0,50000	0,00000	0,00000
anova-2class	60	***0,16522***	0,50000	0,00000	0,00000	60	***0,10782***	0,50000	0,00000	0,00000
correlation	60	0,21502	0,50000	0,00000	0,00000	60	0,17324	0,50000	0,00000	0,00000
entropy	60	0,16584	0,50000	0,00000	0,00000	60	0,12116	0,50000	0,00000	0,00000
joint-entropy	60	0,19374	0,50000	0,00000	0,00000	60	0,14111	0,50000	0,00000	0,00000
simple	60	**0,16533**	0,50000	0,00000	0,00000	60	0,12227	***0,49906***	0,00189	0,00000
sfs	60	0,16749	0,49906	0,00189	0,00000	60	**0,11727**	0,50000	0,00000	0,00000
sbs	60	0,17423	**0,49623**	0,00755	0,00000	60	0,13614	0,50000	0,00000	0,00000

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,09732	0,41894	0,16226	0,00015	60	0,07751	0,17462	0,66981	0,01905
anova	60	0,10315	**0,42077**	0,15849	0,00004	60	0,06761	**0,13545**	0,74340	0,01430
anova-2class	60	0,09403	0,42266	0,15472	0,00004	60	**0,06356**	0,17861	0,65094	0,00816
correlation	60	0,09703	0,43022	0,13962	0,00007	60	0,07685	0,16646	0,67736	0,01027
entropy	60	**0,08889**	0,43115	0,13774	0,00004	60	0,08472	0,23197	0,54340	0,00733
joint-entropy	60	0,09680	0,42549	0,14906	0,00004	60	0,07808	0,18168	0,65660	0,01996
simple	60	0,09990	0,43304	0,13396	0,00004	60	0,05781	0,19118	0,62453	0,00689
sfs	60	***0,08432***	0,36515	0,26981	0,00011	60	***0,05313***	0,11593	0,77925	0,01110
sbs	60	0,09143	***0,32377***	0,35283	0,00036	60	0,06918	***0,06475***	0,88868	0,01818

	Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,09297	0,50000	0,00000	0,00000
anova	60	0,08291	**0,49811**	0,00377	0,00000
anova-2class	60	0,06689	0,49906	0,00189	0,00000
correlation	60	0,09588	0,49906	0,00189	0,00000
entropy	60	***0,05074***	0,50000	0,00000	0,00000
joint-entropy	60	0,09977	**0,49811**	0,00377	0,00000
simple	60	**0,06453**	0,49906	0,00189	0,00000
sfs	60	0,06785	0,49245	0,01509	0,00000
sbs	60	0,07926	***0,49151***	0,01698	0,00000

Konstellation. Tabellen mit den einzelnen Ranking-Werten für Wrapper und Filter sind im Anhang in den Tabellen C.21 bis C.31 zu finden.

Die Abbildungen 7.7 bis 7.17 zeigen, dass beide Klassen von Analyseverfahren die 13 Merkmale mit einer Varianz von Null[1] identifizieren und entsprechend am Ende des Ranking platzieren. Die geringen Abweichungen in der Darstellung kommen zustande, da die Merkmale entsprechend ihrer Reihenfolge platziert werden.

7.3.3 Wrapper-basiertes Ranking

Im folgenden betrachten wir das Ranking der Merkmale entsprechend der verwendeten Kategorien zur Merkmalsanalyse. Der Rang eines Merkmals n_i wird entsprechend der durch das jeweilige

[1] Merkmale n_{19}, n_{20}, n_{21}, n_{22}, n_{24}, n_{25}, n_{57}, n_{108}, n_{109}, n_{120}, n_{127}, n_{128}, n_{129}, siehe auch 5.3.1 *Varianzbasierte Vorselektion*

Analyseverfahren bestimmten Wert ermittelt. Der durch die Wrapper bestimmte Rang ergibt sich aus der Reihenfolge der Selektion der Merkmale. Für die Filter wird der Rang entsprechend des jeweils errechneten Wertes $R(i)$ (vergleiche Abschnitt 2.1.5 *Optimierung basierend auf der Merkmalsebene*) bestimmt. Dabei erhält das Merkmal mit dem größten Rankingwert R den Rang 1, das mit dem zweitgrößten den Rang 2 und so weiter. Die so ermittelten Ränge werden für Wrapper und Filter getrennt für jede Semantik einzeln merkmalweise aufsummiert. Die aus dieser Vorgehensweise entstandenen Rankings sind im Anhang C *Evaluierungsergebnisse* in den Tabellen C.21 bis C.31 angegeben. Eine kompakte Visualisierung der Ergebnisse der Untersuchung des beschriebenen Rankings zeigen jeweils in den folgenden Abschnitten die Abbildungen 7.7 bis 7.17. Dabei sind die Diagramme wie folgt aufgebaut: Die Abszissenachse zeigt aufsteigend den Index i der Merkmale n_i, in unserem Fall also von 1 bis 131. Die Ordinatenachse gibt den kumulierten Ranking-Wert pro Merkmal für die fünf Semantiken wieder. So ergibt sich beispielsweise für Merkmal n_1 in Abbildung C.21 ein kumulierter Ranking-Wert von 431 basierend auf den Rängen 106 (*gegebene PIN*), 102 (*geheime PIN*), 77 (*Pseudonym*), 83 (*Symbol*) und 63 (*Woher*).

Wie im Abschnitt 5.3 *Analyse und Selektion von Merkmalen (A.3)* beschrieben, benötigen Wrapper zur Ermittlung von optimalen Merkmal-Sets den jeweilig verwendeten Algorithmus, um die entsprechende Selektion vornehmen zu können. Aus diesem Grund werden an dieser Stelle die Ergebnisse für die fünf Semantiken basierend auf den beiden Algorithmen und deren jeweiligen fünf Parametrisierungen betrachtet.

Testergebnisse für den Biometric Hash-Algorithmus

In den Abbildungen 7.7 bis 7.11 werden die Ergebnisse der Untersuchung des Rankings basierend auf dem Biometric Hash-Algorithmus und den fünf Parameter-Sets dargestellt. Auch hier zeigen die Ergebnisse, dass die Eignung der Merkmale sowohl von der Parametrisierung als auch von der jeweiligen Semantik abhängig sind. So schneidet beispielsweise Merkmal n_{95} (*minimum Number of Neighbour Points*) mit den Ranking-Werten sechs (*gegebene PIN*), zwei (*geheime PIN*), drei (*Pseudonym*), vier (*Symbol*) und 42 (*Woher*) für das Merkmal-Set *Null* im Schnitt am besten ab (siehe Abbildung 7.7).

Suchen wir nach dem jeweiligen Merkmal, welches den besten Rang für die einzelnen Semantiken ermittelt, stellen wir fest, dass dieser für *geheime PIN* und *Symbol* für das Merkmal n_{89} (*Distance (Start Point; Centroid) * 1000 DIV Distance (End Point; Centroid)*) mit Rang fünf ermittelt wird. Mit Ranking-Platz elf folgt die *gegebene PIN* mit n_{124} (*average Pressure of xy Cluster 4*), während sich das *Pseudonym* mit Merkmal n_{64} (*Area(ConvexHull(Segments)) vs Area(BoundingBox) * 1000*) auf 31 platziert. Aus diesem Blickpunkt schneidet *Woher* mit dem Merkmal n_{73} (*Number of minimum Points in the Y signal*) am schlechtesten ab (Rang 85).

Auch in dieser Evaluierungskonstellation werden die 13 Merkmale ohne Varianz auf den letzten Plätzen gelistet. Dadurch wird deutlich, dass die Wrapper in der Lage sind, diese Merkmale als unbrauchbar zu einzustufen. Die Platzierung dieser Merkmale auf den hinteren Rängen setzt sich auch für die anderen vier Parameter-Sets fort. Allerdings kommt es bei *minCRR*

(Abbildung 7.9), $RR \geq 80\%$ (Abbildung 7.10) und $CR \leq 5\%$ (Abbildung 7.11) dazu, dass die Merkmale vereinzelt vor anderen selektiert werden. Werden die korrespondierenden Ergebnisse für die optimale Anzahl von Merkmalen zum Vergleich herangezogen, stellen wir fest, dass die Anzahl selektierter Merkmale in jedem untersuchten Fall unterhalb des Ranking-Wertes liegt. Daher wurden diese Merkmale bei der Ermittlung der jeweils optimalen Sets durch die Wrapper aussortiert.

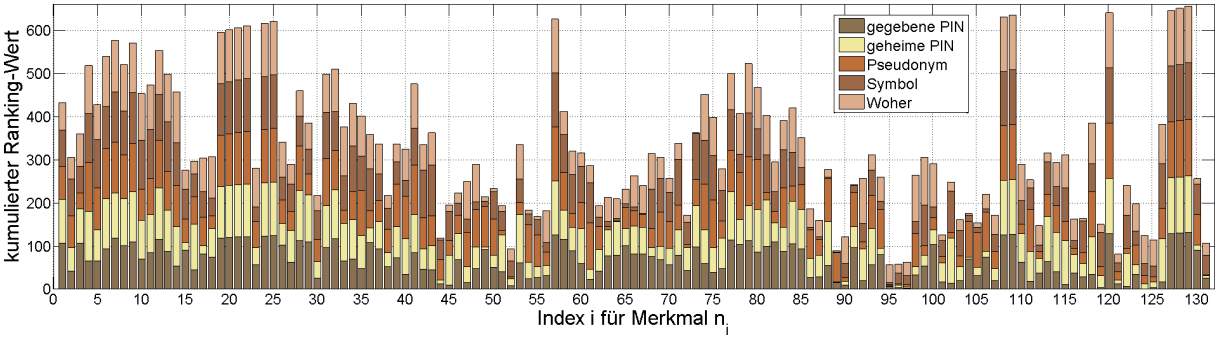


Abbildung 7.7: Biometric Hash: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *Null*

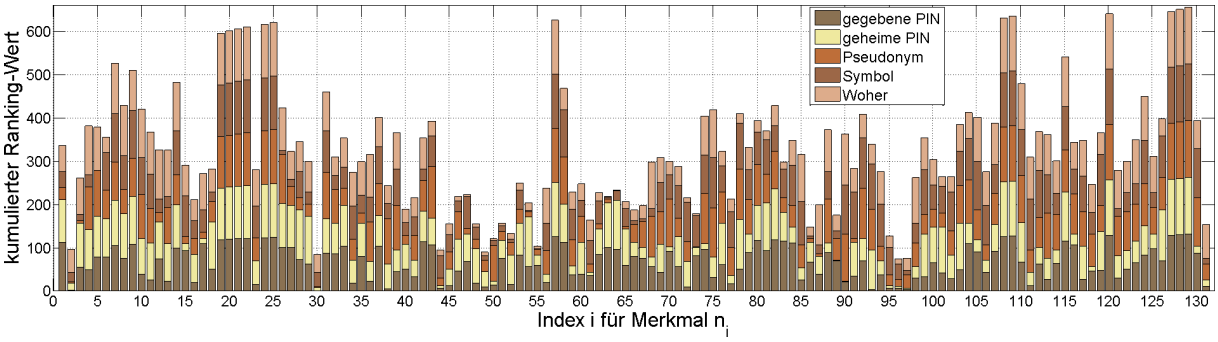


Abbildung 7.8: Biometric Hash: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *minEER*

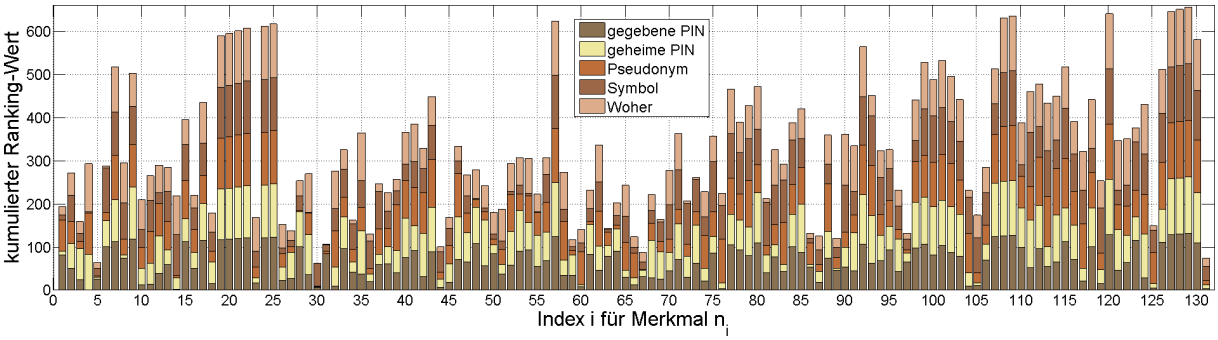


Abbildung 7.9: Biometric Hash: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *minCRR*

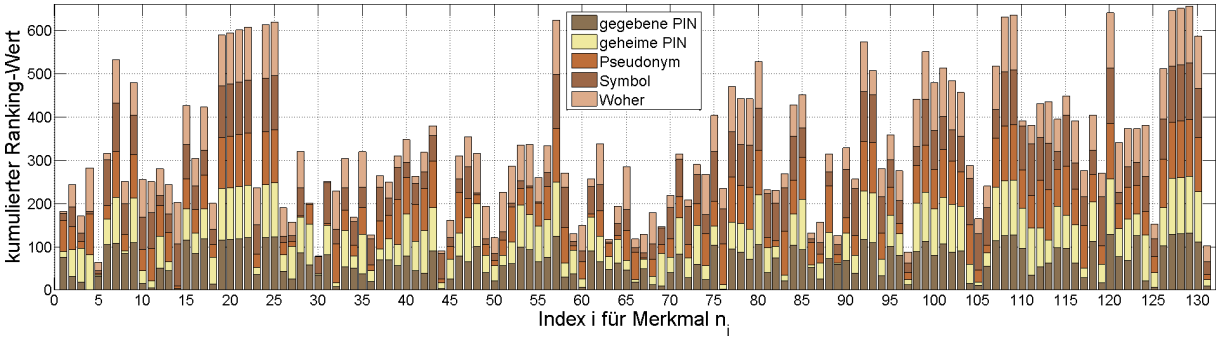


Abbildung 7.10: Biometric Hash: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set $RR \geq 80\%$

Das Merkmal, für welches die meisten ersten Ranking-Plätze bestimmt werden, ist n_{30} (*Total Number of Sample Values*). Dies wird für die Evaluierung des Biometric Hash-Algorithmus unter allen Parameter-Sets und Semantiken sieben mal auf eins gelistet. Das Merkmal n_7 (*Minimum absolute x-velocity during Sample*) erreicht für alle fünf Parameter-Sets jeweils einen kumulierten Ranking-Wert von über 500. Es schneidet folglich sehr häufig mit einem hohen Rang ab, was zur Folge hat, dass es ebenfalls häufig in den optimierten Sets nicht berücksichtigt wird. Der geringste Rang für n_7 wird basierend auf Parameter-Set $CR \leq 5\%$ und Semantik *geheime PIN* mit einem Wert von 73 bestimmt (siehe Abbildung 7.11). Betrachten wir die Standardabweichung, ist eine geringe Varianz zwischen den einzelnen Semantiken erkennbar. Maximal beträgt diese circa 14,02 für die *geheime PIN*, während die *gegebene PIN* mit circa 5,43 die kleinste Standardabweichung aufweist.

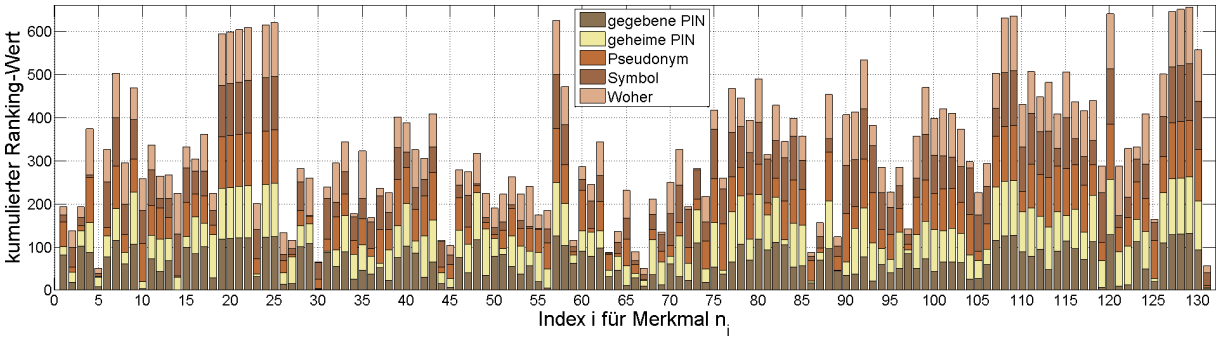


Abbildung 7.11: Biometric Hash: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set $CR \leq 5\%$

Andere Merkmale weisen deutlichere Sprünge bei den ermittelten Ranking-Werten auf, sowohl zwischen den einzelnen Parameter-Sets, als auch zwischen den jeweiligen Semantiken. Beispielhaft kann hier Merkmal n_{73} (*Number of Minimum Points in the Y Signal*) genannt werden, welches maximal um 183 und minimal um 8 Ränge variiert. Die Standardabweichung der über die Resultate der Parameter-Sets summierten Platzierungen beträgt circa 65,68. Betrachten wir die Werte im Kontext der einzelnen Semantiken, zeigt sich, dass die Schwankungen hier ebenfalls unterschiedlich stark ausfallen. Während wir für *geheime PIN* mit rund 31,77 die höchste Standardabweichung bestimmen, weist *Woher* mit 8,95 die geringste auf.

Testergebnisse für den Secure Sketch-Algorithmus

Bei der Betrachtung der Ergebnisse für das Ranking basierend auf dem Secure Sketch-Algorithmus, dargestellt in den Abbildungen 7.12 bis 7.16, können ähnliche Beobachtungen wie beim Biometric Hash-Algorithmus gemacht werden. Hier gibt es ebenfalls teilweise starke Abweichungen zwischen den Semantiken und den einzelnen Merkmalen. Es fällt jedoch auf, dass beispielsweise das Merkmal n_6 (*Number of consecutive pen-down Segments*) für alle fünf Merkmal-Sets im oberen Bereich eingestuft wird. Insgesamt wird es in zehn Fällen auf den ersten Rang gelistet. Merkmal n_6 liegt, basierend auf der Summe der einzelnen Ränge der Semantiken, für *Null* als auch $CR \leq 5\%$ auf Rang drei, für *minEER* auf eins, für *minCRR* auf zwei und für $RR \geq 70\%$ auf elf.

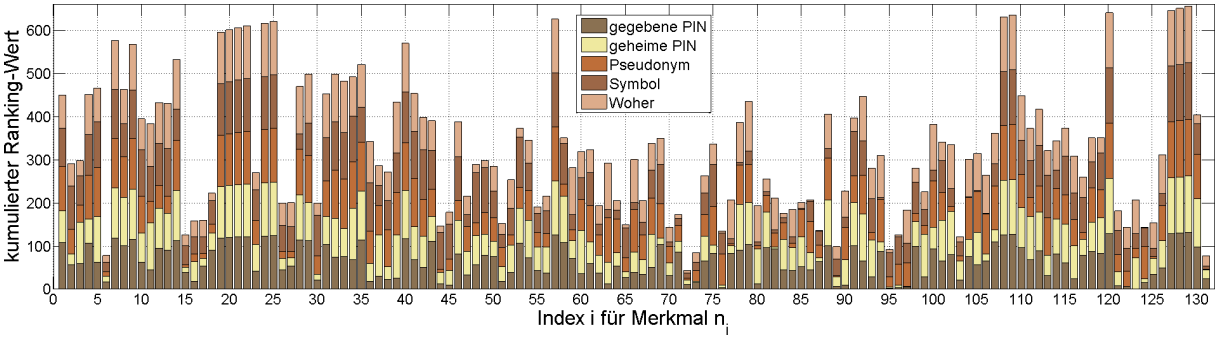


Abbildung 7.12: Secure Sketch: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *Null*

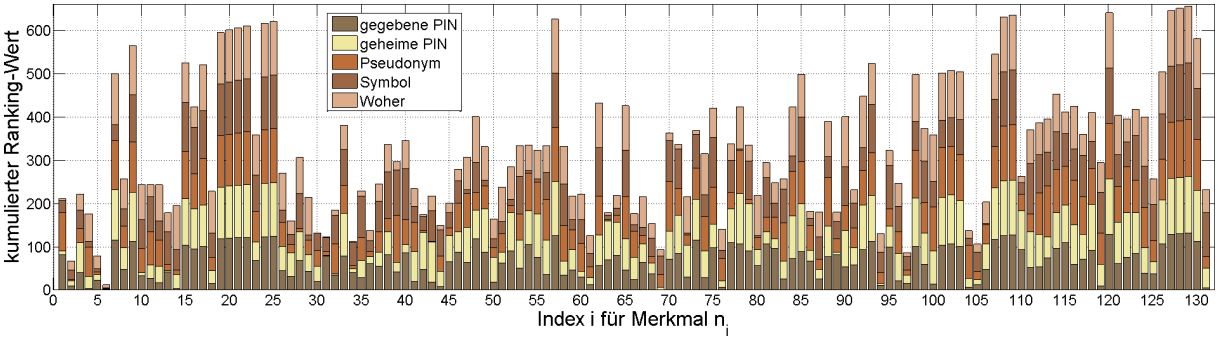


Abbildung 7.13: Secure Sketch: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *minEER*

Betrachten wir die für das Ranking im Zusammenhang mit dem Biometric Hash-Algorithmus beispielhaft detailliert untersuchten Merkmale n_7 und n_{73}, stellen wir fest, dass die Ergebnisse für die hier beschriebene Konstellation anders ausfallen. Einerseits weist Merkmal n_7 (*Minimum absolute x-velocity during Sample*) eine deutlich höhere Varianz in den Standardabweichungen zwischen den Werten der einzelnen Semantiken basierend auf den fünf Parameter-Sets auf. Die größte Standardabweichung wird für das *Symbol* mit 42,56 gemessen, wohingegen die kleinste lediglich 0,84 für die *geheime PIN* beträgt. Auf der anderen Seite zeigen sich für Merkmal n_{73} (*Number of Minimum Points in the Y Signal*) deutlich höhere Werte. Die höchste Standardabweichung ergibt sich hier ebenfalls für die Semantik *geheime PIN*, welche circa 46,15 beträgt.

Die minimale Standardabweichung beträgt etwa 15,47. Bestimmen wir die Standardabweichung für die Ergebnisse der fünf Parameter-Sets ohne einzelne Betrachtung der Semantiken, ergibt sich ein Wert von circa 132,81. Dieser im Vergleich sehr hohe Wert basiert zum einen auf einem niedrigen und einem mittleren über die Semantiken summierten Ranking-Platz von 85 für *Null* (siehe Abbildung 7.12) beziehungsweise 183 für $RR \geq 70\%$ (siehe Abbildung 7.15). Für die verbleibenden drei Parameter-Sets (*minEER*, *minCRR* und $CR \leq 5\%$) werden andererseits hohe summierte Ränge ermittelt. Entsprechend den Abbildungen 7.13, 7.14 und 7.15 betragen diese 368, 373 beziehungsweise 363.

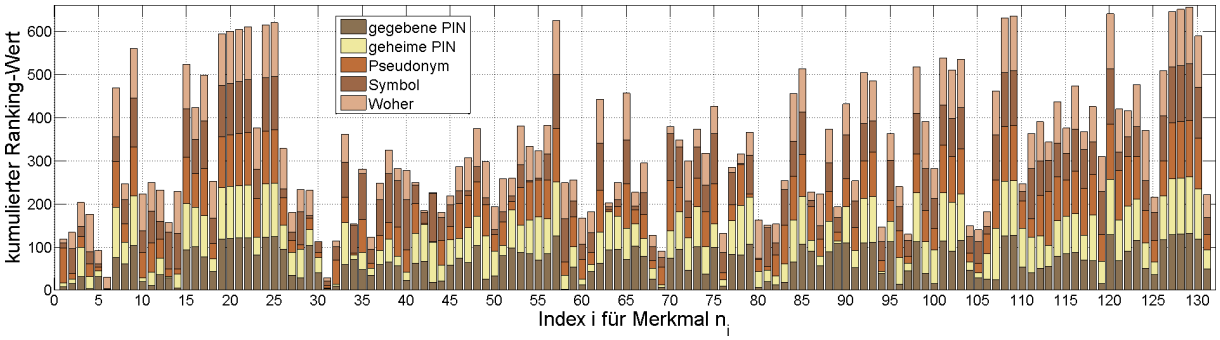


Abbildung 7.14: Secure Sketch: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *minCRR*

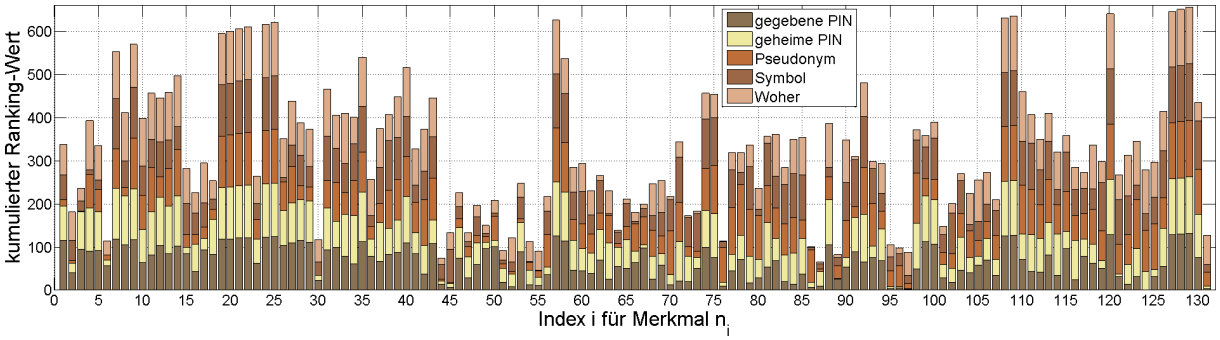


Abbildung 7.15: Secure Sketch: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set $RR \geq 70\%$

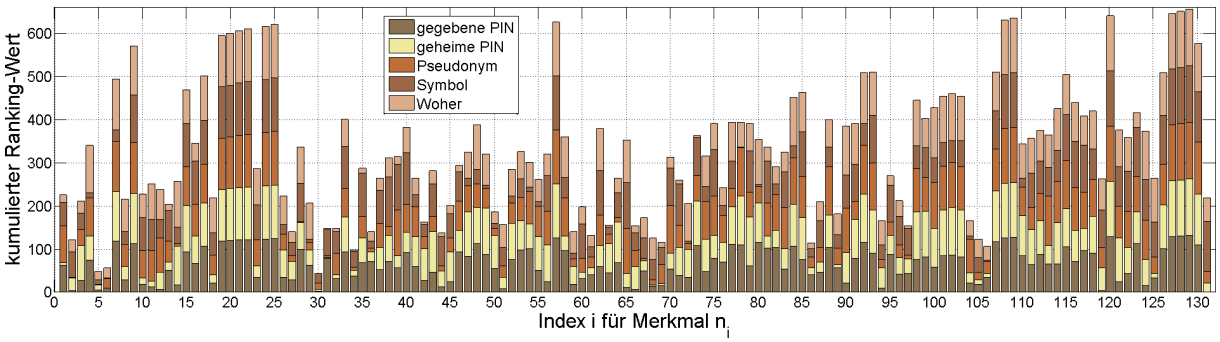


Abbildung 7.16: Secure Sketch: Grafische Darstellung des über die Wrapper-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set $CR \leq 5\%$

7.3.4 Filterbasiertes Ranking

Wie in Abschnitt 5.3 *Analyse und Selektion von Merkmalen (A.3)* beschrieben, ist die Analyse von Merkmalen mittels Filter unabhängig vom verwendeten Algorithmus oder seiner Parametrisierung. Vielmehr basiert sie auf den Eigenschaften der Merkmale beziehungsweise den Beziehungen zwischen ihnen. Das bietet zum einen den Vorteil, dass die Analysen zeitsparender durchgeführt werden können, da die wiederholte Ausführung des Algorithmus mit dem in jedem Schritt aktuellen Merkmal-Set wegfällt, die bei der Nutzung eines Wrapper notwendig ist. Allerdings liegt auf der anderen Seite genau darin auch ein Nachteil, da das auf den jeweiligen Algorithmus bezogene Feedback fehlt.

In der Abbildung 7.17 stellen wir das Ranking vor, das wir für die einzelnen Semantiken über die Resultate der Selektion der fünf Filter ermittelt haben. Dabei wird deutlich, dass es auch hier sowohl Merkmale gibt, die sehr häufig im vorderen Bereich eingestuft wurden, ebenso wie andere Merkmale, die oft schlecht abschneiden. Als bestes Merkmal erweist sich auch hier n_6 (*Number of consecutive pen-down Segments*), welches dreimal Rang eins (*geheime PIN*, *Pseudonym*, *Symbol*) und zweimal Rang zwei (*gegebene PIN*, *Woher*) erreicht.

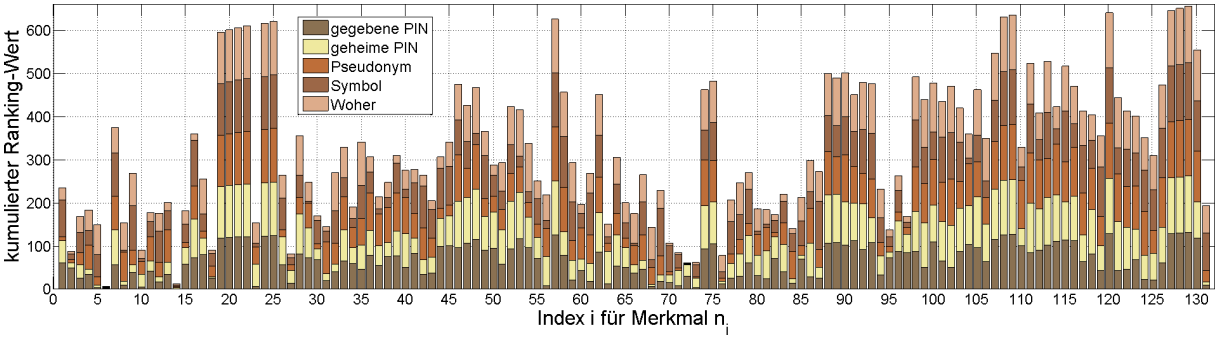


Abbildung 7.17: Grafische Darstellung des über die Filter-Ergebnisse berechneten Rankings für die einzelnen Semantiken und das Parameter-Set *Null*

Ein gutes Ergebnis wird auch für das zweitbeste Merkmal n_{14} (*Maximum absolute Pressure occured during Writing*) mit einem Rang eins (*gegebene PIN*), einem Rang zwei (*geheime PIN*) und dreimal Rang drei (*Pseudonym*, *Symbol*, *Woher*) erreicht. Weniger deutlich fällt der Vergleich beim nächstbesten Merkmal n_{72} (*Number of maximum Points in the Y Signal*) aus. Dieses belegt die Ränge 30 (*gegebene PIN*), 26 (*geheime PIN*), zwei (*Pseudonym*, *Symbol*) und eins (*Woher*). Wie erwartet liegen die 13 Merkmale, deren Varianz in unseren Voruntersuchungen gleich Null war, auf den letzten 13 Rängen.

Bei der Betrachtung der Abbildung 7.17 wird ebenso wie in den vorherigen Abschnitten zur Optimierung durch Merkmalsselektion deutlich, dass die selektierten Merkmale auch von der verwendeten Semantik abhängig sind. Es genügt demzufolge nicht, ein über eine einzelne Semantik bestimmtes Merkmal-Set für andere Schreibinhalte zu nutzen, vielmehr muss hier separat für jede Semantik die entsprechende optimale Menge ermittelt werden.

7.3.5 Häufigkeitsanalyse

In diesem Abschnitt untersuchen wir die Häufigkeiten, mit der die einzelnen Wrapper und Filter das Merkmal-Set bestimmen, welches jeweils in der experimentellen Evaluierung das beste Ergebnis in der Verifikation beziehungsweise Hash-Generierung verursacht hat. Betrachtet werden die einzelnen Fälle für Biometric Hash und Secure Sketch je nach Szenario (*bestEER*, *bestCRR* und *fix60*) und sowohl über alle Semantiken zusammen als auch einzeln. Ziel ist dabei die Ermittlung, welche Verfahren sich besser für die Merkmalsselektion für die dynamische Handschrift eignen beziehungsweise welche weniger. Auch wenn sich diese Aussagen momentan noch auf die Ergebnisse der Evaluierungsszenarien dieser Arbeit beziehen, können sie hilfreich bei der Wahl von Wrappern und Filtern in zukünftigen Forschungsvorhaben sein.

Abbildung 7.18 zeigt in einem Histogramm, mit welcher Häufigkeit die einzelnen Verfahren zur Merkmalsselektion das beste Ergebnis für das Testset und die drei Szenarien bestimmt haben. Eine Übersicht der entsprechenden Häufigkeitswerte ist sowohl in Tabelle 7.15, als auch in Abbildung 7.18 dargestellt. Ausgewertet werden hier die beiden Algorithmen Biometric Hash (BH) und Secure Sketch (SeS) im Kontext der Szenarien *bestEER*, *bestCRR* und *fix60*.

An dieser Stelle unterscheiden wir nicht zwischen den einzelnen Semantiken. Die Analyse der Ergebnisse wird hier unabhängig davon durchgeführt, welchen Wert die jeweilige beste Equal Error Rate beziehungsweise Collision Reproduction Rate hat. Wichtig ist in diesem Zusammenhang nur, dass das Resultat in der jeweilig betrachteten Konstellation den geringsten Betrag aufweist. Dabei ist zu beachten, dass es bei identischer Ermittlung des optimalen Merkmal-Sets zu identischen Ergebnissen bei der Evaluierung kommt. Passiert dies für das beste Ergebnis in der aktuellen Konstellation, wird die Häufigkeit für die betreffenden zwei oder mehr Verfahren erhöht.

Die bestimmten Häufigkeiten zeigen deutlich, dass beim Einsatz von Verfahren zur Merkmalsselektion im Kontext unseres Testset und unserer Methodologie die Wrapper erfolgreicher als die Filter sind. Die drei Wrapper *simple*, *sfs* und *sbs* weisen sowohl zusammen, als auch einzeln meist eine weitaus höhere Häufigkeit auf als die Filter beziehungsweise der Ausgangszustand (*raw*) vor Durchführung der Merkmalsselektion.

Für die Wrapper schneidet *simple* am schlechtesten ab, während *sbs* insgesamt die höchste Häufigkeit aufweist. Einerseits ermitteln wir mittels *sfs* häufiger das beste Ergebnis für den Biometric Hash-Algorithmus als mit anderen Verfahren (in vier von sechs Fällen). Zum anderen werden die besten Ergebnisse für den Secure Sketch-Algorithmus am häufigsten durch das vom *sbs*-Wrapper generierte Merkmal-Set bestimmt. Auf einen interessanten Aspekt weist die Tatsache hin, dass für das Szenario *bestCRR* sowohl für den Biometric Hash als auch für den Secure Sketch zweimal das gesamte Merkmal-Set (*raw*) das beste Ergebnis für die Equal Error Rate liefert. Dies ist in zehn (Biometric Hash) beziehungsweise neun (Secure Sketch) Tests der Fall. Daraus kann geschlussfolgert werden, dass die Merkmale, die für die beste Collision Reproduction Rate selektiert wurden, häufig nicht passend für die Bestimmung der Equal Error Rate sind.

Tabelle 7.15: Anzahl der jeweils durch die einzelnen Merkmalsselektionsmethoden für die Algorithmen Biometric Hash und Secure Sketch ermittelten besten Ergebnisse bezüglich der Szenarien $bestEER$, $bestCRR$ und $fix60$, kumuliert über die fünf Semantiken

Szenario	Algorithmus	Rate	raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
bestEER	Biometric Hash	EER	0	0	1	1	1	0	5	11	8
		CRR	0	0	0	0	0	0	3	11	11
	Secure Sketch	EER	0	1	4	0	1	0	4	1	14
		CRR	0	1	1	0	0	0	0	4	13
bestCRR	Biometric Hash	EER	10	0	0	0	1	1	1	11	2
		CRR	0	0	0	0	0	0	2	11	14
	Secure Sketch	EER	9	0	0	0	0	0	0	7	8
		CRR	0	1	1	0	4	0	5	7	18
fix60	Biometric Hash	EER	0	0	1	0	1	0	6	12	5
		CRR	0	0	0	0	0	0	0	14	11
	Secure Sketch	EER	0	0	3	0	1	0	5	9	7
		CRR	0	0	0	0	0	0	1	0	19

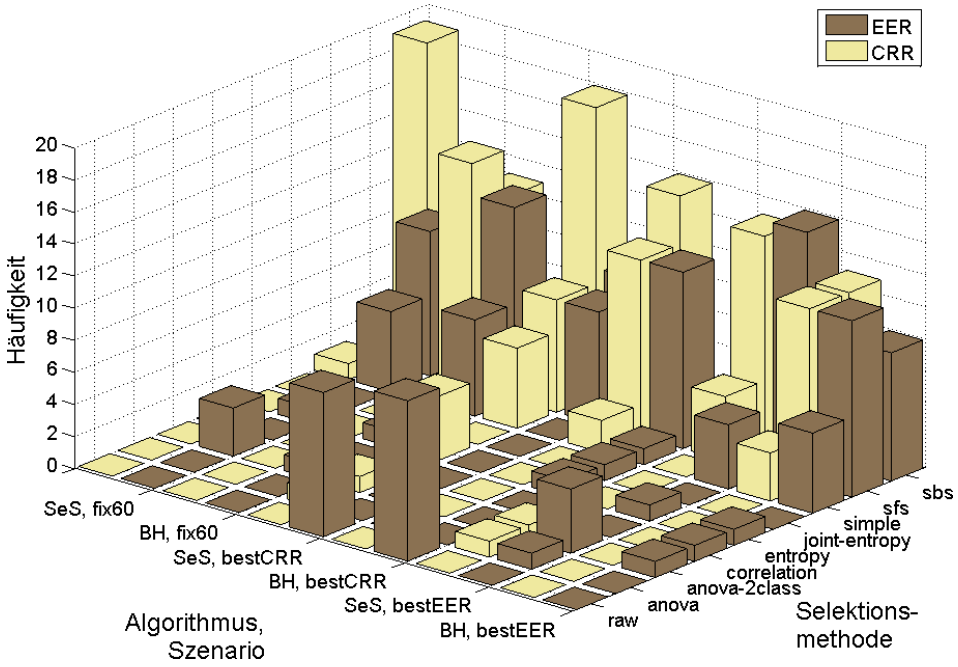


Abbildung 7.18: Grafische Darstellung der Anzahl der jeweils durch die einzelnen Merkmalsselektionsmethoden für die Algorithmen Biometric Hash (BH) und Secure Sketch (SeS) ermittelten besten Ergebnisse bezüglich der Szenarien $bestEER$, $bestCRR$ und $fix60$, kumuliert über die fünf Semantiken

Die Abbildungen 7.19 und 7.20 zeigen den Erfolg der einzelnen Verfahren zur Merkmalsselektion durch die Darstellung der Häufigkeiten mit denen diese das beste Ergebnis in den Modi Verifikation beziehungsweise Hash-Generierung in Bezug auf Equal Error Rate und Collision Reproduction Rate erzielt haben. An dieser Stelle betrachten wir die Resultate in Abhängigkeit von einzelnen Semantiken, Selektionsverfahren und Szenarien jeweils unter Verwendung beider Algorithmen.

Die Ergebnisse der Merkmalsselektion für den Biometric Hash-Algorithmus, welche für das jeweils beste Verifikationsresultat ermittelt wurden (siehe Abbildung 7.19a), zeigen ebenfalls, dass hier die Wrapper erfolgreicher abschneiden. Das ist gut an der Häufung in den letzten drei Reihen des Histogramms (*simple*, *sfs*, *sbs*) zu erkennen. Am erfolgreichsten für den Biometric Hash-Algorithmus erweist sich die Strategie *sfs* mit einer Häufigkeit von 22 gefolgt von *sbs* mit 19 und *simple* mit acht. In nur drei Fällen basiert das beste Ergebnis auf einem der fünf Filter. Auch hier zeigt sich, dass im Szenario *bestCRR* die besten Ergebnisse zum Teil mittels des gesamten Merkmal-Set bestimmt werden (siehe Abbildung 7.19b, *raw*).

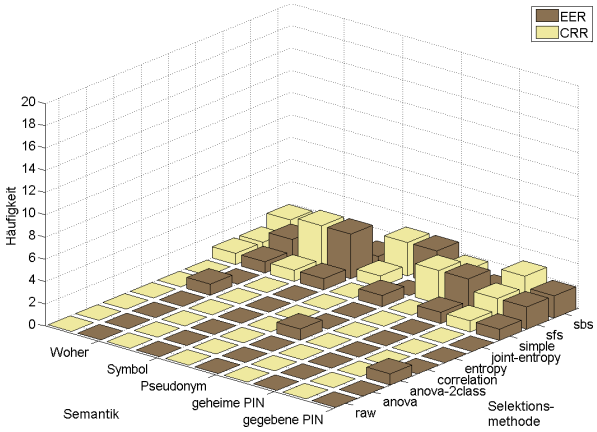
Ähnliche Resultate finden wir auch bei der Betrachtung der Ergebnisse des Secure Sketch-Algorithmus in Abbildung 7.20. Auch hier weist *raw* im Szenario *bestCRR* einige der besten Ergebnisse auf (siehe Abbildung 7.20b, *raw*), wobei die Anzahl pro Semantik zwei beträgt. Eine Ausnahme bildet die Semantik *Symbol* mit eins. Wie schon angedeutet, liefert der Wrapper *simple* am häufigsten die besten Ergebnisse für den Secure Sketch. Dieser weist für die drei Szenarien in 27, 26 beziehungsweise 26 Fällen deutlich öfter das beste Resultat auf als die alternativen Wrapper oder Filter.

7.3.6 Fazit - Evaluierung: A.3 Analyse und Selektion von Merkmalen

Bei der Betrachtung wird deutlich, dass basierend auf den zugrunde liegenden Testdaten und der verwendeten Methodologie die durch die Wrapper selektierten Merkmale wesentlich häufiger die besten Ergebnisse erzielen als die Filter. Einen Hauptgrund stellt dabei die Verwendung des jeweiligen Algorithmus zur Selektion der Merkmale dar. Dies trägt auf der anderen Seite aber auch zum Teil zu einem erheblichen Nachteil der wrapper-basierten Verfahren bei, da für jede ermittelte Menge an Merkmalen der Algorithmus für das Testset ausgeführt werden muss.

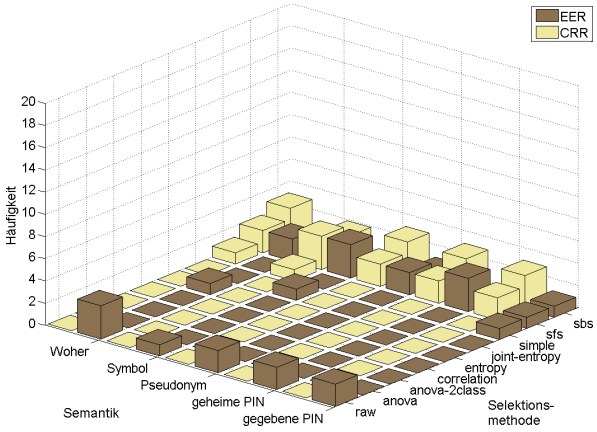
Die kleinste im Verifikationsmodus durch die Merkmalsselektion erreichte Equal Error Rate beträgt 0,03286 und resultiert aus der Kombination von Secure Sketch-Algorithmus, Semantik *Symbol*, Parameter-Set *minEER* ermittelt durch Wrapper *sbs* und Szenario *bestEER*. Das diesem Ergebnis zugrunde liegende Merkmal-Set besteht aus 26 Elementen. Für den Biometric Hash-Algorithmus wird die kleinste EER ebenfalls für das *Symbol* und Parameter-Set *minEER* bestimmt. Weiterhin basiert sie auf 79 Merkmalen, die durch den Wrapper *sfs* zusammengestellt werden. Mit einer Collision Reproduction Rate von 0,02848 für das *Symbol* berechnet der Secure Sketch-Algorithmus das beste Ergebnis innerhalb der Evaluierung der Merkmalsselektion im Hash-Generierungsmodus. Es beruht auf einem vom Wrapper *sbs* selektierten Merkmal-Set mit 22 Elementen, Parameter-Set minEER und dem Szenario *bestCRR*. Die CRR wird für diese Konstellation berechnet aus einer Reproduction Rate von 0,98491 und einer Collision Rate von 0,04187. Eine etwas geringere Reproduction Rate von 0,95660 und eine Collision Rate von 0,04336 bilden die Grundlage der kleinsten CRR für den Biometric Hash-Algorithmus im Kontext der Merkmalsselektion. Sie beträgt mit 0,04338 etwa das 1,5-fache der kleinsten ermittelten Collision Reproduction Rate.

Die Betrachtung der Reihenfolge der selektierten Merkmale pro Wrapper und Filter zeigen, dass



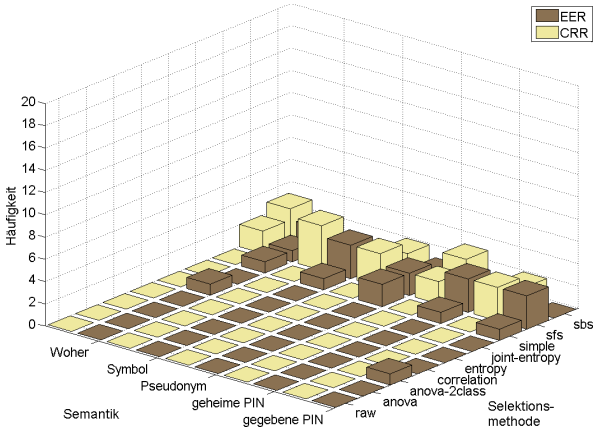
		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	0	0	1	0	0	0	1	2	2
	CRR	0	0	0	0	0	0	1	2	3
geheime PIN	EER	0	0	0	0	0	0	1	3	1
	CRR	0	0	0	0	0	0	0	3	2
Pseudonym	EER	0	0	0	1	0	0	1	0	3
	CRR	0	0	0	0	0	0	0	1	3
Symbol	EER	0	0	0	0	0	0	1	4	1
	CRR	0	0	0	0	0	0	1	4	1
Woher	EER	0	0	0	0	1	0	1	2	1
	CRR	0	0	0	0	0	0	1	1	2

(a) bestEER



		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	2	0	0	0	0	0	1	1	1
	CRR	0	0	0	0	0	0	0	2	3
geheime PIN	EER	2	0	0	0	0	0	0	3	0
	CRR	0	0	0	0	0	0	0	2	3
Pseudonym	EER	2	0	0	0	0	0	0	2	1
	CRR	0	0	0	0	0	0	0	2	3
Symbol	EER	1	0	0	0	0	1	0	3	0
	CRR	0	0	0	0	0	0	1	3	2
Woher	EER	3	0	0	0	1	0	0	2	0
	CRR	0	0	0	0	0	0	1	2	3

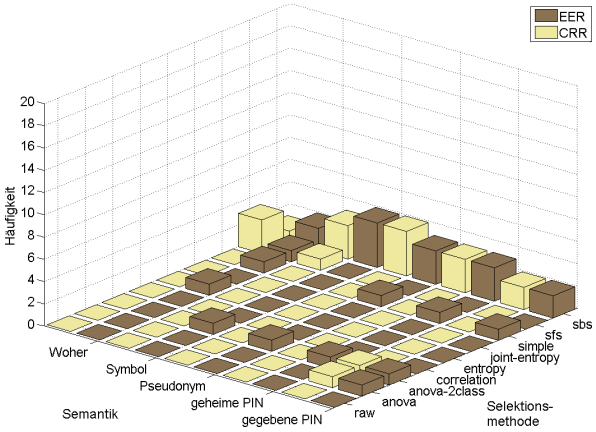
(b) bestCRR



		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	0	0	1	0	0	0	1	3	0
	CRR	0	0	0	0	0	0	0	3	2
geheime PIN	EER	0	0	0	0	0	0	1	3	1
	CRR	0	0	0	0	0	0	0	2	3
Pseudonym	EER	0	0	0	0	0	0	2	2	1
	CRR	0	0	0	0	0	0	0	3	2
Symbol	EER	0	0	0	0	0	0	1	3	1
	CRR	0	0	0	0	0	0	0	4	1
Woher	EER	0	0	0	0	1	0	1	1	2
	CRR	0	0	0	0	0	0	0	2	3

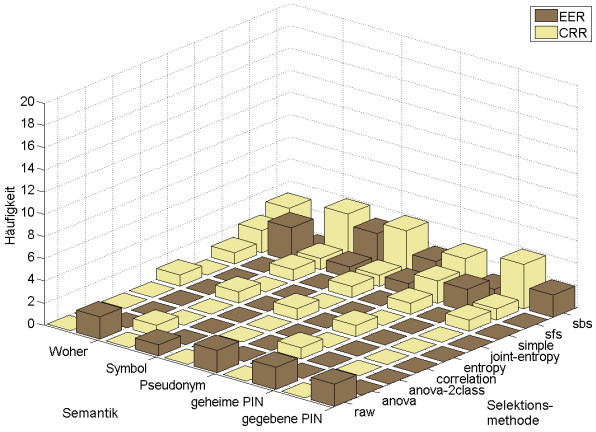
(c) fix60

Abbildung 7.19: Biometric Hash: Anzahl der durch die einzelnen Verfahren zur Merkmalsselektion ermittelten besten Ergebnisse bezüglich der Szenarien (a) bestEER, (b) bestCRR und (c) fix60, jeweils für die fünf Semantiken



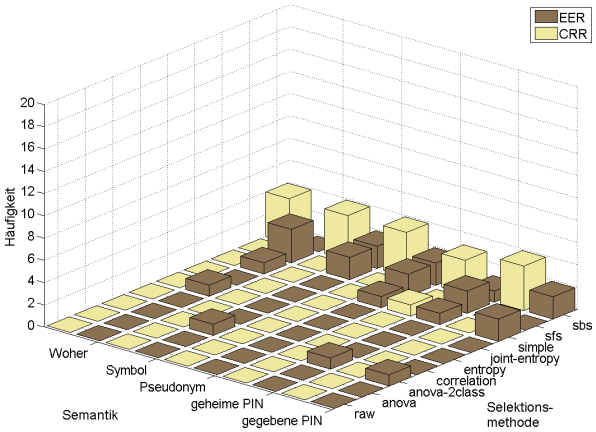
		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	0	1	1	0	0	0	1	0	2
	CRR	0	1	1	0	0	0	0	0	2
geheime PIN	EER	0	0	1	0	0	0	1	0	3
	CRR	0	0	0	0	0	0	0	0	3
Pseudonym	EER	0	0	1	0	0	0	1	0	3
	CRR	0	0	0	0	0	0	0	0	4
Symbol	EER	0	0	1	0	0	0	0	0	4
	CRR	0	0	0	0	0	0	0	1	3
Woher	EER	0	0	0	0	1	0	1	1	2
	CRR	0	0	0	0	0	0	0	3	1

(a) bestEER



		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	2	0	0	0	0	0	0	0	2
	CRR	0	0	0	0	0	0	1	1	4
geheime PIN	EER	2	0	0	0	0	0	0	2	1
	CRR	0	0	1	0	1	0	1	2	3
Pseudonym	EER	2	0	0	0	0	0	0	1	2
	CRR	0	0	0	0	1	0	1	1	4
Symbol	EER	1	0	0	0	0	0	0	1	3
	CRR	0	1	0	0	1	0	1	1	4
Woher	EER	2	0	0	0	0	0	0	3	0
	CRR	0	0	0	0	1	0	1	2	3

(b) bestCRR



		raw	anova	anova-2class	correlation	entropy	joint-entropy	simple	sfs	sbs
gegebene PIN	EER	0	0	1	0	0	0	2	0	2
	CRR	0	0	0	0	0	0	0	0	4
geheime PIN	EER	0	0	1	0	0	0	1	2	1
	CRR	0	0	0	0	0	0	1	0	3
Pseudonym	EER	0	0	0	0	0	0	1	2	2
	CRR	0	0	0	0	0	0	0	0	4
Symbol	EER	0	0	1	0	0	0	0	2	2
	CRR	0	0	0	0	0	0	0	0	4
Woher	EER	0	0	0	0	1	0	1	3	0
	CRR	0	0	0	0	0	0	0	0	4

(c) fix60

Abbildung 7.20: Secure Sketch: Anzahl der durch die einzelnen Verfahren zur Merkmalsselektion ermittelten besten Ergebnisse bezüglich der Szenarien (a) bestEER, (b) bestCRR und (c) fix60, jeweils für die fünf Semantiken

einige Merkmale deutlich häufiger in die selektierten Merkmal-Sets aufgenommen werden als andere. Für die Wrapper basierend auf dem Biometric Hash-Algorithmus wird beispielsweise Merkmal n_{30} (*Total Number of Sample Values*) in sieben Fällen auf Rang eins gelistet, beim Secure Sketch-Algorithmus schneidet Merkmal n_6 (*Number of consecutive pen-down Segments*) mit zehn mal Rang eins am besten ab. Für die Filter ist ebenfalls n_6 mit drei ersten und zwei zweiten Plätzen sehr erfolgreich.

7.4 Evaluierung: A.4 Biometrische Fusion

Im folgenden gehen wir auf die Ergebnisse ein, die durch die Fusion einer Auswahl von biometrischen Komponenten ermittelt wurden. Die von uns getroffene Auswahl der single-modalen Fusionskomponenten und die Bestimmung der jeweiligen Fusionsparameter sind im Abschnitt 5.4 *Biometrische Fusion (A.4)* beschrieben. Die hier untersuchten single-modalen Fusionsstrategien basieren auf der in Abschnitt 6.2.4 *Setup zu A.4 Biometrische Fusion* beschriebenen Vorgehensweise. Wiederum dienen die Samples der ersten und dritten Session der Generierung der Referenzen beziehungsweise als Testdaten für die experimentelle Evaluierung. Die Bestimmung der Gewichte als Fusionsparameter (vergleiche 5.4) erfolgt mittels der Referenzen und Daten aus Session 2, welche als Testdaten zur Ermittlung der Equal Error Rate der einzelnen Fusionskomponenten genutzt werden.

7.4.1 Multi-algorithmic Fusion

In diesem Abschnitt diskutieren wir die Ergebnisse der multi-algorithmischen Fusionen. Grundlage bilden hier die beiden Algorithmen Biometric Hash und Secrue Sketch. Tabelle 7.16 stellt die Gewichte für die einzelnen Strategien dar. Die erste Spalte gibt die jeweilige Semantik an, während die zweite die verwendeten Fusionskomponenten zeigt (hier beide Algorithmen). Spalte *EER* enthält die jeweils mittels Testdaten aus Session 2 ermittelte Equal Error Rate. Die folgenden drei Spalten geben die jeweils basierend auf den EER bestimmten Gewichte entsprechend der einzelnen Fusionsstrategien wieder. In Tabelle 7.17 sind die Ergebnisse dargestellt für die Fusion der beiden Verfahren auf Matching Score Level basierend auf diesen vorab ermittelten algorithmenspezifischen Gewichten (siehe Tabelle 7.16). Da auf diesem Level nur die Bestimmung der Performanz des Verifikationsmodus sinnvoll ist, ist die jeweilige Equal Error Rate (EER) angegeben (siehe auch Abschnitt 5.4.2 *Handschriftbasierte multi-algorithmic Fusion auf Matching Score Level*). Aus dem gleichen Grund haben wir für die Parametrisierung der Algorithmen jeweils lediglich den Tolerance Factor beziehungsweise den Expansion Factor gewählt, welcher für die minimale EER ermittelt wurde (Parameter-Set *minEER*). Die Spalten zwei und drei zeigen die Ergebnisse der beiden Algorithmen ohne Fusion, während die letzten drei Spalten die jeweiligen Ergebnisse der drei Fusionsstrategien darstellen.

Auf der einen Seite kann festgestellt werden, dass im Vergleich der Algorithmen untereinander der Secure Sketch für jede der fünf Semantiken am besten abschneidet. Zusätzlich wird durch ihn für die Semantiken *gegebene PIN* und *Woher* auch das jeweils beste Ergebnis im Vergleich zu den

Fusionen berechnet. Eine eher geringfügige Verbesserung des Verifikationsergebnisses kann für die *geheime PIN* mit 0,10498 durch die gleich gewichtete Fusion erreicht werden. Die Ergebnisse der einzelnen Algorithmen weisen hier Werte von 0,12790 (Biometric Hash) beziehungsweise 0,10927 auf. Für das *Pseudonym* liefert die lineare Strategie mit einer EER von 0,08570 (vor Fusion 0,08943 beziehungsweise 0,08757) das beste Ergebnis und für das *Symbol* die quadratische Fusion mit 0,07047 (vor Fusion 0,07366 beziehungsweise 0,07259).

Tabelle 7.16: Gewichte für multi-algorithmic Fusion basierend auf den Einzelergebnissen beider Algorithmen unter Verwendung der Testdaten aus Session 2

Semantik	Algorithmus	EER	Gewichte		
			equal	linear	quadratic
gegebene PIN	Biometric Hash	0,13988	0,50000	0,49169	0,48338
	Secure Sketch	0,13531	0,50000	0,50831	0,51662
geheime PIN	Biometric Hash	0,09412	0,50000	0,45066	0,40228
	Secure Sketch	0,07722	0,50000	0,54934	0,59772
Pseudonym	Biometric Hash	0,07876	0,50000	0,45470	0,41014
	Secure Sketch	0,06568	0,50000	0,54530	0,58986
Symbol	Biometric Hash	0,06817	0,50000	0,42121	0,34624
	Secure Sketch	0,04961	0,50000	0,57879	0,65376
Woher	Biometric Hash	0,06482	0,50000	0,44055	0,38275
	Secure Sketch	0,05104	0,50000	0,55945	0,61725

Tabelle 7.17: Ergebnisse der multi-algorithmic Fusion im Vergleich zu den Einzelergebnissen der beiden Algorithmen (jeweils angegeben als EER)

Semantik	EER einzeln		EER Fusion		
	Biometric Hash	Secure Sketch	equal	linear	quadratic
gegebene PIN	0,20082	**0,16211**	0,16472	0,16472	0,16381
geheime PIN	0,12790	0,10927	**0,10498**	0,10519	0,10827
Pseudonym	0,08943	0,08757	0,08604	**0,08570**	0,08664
Symbol	0,07366	0,07259	0,07090	0,07177	**0,07047**
Woher	0,08559	**0,06817**	0,07531	0,07697	0,07838

7.4.2 Single-semantic Fusion

Die Ergebnisse der Fusion von zwei Samples derselben Semantik betrachten wir in diesem Abschnitt. Da auch hier die Fusion nach der Bestimmung des Matching Score erfolgt, kann ebenfalls keine Angabe von Werten für den Hash-Generierungsmodus erfolgen. Aus diesem Grund untersuchen wir hier ebenfalls die Verifikation unter Verwendung des Parameter-Set *minEER*.

Die Tabellen 7.18 und 7.20 zeigen die für die Fusion ermittelten Gewichte für den Biometric Hash- beziehungsweise Secure Sketch-Algorithmus. Spalte eins enthält die Semantik, die zweite Spalte verweist auf die Instanz, während die dritte die zugehörige Equal Error Rate angibt, die mittels der Testdaten aus Session 2 ermittelt wurden. In den letzten drei Spalten folgen die Werte für die einzelnen Wichtungsstrategien.

In den Tabellen 7.19 beziehungsweise 7.21 werden die entsprechenden Ergebnisse der Fusion dargestellt. Hier gibt die erste Spalte die Semantik an und die zweite Spalte (*single*) zeigt das

Ergebnis bei der Verwendung einer einzelnen Instanz. Die folgenden drei Spalten enthalten die Ergebnisse für die drei Fusionsstrategien.

Tabelle 7.18: Biometric Hash: Gewichte für single-semantische Fusion basierend auf den Einzelergebnissen beider Instanzen unter Verwendung der Testdaten aus Session 2

Semantik	Instanz	EER	Gewichte equal	linear	quadratic
gegebene PIN	1	0,14013	0,50000	0,49942	0,49883
	2	0,13980	0,50000	0,50058	0,50117
geheime PIN	1	0,09935	0,50000	0,47029	0,44078
	2	0,08821	0,50000	0,52971	0,55922
Pseudonym	1	0,07789	0,50000	0,50309	0,50618
	2	0,07886	0,50000	0,49691	0,49382
Symbol	1	0,06935	0,50000	0,49161	0,48323
	2	0,06706	0,50000	0,50839	0,51677
Woher	1	0,06431	0,50000	0,49250	0,48500
	2	0,06241	0,50000	0,50750	0,51500

Durch die Fusion basierend auf mehreren Samples einer Semantik und dem Biometric Hash-Algorithmus (siehe Tabelle 7.19) kann abhängig von der Strategie in fast jedem Fall eine Verbesserung gegenüber der Verwendung einer einzelnen Probe erreicht werden. Das beste Ergebnis wird für die Semantik *Symbol* mit der gleich gewichteten Fusionsstrategie (siehe Spalte *equal*, EER=0,07183) ermittelt. Das entspricht einer Verbesserung um etwa 2,5% verglichen mit dem entsprechenden Ausgangswert von 0,07366. Die größte Verbesserung zwischen der durch eine einzelne Instanz und eine Fusion erreicht wurde liegt bei circa 6,5% für die *gegebene PIN* und Strategie *equal*. Hier wurde die Equal Error Rate von ursprünglich 0,20082 auf 0,18769 gesenkt. Beim *Symbol* kam es sowohl bei *linear* als auch bei *quadratic* durch die Fusion zu einer Verschlechterung von etwa 1,5%.

Tabelle 7.19: Biometric Hash: Ergebnisse der single-semantic Fusion im Vergleich zu den Einzelergebnissen (jeweils angegeben als EER)

Semantik	EER single	EER Fusion equal	linear	quadratic
gegebene PIN	0,20082	**0,18769**	0,18906	0,18906
geheime PIN	0,12790	**0,12005**	0,12223	0,12227
Pseudonym	0,08943	0,08674	**0,08628**	**0,08628**
Symbol	0,07366	**0,07183**	0,07474	0,07474
Woher	0,08559	0,08302	**0,08250**	**0,08250**

Wie Tabelle 7.21 zeigt, werden durch die single-semantic Fusion basierend auf dem Secure Sketch-Algorithmus im jedem hier untersuchten Fall bessere Ergebnisse erreicht, als bei der Nutzung eines einzelnen Sample. Hier beträgt die beste EER 0,06097, ermittelt für die Semantik *Woher* und die Fusionsstrategie *linear*, was einer Verbesserung um circa 10,6% gegenüber dem Ausgangswert (0,06817) entspricht. Diese Equal Error Rate stellt auch die höchste Verbesserung im Vergleich der fünf Semantiken und drei Fusionsstrategien dar.

Tabelle 7.20: Secure Sketch: Gewichte für single-semantische Fusion basierend auf den Einzelergebnissen beider Instanzen unter Verwendung der Testdaten aus Session 2

Semantik	Instanz	EER	Gewichte equal	linear	quadratic
gegebene PIN	1	0,13064	0,50000	0,50016	0,50033
	2	0,13073	0,50000	0,49984	0,49967
geheime PIN	1	0,07567	0,50000	0,49756	0,49512
	2	0,07493	0,50000	0,50244	0,50488
Pseudonym	1	0,06236	0,50000	0,52246	0,54483
	2	0,06822	0,50000	0,47754	0,45517
Symbol	1	0,05072	0,50000	0,48813	0,47628
	2	0,04837	0,50000	0,51187	0,52372
Woher	1	0,04802	0,50000	0,51962	0,53918
	2	0,05194	0,50000	0,48038	0,46082

Tabelle 7.21: Secure Sketch: Ergebnisse der single-semantischen Fusion im Vergleich zu den Einzelergebnissen (jeweils angegeben als EER)

Semantik	EER single	EER Fusion equal	linear	quadratic
gegebene PIN	0,16211	**0,14937**	0,15262	0,15262
geheime PIN	0,10927	0,10598	**0,10562**	**0,10562**
Pseudonym	0,08757	0,08619	0,08586	**0,08585**
Symbol	0,07259	**0,07074**	0,07119	0,07119
Woher	0,06817	0,06180	**0,06097**	0,06106

7.4.3 Multi-instance Fusion

In diesem Abschnitt beschäftigen wir uns mit der multi-instance Fusion. Hier werden, wie in Abschnitt 5.4.3 *Handschriftbasierte multi-instance Fusion auf Matching Score Level* beschrieben, jeweils unterschiedliche Instanzen der Handschrift einer Person verwendet, wobei diese Instanzen in unserem Kontext Semantiken sind. Daher kann auch von einer multi-semantic Fusion [SVF11] gesprochen werden. Die Untersuchung der multi-instance Fusion ist auf die Verifikation beschränkt (vergleiche Abschnitt 5.4.2 *Handschriftbasierte multi-algorithmic Fusion auf Matching Score Level*).

In den Tabellen 7.22 und 7.24 sind für Biometric Hash und Secure Sketch die Gewichte aufgelistet, die für die einzelnen Strategien entsprechend der Beschreibung in Abschnitt 5.4.1 *Bestimmung von Parametern für die Fusion auf Matching Score Level* bestimmt wurden. Die Zeile *single* gibt die Equal Error Rate für die entsprechende Semantik basierend auf den Testdaten aus Session 2 an. Diese werden genutzt, um die Gewichte für die Fusion zu bestimmen, welche jeweils paarweise in den Spalten angegeben werden. Der Wert direkt in der Spalte unter der Semantik gibt deren Wichtung an, während der rechts daneben stehende Wert das Gewicht zur Semantik in der jeweiligen Zeile darstellt.

Tabelle 7.22: Biometric Hash: Gewichte für multi-instance Fusion basierend auf den Einzelergebnissen beider Semantiken unter Verwendung der Testdaten aus Session 2

Semantik	Fusion	Geg. PIN		Geh. PIN		Pseu-donym		Sym-bol		Woher	
	single	0,13988		0,09412		0,07876		0,06817		0,06482	
gegebene PIN	equal	-	-	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000
	linear	-	-	0,59778	0,40222	0,63977	0,36023	0,67236	0,32764	0,68336	0,31664
	quadr.	-	-	0,68836	0,31164	0,75928	0,24072	0,80811	0,19189	0,82325	0,17675
geheime PIN	equal	-	-	-	-	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000
	linear	-	-	-	-	0,54442	0,45558	0,57997	0,42003	0,59219	0,40781
	quadr.	-	-	-	-	0,58814	0,41186	0,65596	0,34404	0,67832	0,32168
Pseudonym	equal	-	-	-	-	-	-	0,50000	0,50000	0,50000	0,50000
	linear	-	-	-	-	-	-	0,53607	0,46393	0,54857	0,45143
	quadr.	-	-	-	-	-	-	0,57176	0,42824	0,59623	0,40377
Symbol	equal	-	-	-	-	-	-	-	-	0,50000	0,50000
	linear	-	-	-	-	-	-	-	-	0,51259	0,48741
	quadr.	-	-	-	-	-	-	-	-	0,52516	0,47484

Tabelle 7.23: Biometric Hash: Ergebnisse der multi-instance Fusion im Vergleich zu den Einzelergebnissen (jeweils angegeben als EER)

Semantik	Fusion	Gegebene PIN	Geheime PIN	Pseudonym	Symbol	Woher
	single	0,20082	0,12790	0,08943	0,07366	0,08559
gegebene PIN	equal	-	0,11849	0,07832	0,12922	0,06861
	linear	-	0,11576	0,07398	0,10178	0,06674
	quadratic	-	0,11720	0,07774	0,07672	0,07642
geheime PIN	equal	-	-	0,06142	0,08440	0,05780
	linear	-	-	0,06003	0,07544	0,06168
	quadratic	-	-	0,06255	0,06832	0,06805
Pseudonym	equal	-	-	-	0,06310	**0,04612**
	linear	-	-	-	0,06284	0,05140
	quadratic	-	-	-	0,06015	0,05322
Symbol	equal	-	-	-	-	0,06627
	linear	-	-	-	-	0,06774
	quadratic	-	-	-	-	0,06685
Woher	equal	-	-	-	-	-
	linear	-	-	-	-	-
	quadratic	-	-	-	-	-

Die Ergebnisse für die multi-semantic Fusion basierend auf dem Biometric Hash-Algorithmus sind in Tabelle 7.23 angegeben. Mit Ausnahme von fünf Fusionen kommt es dabei zur Verbesserung, verglichen mit der Performanz der jeweils besseren Semantik. Die kleinste Equal Error Rate wird für die Kombination von *Pseudonym* und *Woher* unter Nutzung der Wichtung *equal* ermittelt. Dabei wird eine EER von 0,04612 berechnet, was einer Verbesserung von circa 46% verglichen mit der geringeren EER der beteiligten Semantiken *Pseudonym* (0,08943) und *Woher* (0,08559) entspricht.

Basierend auf dem Secure Sketch-Algorithmus kann im besten Fall eine Verbesserung um circa 56% erreicht werden (siehe Tabelle 7.25). Bei der Fusion der Semantiken *Symbol* (EER=0,07366) und *Woher* (EER=0,06817) wird eine Equal Error Rate von 0,03002 erreicht. Grundlage dafür bildet die quadratische Wichtungsstrategie.

Tabelle 7.24: Secure Sketch: Gewichte für multi-instance Fusion basierend auf den Einzelergebnissen beider Semantiken unter Verwendung der Testdaten aus Session 2

Semantik	Fusion	Geg. PIN		Geh. PIN		Pseu-donym		Sym-bol		Woher	
	single	0,13531		0,07722		0,06568		0,04961		0,05104	
gegebene PIN	equal	-	-	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000
	linear	-	-	0,63668	0,36332	0,67322	0,32678	0,73173	0,26827	0,72611	0,27389
	quadr.	-	-	0,75435	0,24565	0,80932	0,19068	0,88152	0,11848	0,87544	0,12456
geheime PIN	equal	-	-	-	-	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000
	linear	-	-	-	-	0,54037	0,45963	0,60885	0,39115	0,60204	0,39796
	quadr.	-	-	-	-	0,58022	0,41978	0,70785	0,29215	0,69592	0,30408
Pseudonym	equal	-	-	-	-	-	-	0,50000	0,50000	0,50000	0,50000
	linear	-	-	-	-	-	-	0,56970	0,43030	0,56270	0,43730
	quadr.	-	-	-	-	-	-	0,63675	0,36325	0,62347	0,37653
Symbol	equal	-	-	-	-	-	-	-	-	0,50000	0,50000
	linear	-	-	-	-	-	-	-	-	0,49288	0,50712
	quadr.	-	-	-	-	-	-	-	-	0,48576	0,51424

Tabelle 7.25: Secure Sketch: Ergebnisse der multi-instance Fusion im Vergleich zu den Einzelergebnissen (jeweils angegeben als EER)

Semantik	Fusion	Gegebene PIN	Geheime PIN	Pseudonym	Symbol	Woher
	single	0,16211	0,10927	0,08757	0,07259	0,06817
gegebene PIN	equal	-	0,12290	0,07997	0,08919	0,05411
	linear	-	0,11899	0,07015	0,05929	0,05141
	quadratic	-	0,10686	0,08133	0,06162	0,06439
geheime PIN	equal	-	-	0,05480	0,05216	0,04091
	linear	-	-	0,05489	0,04347	0,04411
	quadratic	-	-	0,05646	0,04604	0,04547
Pseudonym	equal	-	-	-	0,05417	0,04028
	linear	-	-	-	0,04307	0,03962
	quadratic	-	-	-	0,03708	0,04467
Symbol	equal	-	-	-	-	0,03178
	linear	-	-	-	-	0,03176
	quadratic	-	-	-	-	**0,03002**
Woher	equal	-	-	-	-	-
	linear	-	-	-	-	-
	quadratic	-	-	-	-	-

7.4.4 Fazit - Evaluierung: A.4 Biometrische Fusion

Zusammenfassend kann für die hier untersuchten Methoden und Strategien gesagt werden, dass sowohl für beide Algorithmen als auch für die Fusion zwischen ihnen häufig bessere Ergebnisse erzielt werden können als ohne Kombination. Betrachten wir die Ergebnisse der Fusionen auf Matching Score Level, zeigt sich ein größerer Erfolg vom Blickpunkt der Optimierung. In den 35 untersuchten Fällen wurde 32 Mal das beste Ergebnis durch eine der drei Fusionsstrategien (equal, linear, quadratic) ermittelt. Die beste Equal Error Rate mit 0,03002 wird durch die multi-instance Fusion der Semantiken *Symbol* (EER=0,07259) und *Woher* (EER=0,06817) unter Verwendung des Secure Sketch-Algorithmus bestimmt. Basierend auf diesen vielversprechenden Ergebnissen sollte in zukünftigen Arbeiten die Kombination unterschiedlicher Methoden der single-sematic Fusion untersucht werden. Basierend auf den guten Ergebnissen und den Daten unserer Methodologie

bietet sich dafür beispielsweise die multi-algorithmic Fusion unter Verwendung unterschiedlicher Semantiken für beide Algorithmen an.

7.5 Evaluierung: B Alterung biometrischer Daten

In diesem Abschnitt beschreiben wir die Ergebnisse zur experimentellen Untersuchung möglicher Effekte des zeitlichen Abstandes zwischen Referenz- und Verifikationsdaten. Dabei wird keine spezielle Parametrisierung der beiden Algorithmen genutzt, da die Ermittlung der Werte durch die in Abschnitt 6.2.5 *Setup zu B Alterung biometrischer Daten* beschriebene Konstellation des Setup zur Evaluierung wenig sinnvoll erscheint. Diese basiert auf drei Sessions mit jeweils zehn Samples pro Person. Wir teilen die zehn Schriftproben jeder Session jeweils auf, um Referenz- (erste fünf) und Testdaten (letzten fünf) zu erhalten. Damit verbleiben keine Daten, um Parameter zu bestimmen beziehungsweise eine Merkmalsanalyse durchzuführen. Eine Verwendung der in den vorhergehenden Abschnitten 7.2 bis 7.4 ermittelten und verwendeten Parameter, Merkmal-Sets oder Fusionsansätze schließen wir aus, da diese mithilfe der Daten aus den Sessions 1 beziehungsweise 2 ermittelt wurden. Das würde zu einer Verfälschung der Ergebnisse führen, da wir diese Informationen für die Evaluierung nutzen. Folglich verwenden wir hier 0 als Wert sowohl für den Tolerance Factor TF als auch für den Expansion Factor EF. Es hat sich bei der Bestimmung der Performanzmaße für den Hash-Generierungsmodus gezeigt, dass in allen Fällen die Resultate für die Reproduction Rate und die Collision Rate 0 beträgt, die daraus ermittelte Collision Reproduction Rate ist dementsprechend 0,5. Aus diesem Grund werden diese Werte in der folgenden Diskussion nicht berücksichtigt.

7.5.1 Testergebnisse für den Biometric Hash-Algorithmus

Die Ergebnisse für die Betrachtung der Auswirkung eines zunehmenden zeitlichen Abstandes zwischen der Erzeugung der Referenz- und Verifikationsdaten unter Verwendung des Biometric Hash sind in den Tabellen 7.26 bis 7.28 und der Abbildung 7.21 dargestellt. In den Tabellen werden ab der zweite Spalte die jeweiligen Resultate für die Verwendung der Referenzdaten R_i und Verifikationsdaten V_j angeben. Die Indizes i und j können dabei jeweils die Werte 1, 2 und 3 annehmen und verweisen damit auf die entsprechende Session, in der sie erfasst wurden.

Tabelle 7.26 zeigt die Ergebnisse für die einzelnen Sessions wobei jeweils nur Referenz- und Testdaten für die Verifikation herangezogen werden, die in derselben Session aufgenommen wurden. Das insgesamt beste Ergebnis weist die Verifikation mittels Semantik *Woher* in der zweiten Session mit einer EER von 0,01509 auf. Für dieselbe Session und die Semantik *gegebene PIN* wird das schlechteste Resultat ermittelt, hier beträgt die EER 0,06460. Für keine der drei Sessions kann festgestellt werden, dass hier alle besten beziehungsweise schlechtesten Ergebnisse ermittelt wurden. Daher kann angenommen werden, dass die Daten keiner der drei Sessions wesentlich mehr oder weniger für die Verifikation geeignet ist.

Tabelle 7.26: Biometric Hash: Verifikationsperformanz basierend auf Referenz- und Verifikationsdaten aus jeweils derselben Session in Abhängigkeit von der Semantik

	EER		
Semantik	R_1 **vs.** V_1	R_2 **vs.** V_2	R_3 **vs.** V_3
gegebene PIN	0,05882	0,06460	0,04939
geheime PIN	0,06038	0,04576	0,03459
Pseudonym	0,03819	0,03309	**0,02026**
Symbol	**0,02073**	0,03230	0,02918
Woher	0,03905	**0,01509**	0,02027

Die Verwendung von Verifikationsdaten aus den einzelnen Sessions mit den Referenzen aus der ersten wird in Tabelle 7.27 gezeigt. Hier wird deutlich, dass die Authentifikationsperformanz des Systems signifikant abnimmt, wenn zwischen der Erfassung der Referenzen und Testdaten ein Monat (Spalte R_1 *vs.* V_2) beziehungsweise zwei Monate (Spalte R_1 *vs.* V_3) liegen. Verglichen mit dem besten Wert aus R_1 *vs.* V_1 von 0,02073 für das *Symbol* verschlechtert sich die Equal Error Rate für dieselbe Semantik nach einem Monat auf 0,11075. Das entspricht einer Erhöhung der EER auf das mehr als fünffache. Nach einem weiteren Monat ergibt sich für die EER der Semantik *Symbol* mit einem Wert von 0,14595 (siehe Spalte R_1 *vs.* V_3) eine Vergrößerung auf etwa das siebenfache.

Tabelle 7.27: Biometric Hash: Verifikationsperformanz basierend auf Referenzdaten der ersten und Verifikationsdaten aus der ersten und den folgenden Sessions in Abhängigkeit von der Semantik

	EER		
Semantik	R_1 **vs.** V_1	R_1 **vs.** V_2	R_1 **vs.** V_3
gegebene PIN	0,05882	0,19140	0,23467
geheime PIN	0,06038	0,13566	0,16221
Pseudonym	0,03819	0,13209	0,14007
Symbol	**0,02073**	0,11075	0,14595
Woher	0,03905	**0,09107**	**0,13321**

Dieser Trend bestätigt sich auch bei der Betrachtung der Verifiaktionsergebnisse basierend auf den Referenzdaten der zweiten Session und den Testdaten der dritten Session in Tabelle 7.28 in Spalte R_2 *vs.* V_3. Auch bei der Verifikation mittels der Referenzdaten der zweiten Session mit den Testdaten der dritten Session ist eine deutliche Vergrößerung der EERs festzustellen, verglichen mit den Ergebnissen in Spalte R_2 *vs.* V_2. Das in diesem Fall beste Ergebnis ist eine EER von 0,01509 bestimmt für die Semantik *Woher*. Diese verschlechtert sich mit einem Wert von 0,07253 auf das etwa fünffache.

Tabelle 7.28: Biometric Hash: Verifikationsperformanz basierend auf Referenzdaten der zweiten und Verifikationsdaten aus der letzten Session in Abhängigkeit von der Semantik

	EER	
Semantik	R_2 **vs.** V_2	R_2 **vs.** V_3
gegebene PIN	0,06460	0,14906
geheime PIN	0,04576	0,12406
Pseudonym	0,03309	0,08369
Symbol	0,03230	0,11056
Woher	**0,01509**	**0,07253**

Die Unterschiede der Ergebnisse der Testdaten der einzelnen Sessions und der Referenzdaten derselben beziehungsweise der älteren Sessions werden in Abbildung 7.21 grafisch dargestellt. Deutlich wird hier zum einen die ähnliche Verifikationsperformanz bei den einzelnen Sessions, wenn Referenz- und Testdaten zusammen aufgenommen wurden. Auf der anderen Seite ist die zum Teil erhebliche Verschlechterung der Performanz zu erkennen, verwendet man Testdaten zusammen mit den Referenzdaten älterer Sessions.

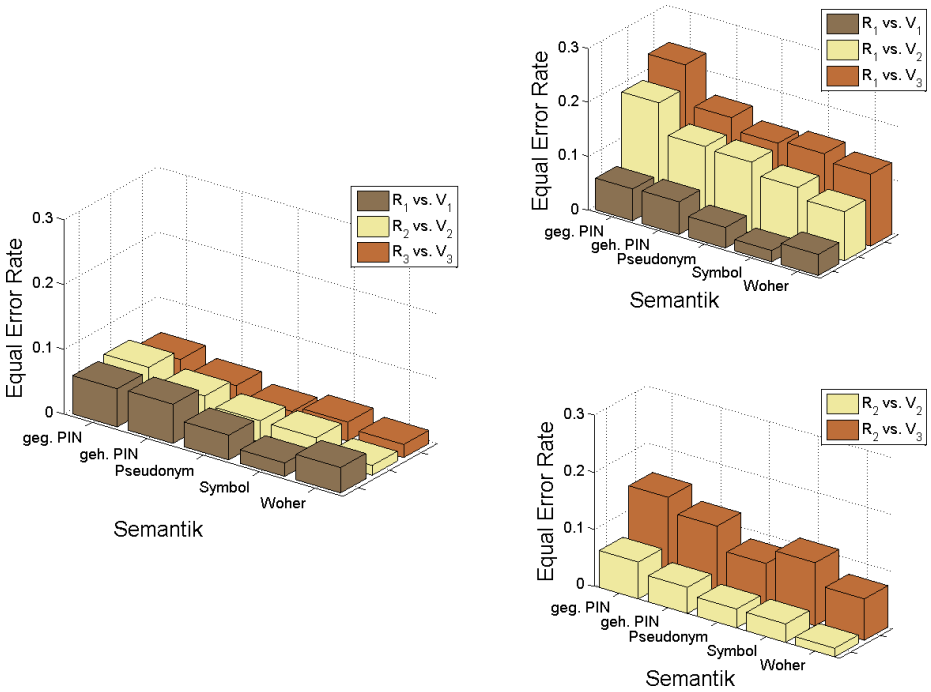


Abbildung 7.21: Biometric Hash: Grafische Darstellung der Verifikationsperformanz basierend auf: Referenz- und Verifikationsdaten derselben Session (*links*), Referenzdaten der ersten und Verifikationsdaten der drei Sessions (*rechts oben*) und Referenzdaten der zweiten und Verifikationsdaten der zweiten und dritten Session (*rechts unten*) in Abhängigkeit von der Semantik

7.5.2 Testergebnisse für den Secure Sketch-Algorithmus

Auch bei der Betrachtung der Ergebnisse für den Secure Sketch-Algorithmus ist die Tendenz deutlich zu erkennen, dass mit zunehmender zeitlicher Entfernung von Referenz- und Testdaten mit abnehmender Verifikationsperformanz zu rechnen ist. Die Resultate der entsprechenden Tests sind in den Tabellen 7.29 bis 7.31 dargestellt. Die Equal Error Rates, die jeweils für die einzelnen Semantiken mit Daten derselben Session bestimmt werden, liegen auch hier relativ dicht beieinander (siehe Tabelle 7.29). Die besten Ergebnisse werden hier für die erste und zweite Session mittels Semantik *Woher* (0,05170 beziehungsweise 0,03580) bestimmt, während in der dritten Session das *Pseudonym* mit 0,04261 am besten abschneidet.

In Tabelle 7.30 dienen jeweils R_1 als Referenzdaten für den Vergleich mit den Testdaten der einzelnen Sessions. Hier werden in allen drei Fällen die geringsten EERs für die Semantik *Woher*

bestimmt. Dabei ist die Equal Error Rate für die Testdaten der zweiten und dritten Session mit mehr als dem Doppelten deutlich höher als bei den in der ersten Session aufgenommenen Referenz- und Testdaten. Ähnlich sieht es bei den anderen vier Semantiken aus, wobei sich die EERs von *gegebene PIN* und *Pseudonym* von Session eins (0,08581 beziehungsweise 0,05646) zu Session 3 nahezu verdreifachen (0,24039 beziehungsweise 0,15726).

Tabelle 7.29: Secure Sketch: Verifikationsperformanz basierend auf Referenz- und Verifikationsdaten aus derselben Session in Abhängigkeit von der Semantik

Semantik	**EER** R_1 **vs.** V_1	R_2 **vs.** V_2	R_3 **vs.** V_3
gegebene PIN	0,08581	0,09639	0,08707
geheime PIN	0,09163	0,06277	0,06021
Pseudonym	0,05646	0,04592	**0,04261**
Symbol	0,06258	0,07558	0,06942
Woher	**0,05170**	**0,03580**	0,05250

Wie in Tabelle 7.31 dargestellt, zeigt sich der Einfluss der Alterung ebenfalls bei der Verwendung der Referenzdaten aus der zweiten Session (R_2) und der Testdaten aus Session 3 (V_3). Hier erhöht sich beispielsweise die Equal Error Rate basierend auf der Semantik *Woher* von 0,03580 auf mehr als das Doppelte (0,08172). Diese Tendenz ist auch bei den anderen Semantiken zu beobachten.

Tabelle 7.30: Secure Sketch: Verifikationsperformanz auf Referenzdaten der ersten und Verifikationsdaten aus der ersten und den folgenden Sessions in Abhängigkeit von der Semantik

Semantik	**EER** R_1 **vs.** V_1	R_1 **vs.** V_2	R_1 **vs.** V_3
gegebene PIN	0,08581	0,21189	0,24039
geheime PIN	0,09163	0,15158	0,16331
Pseudonym	0,05646	0,13962	0,15726
Symbol	0,06258	0,13285	0,15204
Woher	**0,05170**	**0,11708**	**0,11548**

Tabelle 7.31: Secure Sketch: Verifikationsperformanz basierend auf Referenzdaten der zweiten und Verifikationsdaten aus der letzten Session in Abhängigkeit von der Semantik

Semantik	**EER** R_2 **vs.** V_2	R_2 **vs.** V_3
gegebene PIN	0,09639	0,16153
geheime PIN	0,06277	0,12710
Pseudonym	0,04592	0,09669
Symbol	0,07558	0,11924
Woher	**0,03580**	**0,08172**

Die entsprechenden Ergebnisse für die Equal Error Rate basierend auf den drei Sessions und fünf Semantiken unter Verwendung des Secure Sketch-Algorithmus werden in Abbildung 7.22 grafisch aufbereitet dargestellt. Gut sind dabei die ähnlichen Ergebnisse der einzelnen Sessions für sich genommen zu erkennen (siehe Diagramm links). Die beiden Diagramme auf der rechten Seite zeigen deutlich den Einfluss älterer Referenzdaten auf das Verifikationsergebnis.

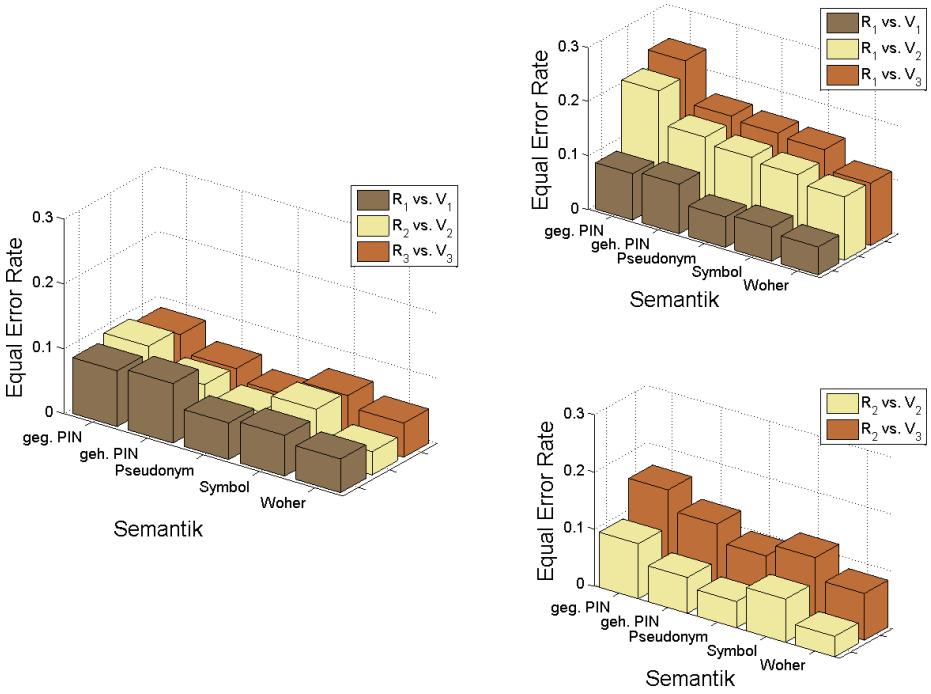


Abbildung 7.22: Secure Sketch: Grafische Darstellung der Verifikationsperformanz basierend auf: Referenz- und Verifikationsdaten derselben Session (*links*), Referenzdaten der ersten und Verifikationsdaten aller drei Sessions (*rechts oben*) und Referenzdaten der zweiten und Verifikationsdaten der zweiten und dritten Session (*rechts unten*) in Abhängigkeit von der Semantik

7.5.3 Fazit - Evaluierung: B Alterung biometrischer Daten

Die Evaluierung der Auswirkung der kurzfristigen Alterung der Referenzdaten hat basierend auf unserem Testset und der angewendeten Methodologie gezeigt, dass Unterschiede bei der Verifikationsperformanz auftreten. Dabei wird sowohl für den Biometric Hash- als auch für den Secure Sketch-Algorithmus in jedem untersuchten Fall eine deutliche Erhöhung der Equal Error Rate ermittelt, wenn Testdaten mit Referenzdaten älterer Sessions verglichen werden.

In unseren Tests wurde die größte Veränderung für die Semantik *Symbol* und den Biometric Hash-Algorithmus gemessen. Hier erhöht sich die Equal Error Rate von 0,02073 auf das Siebenfache (0,14595), wenn zwischen der Aufnahme der Referenz- und Testdaten zwei Monate liegen. Vergleicht man die Ergebnisse beider Algorithmen miteinander, fällt auf, dass der Einfluss der Alterung bei Verwendung des Biometric Hash-Algorithmus größer ist. So liegt der durchschnittliche Faktor für die fünf Semantiken, um den sich die Equal Error Rate verschlechtert, bei 3,33 (R_1 *vs.* V_2), 4,16 (R_1 *vs.* V_3) beziehungsweise 3,16 (R_2 *vs.* V_3). Für den Secure Sketch-Algorithmus fallen diese Werte deutlich geringer aus: 2,20 (R_1 *vs.* V_2), 2,41 (R_1 *vs.* V_3) und 1,93 (R_2 *vs.* V_3). Aus diesem Blickwinkel wird die Verwendung des Secure Sketch-Algorithmus interessant, obwohl für diesen bei den Tests in diesem Abschnitt die schlechteren EERs im Verifikationsmodus erzielt wurden. So nähren sich mit zunehmendem Alter der Referenzdaten die EERs beider Verfahren für die jeweiligen Semantiken immer mehr an. Dies spricht für eine höhere Toleranz des Secure

Sketch gegenüber der Alterung im Kontext unserer Daten und Methodologie.

In zukünftigen Arbeiten sollte überprüft werden, ob sich dieser Trend bestätigt beziehungsweise fortsetzt, wenn der Zeitraum zwischen Erfassung der Referenz- und Testdaten vergrößert wird. Einen weiteren Ansatzpunkt für ergänzende Untersuchungen bildet die Betrachtung der einzelnen Merkmale in Bezug auf ihre Empfindlichkeit gegenüber dem Faktor Alterung der Referenzdaten. Werden dabei Regelmäßigkeiten für bestimmte Merkmale festgestellt, kann basierend auf dieser Betrachtung eine weiter Methode zur Selektion von Merkmalen abgeleitet werden. Ebenso ist auch der Einsatz biometrischer Fusion zur Minderung der durch Alterung der Referenzdaten hervorgerufenen Effekte denkbar.

8 Fazit und Ausblick

In diesem Kapitel geben wir eine Zusammenfassung der im Rahmen dieser Arbeit durchgeführten Forschung und der wichtigsten daraus resultierenden Ergebnisse beziehungsweise Erkenntnisse. Daran anschließend folgt ein Überblick über durch die Vorgehensweise und Ergebnisse dieser Arbeit beeinflussten Forschungs- und Anwendungsbereiche sowie ein Ausblick mit Empfehlungen für zukünftige Arbeiten, die durch die hier vorgestellte Forschung motiviert beziehungsweise ermöglicht werden.

8.1 Fazit

In dieser Arbeit untersuchen wir verschiedene Ansätze zur Optimierung zweier Algorithmen, die auf der biometrischen Modalität der Handschrift beruhen. Zusätzlich wird für die Vielzahl sehr unterschiedlicher Ansätze ein einheitliches Evaluierungsszenario entwickelt und angewandt. Beide Algorithmen sind in der Lage, entweder in einem Verifikations- oder Hash-Generierungsmodus ausgeführt zu werden. Dabei ist die Verifikation ein typischer Anwendungsfall der biometrischen Benutzerauthentifikation, bei dem die vorgegebene Identität einer Person überprüft wird. Ein biometrischer Hash ist ein individueller stabiler Wert, der basierend auf variierenden biometrischen Eingangsdaten berechnet wird. Beide Algorithmen verwenden dasselbe Modul für die Merkmalsextraktion und werden mit identischen Daten beziehungsweise Methodologien evaluiert.

Die zur Optimierung untersuchten Ansätze lassen sich in die Kategorien merkmalbasiert, parameterbasiert und fusionbasiert einteilen. Dazu fügen wir dem vorhandenen statistischen Merkmal-Set (69 Elemente) im ersten Schritt 62 neue Merkmale hinzu. Für die Untersuchung der Auswirkung von an den Verwendungszweck angepassten Parametern wählen wir im zweiten Schritt für beide Algorithmen je einen Parameter zur Beeinflussung der Quantisierung. Dabei handelt es sich um den Tolerance Factor TF (Biometric Hash) und den Expansion Factor EF (Secure Sketch). Zur Untersuchung der Eignung der 131 Merkmale für die Anwendungsbereiche Verifikation und Hash-Generierung verwenden wir eine Auswahl von drei Wrapper- und fünf Filter-Verfahren. Aufbauend auf diese Analyse selektieren wir verschiedene Merkmal-Sets in Abhängigkeit von Anwendungsszenarien. Als weiteres Optimierungsverfahren wählen wir unterschiedliche Ansätze zur Kombination biometrischer Komponenten. Dabei beschränken wir uns auf single-modal Fusion-Verfahren, die ausschließlich die biometrische Handschrift nutzen. Wir kombinieren jeweils auf Matching Score Level beide Algorithmen miteinander sowie je zwei Instanzen der selben Semantik beziehungsweise je zwei Semantiken. Im folgenden gehen wir auf wesentliche Ergebnisse in Bezug auf die Aufgabenstellungen *A.1* - *A.4* und *B* ein.

8.1.1 Fazit: A.1 Hinzufügen zusätzlicher Merkmale

Für das Hinzufügen neuer Merkmale zu den existierenden 69 als Grundlage beider Algorithmen konnten wir zeigen, dass sich dadurch die Performanz der Verifikation erhöhen lässt. Die Hash-Generierung konnte hingegen nicht messbar verbessert werden, hier bleibt für beide Algorithmen die Collision Reproduction Rate auf 0,5 basierend auf je Null für Reproduction Rate und Collision Rate. Im Verifikationsmodus des Biometric Hash-Algorithmus wird die kleinste Equal Error Rate in drei von fünf Fällen (*gegebene PIN*, *Symbol*, *Woher*) für die Konkatenation der vier Sets mit insgesamt 131 Merkmalen bestimmt. Für die Semantiken *geheime PIN* und *Pseudonym* reichen die ersten drei Sets mit 103 Merkmal aus. Etwas anders stellt sich die Situation beim Secure Sketch-Algorithmus dar, bei dem lediglich für die *gegebene PIN* alle Merkmale für die geringste EER benötigt werden. Die Semantiken *geheime PIN* und *Pseudonym* benötigen drei, während *Symbol* und *Woher* nur die ersten beiden Merkmal-Sets zur Bestimmung der jeweils kleinsten EER verwenden. Betrachten wir die Werte der Equal Error Rate für die einzelnen Merkmal-Sets und Semantiken, stellen wir für den Biometric Hash-Algorithmus fest, dass die höchste Verbesserung für das *Symbol* erreicht wird. Hier wird der Ausgangswert von 0,16311 unter Verwendung aller 131 Merkmale um circa 23,3% auf 0,12512 gesenkt. Ebenfalls das *Symbol* liefert für das Secure Sketch-Verfahren die größte Verringerung der Equal Error Rate, in diesem Fall allerdings basierend auf lediglich den ersten beiden Merkmal-Sets (insgesamt 89 Merkmale). Hier wird eine Senkung der EER um 38,3% von 0,19680 auf 0,12141 erreicht. Damit ist das Erweitern des Merkmal-Sets zumindest im Kontext der Verifikation als erfolgreich anzusehen.

8.1.2 Fazit: A.2 Optimierung von Parametern

Vor der Analyse der Merkmale haben wir basierend auf dem kompletten Merkmal-Set mit 131 Elementen eine Optimierung der algorithmenspezifischen Parameter durchgeführt. Dazu wurden die Parameter Tolerance Factor TF (Biometric Hash) und Expansion Factor EF (Secure Sketch) schrittweise um jeweils 0,25 erhöht und die Verifikations- beziehungsweise Hash-Generierungsperformanz ermittelt. Darauf aufbauend haben wir je Algorithmus und Semantik fünf Parameter-Sets ermittelt: *Null*, *minEER*, *minCRR*, $RR \geq 80\%$ (Biometric Hash) beziehungsweise $RR \geq 70\%$ (Secure Sketch) und $CR \leq 5\%$. Für den Verifikationsmodus des Biometric Hash-Algorithmus wird die kleinste EER (0,09048) für die Semantik *Symbol* und *minEER* ermittelt. Das entspricht einer Verringerung um circa 41% im Vergleich zum nicht parametrisierten Durchlauf *Null* mit einer Equal Error Rate von 0,12512. Ebenfalls für das *Symbol* wird mit 0,09048 die beste Collision Reproduction Rate unter Verwendung des Parameter-Set *minCRR* berechnet, was im Vergleich mit 0,5 ohne optimierte Parametrisierung einer Verkleinerung um circa 82% bedeutet.

Für den Secure Sketch-Algorithmus wird die geringste Equal Error Rate mittels Parameter-Set *minEER* und der Semantik *Woher* bestimmt. Hier ergibt sich eine Verbesserung von circa 44% mit einer EER von 0,06817 gegenüber 0,12147 ohne Parameteroptimierung. Auch die Performanz der Hash-Generierung basierend auf dem Secure Sketch kann deutlich verbessert werden. Die

kleinste Collision Reproduction Rate liefert das *Symbol* mit 0,17311 unter Verwendung des Parameter-Set $CR \leq 5\%$. Verglichen mit dem Ausgangswert von 0,5 basierend auf *Null* entspricht dies einer Verringerung der CRR um circa 65%. Davon ausgehend, dass bei der Verwendung von 131 Merkmalen in neun (Verifikation) beziehungsweise in zehn von zehn (Hash-Generierung) Fällen das beste Ergebnis mittels eines optimierten Merkmal-Set ermittelt wurde, kann der hier durchgeführte Ansatz zur Parameteroptimierung unter Verwendung unserer Testdaten und unserer Methodologie als erfolgreich betrachtet werden.

8.1.3 Fazit: A.3 Analyse und Selektion von Merkmalen

Vor dem Hinzufügen wurden weder die vorhandenen noch die neuen Merkmale auf ihre Eignung bezüglich der Anwendungsszenarien Verifikation und Hash-Generierung hin untersucht. Aus diesem Grund haben wir in dieser Arbeit die Merkmale in Abhängigkeit vor den jeweiligen Applikationen analysiert, dies resultiert in einer Selektion derer, die eine Erhöhung der jeweiligen Performanz ermöglichen. Zum Zweck der Analyse der Merkmale setzen wir Wrapper und Filter ein, um sowohl für beide Algorithmen und die zwei Anwendungsszenarien, als auch für die je fünf Semantiken und die je fünf Parameter-Sets optimierte Merkmal-Sets zu ermitteln. Dabei kann nicht davon ausgegangen werden, die jeweils optimale Zusammenstellung von Merkmalen zu finden, da auf der einen Seite die Wrapper aufgrund der hohen Rechenaufwandes nicht jede mögliche Kombination berücksichtigen. Demgegenüber führen Filter, anders als Wrapper, die Analyse der Merkmale unabhängig vom jeweiligen Algorithmus durch. Dies hat den Vorteil, dass die Analyse schneller verläuft, allerdings gegebenenfalls Besonderheiten des Algorithmus nicht einbezogen werden.

Eine wichtige Beobachtung basierend auf den Ergebnissen der experimentellen Evaluierung in diesem Kontext ist, dass diejenigen Merkmal-Sets, die mittels Wrapper erstellt wurden, mit einer höheren Wahrscheinlichkeit zum besten Ergebnis führen. Für das Szenario *bestCRR* für den Biometric Hash-Algorithmus erzielen beispielsweise die Merkmal-Sets der drei Wrapper zusammen 49 der besten Equal Error Rate beziehungsweise Collision Reproduction Rate, während die fünf Filter gemeinsam nur auf drei kommen. Ähnlich sieht es für dieselbe Konstellation beim Secure Sketch-Algorithmus aus. Hier werden 36 der besten Resultate für Equal Error Rate und Collision Reproduction Rate durch die Merkmal-Sets der drei Wrapper und acht durch die der fünf Filter ermittelt.

Bei der Betrachtung der einzelnen Ergebnisse fällt auf, dass die kleinste Equal Error Rate von 0,03268 basierend auf dem Secure Sketch-Algorithmus berechnet wird. Die Grundlage bildet die folgende Konstellation: Semantik *Symbol*, Szenario *bestEER*, Parameter *minEER* und 26 Merkmale bestimmt von Wrapper *sbs*. Im Vergleich zum nicht parametrisierten und basierend auf allen 131 Merkmalen bestimmten Ausgangswert von 0,16158 entspricht dies einer Verbesserung um etwa 80%. Die höchte Verbesserung der Equal Error Rate auf Basis des Biometric Hash-Algorithmus fällt mit circa 65,32% geringer aus. Hier wird der Ausgangswert von 0,12512 auf 0,04339 gesenkt (Semantik *Symbol*, Szenario *bestEER*, Parameter *minEER*, 79 Merkmale,

Wrapper *sfs*).

Die kleinste Collision Reproduction Rate basiert ebenfalls auf dem Secure Sketch-Algorithmus und dem *Symbol*, weiterhin auf Szenario *bestCRR*, Parameter *minEER* und 22 Merkmalen ermittelt von Wrapper *sbs*. Ausgehend vom unparametrisierten Wert am Anfang der Untersuchungen von 0,5 kann die CRR auf 0,02848 gesenkt werden. Dies entspricht einer Verbesserung um circa 94%. Dem Wert liegen eine Reproduction Rate von 0,98491 und eine Collision Rate von 0,04187 zugrunde. Mittels Biometric Hash-Algorithmus wird die kleinste CRR von 0,03672 für das *Symbol*, Szenario *bestCRR*, Parameter *minEER*, Wrapper *sbs*, 46 Merkmale gebildet. Hier betragen Reproduction Rate und Collision Rate 0,99434 beziehungsweise 0,06778.

Entsprechend der genannten Ergebnisse betrachten wir die von uns untersuchten und evaluierten Ansätze zur Merkmalsanalyse und -selektion basierend auf den genutzten Testdaten und der Methodologie als erfolgreich. Dabei schneiden die eingesetzten Wrapper deutlich besser ab als die Filter, letztere weisen jedoch eine signifikant kürzere Laufzeit für die Analyse auf.

8.1.4 Fazit: A.4 Biometrische Fusion

Für die Ergebnisse der Fusionen können wir feststellen, dass in vielen der untersuchten Fällen ein besseres Resultat im Vergleich zu den Einzelergebnissen erzielt werden konnte. Das beste Verifikationsergebnis für die von uns untersuchten Fusionsmöglichkeiten haben wir für die multi-instance Fusion der Semantiken *Symbol* und *Woher* für den Secure Sketch-Algorithmus bestimmt. Der Wert für die Equal Error Rate sinkt durch die Fusion auf 0,03002, ausgehend von 0,07259 für *Symbol* beziehungsweise 0,06817 für *Woher* als Einzelresultate.

8.1.5 Fazit: B Alterung biometrischer Daten

Die experimentelle Evaluierung zur Untersuchung der Auswirkungen der kurzfristigen Alterung der Referenzdaten zeigt, dass es basierend auf den verwendeten Testdaten und der Methodologie erkennbare Effekte gibt. Wir können zeigen, dass sich die Equal Error Rate schon bei einem Abstand von einem Monat zwischen Aufnahme der Referenzen und der Testdaten deutlich und für jeden untersuchten Fall verschlechtert. Diese Beobachtung bestätigt sich, wenn wir unsere Untersuchungen auf zwei Monate zeitlichen Abstand erweitern. Die Equal Error Rate für den Biometric Hash-Algorithmus verschlechtert sich zum Beispiel für das *Symbol* nach einem Monat um circa 434%, wobei der Ausgangswert 0,02073 und der Wert nach einem Monat 0,11075 betragen. Nach einem weiteren Monat erhöht sich die EER auf 0,14595, was einer Erhöhung um etwa 604% entspricht. Die Betrachtung des Secure Sketch-Algorithmus führt zu einem ähnlichen Resultat, hier ist das Ausmaß der Verschlechterungen allerdings geringer. So erhöht sich die Equal Error Rate für das *Pseudonym* um etwa 147% von 0,05646 auf 0,13962. Nach einem weiteren Monat wird eine EER von 0,15726 ermittelt, das ist mit einer Erhöhung um circa 178% gleichzusetzen.

8.2 Einfluss auf andere Forschungs- und Anwendungsbereiche

Viele der Ergebnisse beziehungsweise Erkenntnisse dieser Arbeit beschränken sich nicht auf die biometrische Modalität der dynamischen Handschrift. In diesem Abschnitt gehen wir darauf ein, inwiefern die Resultate Einfluss auf andere Bereiche nehmen können. Die hier gezeigten Ergebnisse und die daraus gezogen Schlussfolgerungen basieren zwar auf speziellen Daten und darauf aufbauenden eindeutig definierten Szenarien, können aber eingeschränkt auf andere Bereiche übertragen werden.

In Abschnitt 7.1 *Evaluierung: Ausgangssituation und A.1 Hinzufügen zusätzlicher Merkmale* konnte gezeigt werden, dass das Hinzufügen neuer statistischer Merkmale teilweise zu signifikanten Verbesserungen der Verifikationsperformanz führen kann. Dadurch wird auch für andere handschriftbasierte Algorithmen beziehungsweise biometrische Modalitäten die Erweiterung des vorhandenen Merkmal-Sets motiviert. Zusätzlich können einige der neuen Merkmale auch für andere Modalitäten eingesetzt werden, zum Teil ist hierzu eine entsprechende Anpassung an die zugrunde liegenden Daten notwendig.

Die Anpassung eines einzelnen Parameters pro Algorithmus hat gezeigt, dass dadurch eine deutliche Verbesserung erreicht werden kann (vergleiche Abschnitt 7.2 *Evaluierung: A.2 Optimierung von Parametern*). Dies verdeutlicht, dass eine vom Verwendungszweck abhängige Optimierung der Parameter zu einer besseren Performanz der Verifikation und Hash-Generierung führen kann. Die für unsere experimentelle Evaluierung zugrunde gelegte Herangehensweise zur Anpassung der Parameter des Quantisierungsintervalls lässt sich auf ähnliche Verfahren übertragen. Voraussetzung ist die Möglichkeit, die Breite des Intervalls parametrisiert schrittweise zu verändern und die zu optimierenden Werte in Abhängigkeit dazu auszuwerten. Wird einer der hier untersuchten Algorithmen in zukünftigen Arbeiten mit alternativen Daten verwendet, sollten die ermittelten Parameter daraufhin untersucht werden, ob sie in Abhängigkeit vom Verwendungszweck und der verwendeten Semantik unverändert einsetzbar sind beziehungsweise gegebenenfalls neu bestimmt werden müssen.

Die in dieser Arbeit durchgeführte Analyse und die daraus resultierende Selektion von Merkmalen (siehe Abschnitt 7.3 *Evaluierung: A.3 Analyse und Selektion von Merkmalen*) weist im Kontext der Verbesserung von Verifikations- und Hash-Generierungsperformanz ein hohes Potential auf und motiviert dadurch auch für anderen Bereiche der Biometrie die Prüfung vorhandener Merkmale auf Eignung für das jeweilige Anwendungsszenario. Basierend auf den Ergebnissen weisen vor allem die basierend auf den gewählten Wrappern bestimmten Merkmal-Sets häufig die im Vergleich zu den Filtern besten Ergebnisse auf. Dieses Wissen erleichtert in zukünftigen Arbeiten eventuell die Wahl zwischen beiden Kategorien im Allgemeinen und den einzeln Verfahren im speziellen. Die die im Vergleich zu den Filtern besseren Resultate der Wrapper werden in unseren Tests erkauft durch einen wesentlich höheren zeitlichen Aufwand bei der Bestimmung des für den einzelnen Anwendungsfall optimalen Merkmal-Sets.

Die Herangehensweise bei der in 7.4 *Evaluierung: A.4 Biometrische Fusion* durchgeführten single-

modalen Fusion ist auch auf andere Modalitäten beziehungsweise multi-biometrische Systeme möglich. Die vorgestellten Wichtungsstrategien sind abhängig von der Equal Error Rate der beteiligten biometrischen Komponenten und lassen sich auch entsprechend auf mehr als zwei anwenden. Die gezeigten Ansätze lassen sich auch in anderen Bereichen anwenden, indem nach der Bestimmung geeigneter Fusionsparameter beispielsweise Klassifikatoren miteinander kombiniert werden.

Die anhand unseres Testszenarios zur Untersuchung der Kurzzeitalterung basierend auf den im Rahmen dieser Arbeit erhobenen Daten nachgewiesene Abhängigkeit motiviert weitere Forschung in diesem Bereich. Dazu gehört neben der Betrachtung weiterer biometrischer Modalitäten bezüglich der Alterung sowohl die Ausdehnung der zeitlichen Abstände als auch die Erforschung geeigneter Verfahren, um der Alterung entgegenzuwirken wie beispielsweise durch Update-Strategien.

Ein wichtiger Beitrag für zukünftige Arbeiten stellt der Aufbau und die Verwendung der Testdatenbank dar (vergleiche Kapitel 6 *Aufbau und Methodologie der experimentellen Evaluierung*). Die zeitversetzte Erfassung der Daten in drei Sessions mit dem Abstand von je einem Monat stellt zum einen ausreichend Daten für die Bestimmung von Referenzdaten zur Verfügung (Daten aus Session 1). Auf der anderen Seite können die Daten der Session 2 genutzt werden,um Parameter zu bestimmen, die Optimierung durchzuführen und für die Fusion die Wichtungsparameter zu ermitteln. Ergänzend stehen mit den Daten der dritten Session pro Person mehrere Testdaten zur Verfügung mit denen die experimentelle Überprüfung der Auswirkungen der genannten Optimierungsansätze durchgeführt werden kann. Zusätzlich bietet die Art der Datenerfassung auch die Möglichkeit zur zeitabhängigen Untersuchung der Algorithmen.

In Abschnitt 2.1.7 *Reproduktion und Kollision im Hash-Generierungsmodus* haben wir Maße für biometrische Hashes mit dem Ziel der Feststellung der Reproduzierbarkeit (Reproduction Rate) und Kollisionswahrscheinlichkeit (Collision Rate) sowie der Collision Reproduction Rate als deren Trade Off vorgestellt. Diese Maße sind ohne zusätzliche Anpassungen für Hash-Verfahren sowohl basierend auf anderen biometrischen Modalitäten als auch in weiteren Bereichen mit unscharfen Eingabedaten einsetzbar.

Zusätzlich zu den hier genannten Beiträgen ergeben sich durch diese Arbeit und deren Ergebnisse weitere Forschungen beziehungsweise werden diese dadurch ermöglicht. Auf diesen Forschungsbedarf gehen wir im nächsten Abschnitt ein.

8.3 Weiterer Forschungsbedarf

Im Rahmen dieser Arbeit haben wir ausschließlich Daten verwendet, die mit einem einzelnen Sensor (Toshiba Pórtegé M200) erfasst wurden. Ein weiterer Ansatzpunkt für künftige Forschung bietet die Untersuchung der Hardware-Interoperabilität im Zusammenhang mit den hier herangezogenen Verfahren zur Optimierung. Hardware-Interoperabilität meint in diesem Kontext die Verwendung unterschiedlicher Sensoren für die Erstellung der Referenz- beziehungsweise

Hilfsdaten und der Testdaten. Dies kann vor allem dann ein Problem darstellen, wenn sich die physikalischen Eigenschaften (beispielsweise Auflösung von X, Y oder Druck) der jeweiligen Geräte zu stark unterscheiden. In diesem Zusammenhang sollte erforscht werden, ob eine Optimierung basierend auf der Anpassung von algorithmenspezifischen Parametern oder der Selektion von Merkmalen zu besseren Ergebnissen führen kann.

Diese Ansätze sollten ebenfalls bei der Kompensation der Auswirkungen der Alterung untersucht werden. So bewirken Tolerance Factor und Tolerance Vector des Biometric Hash-Algorithmus die Verkleinerung beziehungsweise Vergrößerung des Mapping-Intervalls bei der Hash-Generierung, bei entsprechender Bestimmung auch individuell für einzelne Nutzer oder Nutzergruppen. Darauf aufbauend ist folglich auch eine Anpassung der Parameter denkbar, die den hier identifizierten Effekten der Alterung entgegen wirkt. Unter diesem Aspekt erscheint ebenfalls die Nutzung von Verfahren zur Merkmalsselektion sinnvoll. Durch diese können gegebenenfalls Merkmale, die besonders stark durch die zunehmende Zeitspanne zwischen der Erfassung der Referenzen und der Testdaten beeinflusst werden, erkannt und aussortiert werden. An dieser Stelle bieten sich auch Tests an, die die Merkmale der Daten vergleichen, die in den unterschiedlichen Sessions akquiriert wurden. Auf diese Weise können eventuell diejenigen Merkmale identifiziert werden, auf denen die negativen Effekte der Alterung der Referenzen beruhen.

Neben der Untersuchung der optimalen Parametrisierung beziehungsweise der Merkmalsselektion in Verbindung mit der Alterung stellt auch die Verwendung der biometrischen Fusion in diesem Kontext ein potentielles zukünftiges Forschungsfeld dar. Hier erscheint beispielsweise die Verwendung von multiplen Sensoren, Semantiken oder Algorithmen beziehungsweise deren Kombination vielversprechend.

Da wir anhand unserer Testdaten und der verwendeten Methodologie zeigen konnten, dass schon nach einer kurzen Zeit von einem beziehungsweise zwei Monaten eine signifikante Verschlechterung der Verifikationsperformanz eintritt, ist die weitere Untersuchung der Alterung biometrischer Referenzdaten empfehlenswert. So kann sowohl die Anzahl als auch die Zeitspanne zwischen den einzelnen Sessions beispielsweise erhöht werden, um festzustellen, ob sich die Verschlechterung der Performanz weiter fortsetzt und gegebenenfalls bis zu welchem Grad.

Literaturverzeichnis

[Alb94] Albert, B.: *Verfahren zur Erzeugung einer digitalen Signatur mit Hilfe eines biometrischen Merkmals.* 1994

[Ano11] Anoto Group AB: *Anoto - Start.* Version: 2011. `http://www.anoto.com/`, Abruf: 2014-10-06

[BAN+10] Busch, Christoph ; Abt, Sebastian ; Nickel, Claudia ; Korte, Ulrike ; Zhou, Xuebing: Biometrische Template-Protection-Verfahren und Interoperabilitätsstrategien. In: *Sicherheit 2010*, 2010, S. 1–12

[BB51] Birren, J.E. ; Botwinick, J.: The relation of writing speed to age and to the senile psychoses. In: *Journal of Consulting Psychology* 15 (1951), S. 243–249

[BCK09] Bringer, Julien ; Chabanne, Herv´e ; Kindarji, Bruno: Anonymous Identification with Cancelable Biometrics. In: *Image and Signal Processing and Analysis (ISPA 2009)*, 2009, S. 494 – 499

[[BI08] *BIODEVII (Biometric Data Experimented in Visas), Children fingerprinting.* Version: 2008. `www.statewatch.org/news/2008/oct/eu-com-fp-children-rep.pdf`, Abruf: 2012-05-12

[BSI05] BSI: *BioP II - Untersuchung der Leistungsfähigkeit von biometrischen Verifikationssystemen, Version 2.0.* Version: 2005. `https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/BioP/biopabschluss2_pdf.pdf?__blob=publicationFile`, Abruf: 2014-10-06

[CBF03] Chang, Kyong I. ; Bowyer, Kevin W. ; Flynn, Patrick J.: Face recognition using 2D and 3D facial data. In: *ACM Workshop on Multimodal User Authentication*, 2003, S. 25–32

[CLRS09] Cormen, Thomas H. ; Leiserson, Charles E. ; Rivest, Ronald L. ; Stein, Clifford: *Introduction to Algorithms, Third Edition.* 3rd. The MIT Press, 2009

[CS09] Cavoukian, Ann ; Stoianov, Alex: Biometric Encryption. In: Li, S. (Hrsg.) ; Jain, A. K. (Hrsg.): *Encyclopedia of Biometrics.* Springer US, 2009 (LNCS), Kapitel E, S. 260–269

[CSS+06] Corby, P.M. ; Schleyer, T. ; Spallek, H. ; Hart, T.C. ; Weyant, R.J. ; Corby, A.L. ; Bretz, W.A.: Using biometrics for participant identification in a research

study: a case report. In: *Journal of the American Medical Informatics Association* 12 (2006), Nr. 2, S. 233–235

[CTGN04] Connie, Tee ; Teoh, Andrew ; Goh, Michael ; Ngo, David: PalmHashing: A novel approach for dual-factor authentication. In: *Pattern Anal. Appl.* 7 (2004), Nr. 3, S. 255–268

[Dau08] Daugman, John: *Iris Recognition.* Version: 2008. `http://www.icdri.org/biometrics/iris_biometrics.htm`, Abruf: 2012-05-12

[Dit00] Dittmann, Jana: *Digitale Wasserzeichen - Grundlagen, Verfahren, Anwendungsgebiete.* Springer, 2000 (Xpert.press). – I–IX, 1–183 S.

[DK07] Delfs, Hans ; Knebl, Helmut: *Introduction to Cryptography: Principles and Applications.* Second, Extended Edition. Springer Berlin Heidelberg, 2007 (Information Security and Cryptography). – 367 S.

[DKDS11] Dwivedi, Akhilesh ; Kumar, Suresh ; Dwivedi, Abhishek ; Singh, Manjeet: Cancellable Biometrics for Security and Privacy Enforcement on Semantic Web. In: *International Journal of Computer Applications* 21 (2011), May, Nr. 8, S. 1–8

[DKMV07] Draper, Stark C. ; Khisti, Ashish ; Martinian, Emin ; Vetro, Anthony: Secure storage of fingerprint biometrics using Slepian-Wolf codes. In: *in Inform. Theory and Apps. Work., UCSD*, 2007

[DRS04] Dodis, Yevgeniy ; Reyzin, Leonid ; Smith, Adam: Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. In: *EUROCRYPT*, 2004, S. 523–540

[Eck05] Eckert, Claudia: *IT-Sicherheit: Konzept - Verfahren - Protokolle.* Studienausgabe. Oldenbourg Verlag München Wien, 2005. – 412 S.

[EF12] Erbilek, M. ; Fairhurst, Michael: Framework for managing ageing effects in signature biometrics. In: *Biometrics, IET* 1 (2012), Nr. 2, S. 136–147

[EFG+09] Erkin, Zekeriya ; Franz, Martin ; Guajardo, Jorge ; Katzenbeisser, Stefan ; Lagendijk, Inald ; Toft, Tomas: Privacy-Preserving Face Recognition. In: *Privacy Enhancing Technologies Symposium (PETS 2009)*, 2009

[EKS09] Elliott, S.J. ; Kukula, E.P. ; Sickler, N.C.: The Challenges of the Environment and the Human / Biometric Device Interaction on Biometric System Performance. In: *International Workshop on Biometric Technologies-Special forum on Modeling and Simulation in Biometric Technology, Calgary, Alberta, Canada*, 2009

[FEL14] Fairhurst, Michael ; Erbilek, Meryem ; Li, Cheng: Enhancing the forensic value of handwriting using emotion prediction. In: *Biometrics and Forensics (IWBF), 2014 International Workshop on*, 2014, S. 1–6

[FFGO07] Freire, Manuel R. ; Fiérrez-Aguilar, Julian ; Galbally, Javier ; Ortega-Garcia, Javier: Biometric Hashing Based on Genetic Selection and Its Application to On-Line Signatures. In: *ICB'07*, 2007, S. 1134–1143

[FHK+10] Färberböck, Peter ; Hämmerle-Uhl, Jutta ; Kaaser, Dominik ; Pschernig, Elias ; Uhl, Andreas: Transforming Rectangular and Polar Iris Images to Enable Cancelable Biometrics. In: Campilho, A. C. (Hrsg.) ; Kamel, M. S. (Hrsg.): *ICIAR (2)*, Springer, 2010 (LNCS 6112), S. 276–286

[FHP08] Franke, Dirk ; Herrmann, Enrico ; Paatz, Martin: *Abschlussbericht: Team BioHash: Sicherheit biometrischer Hashfunktionen am Beispiel von Handschrift.* Multi-modal Data Analysis Project: Biometrics, Otto-von-Guericke Universität Magdeburg, Januar 2008

[FLC10] Fang, Chengfang ; Li, Qiming ; Chang, Ee-Chien: Secure sketch for multiple secrets. In: *Proceedings of the 8th international conference on Applied cryptography and network security*, Springer Verlag Berlin Heidelberg, 2010, S. 367–383

[FOB13] Fenker, Samuel P. ; Ortiz, Estefan ; Bowyer, Kevin W.: Template Aging Phenomenon in Iris Recognition. In: *Access, IEEE* 1 (2013), S. 266–274

[FR11] Fratric, Ivan ; Ribaric, Slobodan: Local Binary LDA for Face Recognition. In: Vielhauer, C. (Hrsg.) ; Dittmann, J. (Hrsg.) ; Drygajlo, A. (Hrsg.) ; Juul, N. C. (Hrsg.) ; Fairhurst, M. C. (Hrsg.): *Biometrics and ID Management - COST 2101 European Workshop, BioID 2011*, Springer Berlin Heidelberg, 2011 (LNCS 6583), S. 144–155

[Fra03] Fraser, Fiona: Exploring the use of face recognition technology for border control applications - australia's experience. In: *Biometric Consortium Conference*, 2003

[GBDB02] Givens, Geof ; Beveridge, J R. ; Draper, Bruce A. ; Bolme, David: A statistical assessment of subject factors in the pca recognition of human faces. In: *Proceedings of CVPR Workshop: Statistical Analysis in Computer Vision*, 2002

[GE03] Guyon, Isabelle ; Elisseeff, André: An introduction to variable and feature selection. In: *Journal of Machine Learning Research* 3 (2003), S. 1157–1182

[GFA+07] Garcia-Salicetti, S. ; Fierrez-Aguilar, J. ; Alonso-Fernandez, F. ; Vielhauer, C. ; Guest, R. ; Allano, L. ; Doan Trung, T. ; Scheidat, T. ; Ly Van, B. ; Dittmann, J. ; Dorizzi, B. ; Ortega-Garcia, J. ; Gonzalez-Rodriguez, J. ; Bacile di Castiglione, M. ; Fairhurst, M.: Biosecure Reference Systems for On-Line Signature Verification: A Study of Complementarity. In: *Annals of Telecommunications, Special Issue on Multimodal Biometrics* 62 (2007), Nr. 1-2

[GFMP11] Galbally, J. ; Fierrez, J. ; Martinez-Diaz, M. ; Plamondon, R.: Quality Analysis of Dynamic Signature Based on the Sigma-Lognormal Model. In: *Document*

Analysis and Recognition (ICDAR), 2011 International Conference on, 2011, S. 633–637

[GKHK69] Gellert, W. (Hrsg.) ; Küstner, H. (Hrsg.) ; Hellwich, M. (Hrsg.) ; Kästner, H. (Hrsg.): *Kleine Enzyklopädie Mathematik*. 4. Auflage. VEB Bibligraphisches Institut Leipzig, 1969. – 837 S.

[Gue06] Guest, Richard: Age dependency in handwritten dynamic signature verification systems. In: *Pattern Recogn. Lett.* 27 (2006), July, S. 1098–1104

[Guz07] Guzmán, Heriberto Adolfo R.: *Discriminatory Power and Reproducibility of biometric Hashes - A statistical feature analysis*, Otto-von-Guericke Universität Magdeburg, Diplomarbeit, September 2007

[Haa07] Haarmann, Harald: *Geschichte der Schrift - Von den Hieroglyphen bis heute*. Verlag C.H. Beck, 2007 (besck'sche Reihe). – 128 S.

[HAD05] Hao, Feng ; Anderson, Ross ; Daugman, John: Combining Crypto with Biometrics Effectively. In: *IEEE Transactions on Computers* 55 (2005), Nr. 9, S. 1081–1088

[Ham50] Hamming, Richard W.: Error detecting and error correcting codes. In: *Bell System Technical Journal* 29 (1950), April, Nr. 2, S. 147–160

[HBF09] Hollingsworth, Karen P. ; Bowyer, Kevin W. ; Flynn, Patrick J.: The Best Bits in an Iris Code. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31 (2009), Nr. 6, S. 964–973

[HHH09] Hellmich, Martin ; Hintsch, Johannes ; Holthusen, Sönke: *Abschlussbericht: Projekt OptimizedBioHashing*. Projektvorlesung Secure Infrastructure Project, Otto-von-Guericke Universität Magdeburg, Juni 2009

[HLP12] Hsu, Chao-Yung ; Lu, Chun-Shien ; Pei, Soo-Chang: Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT. In: *Image Processing, IEEE Transactions on* 21 (2012), Nov, Nr. 11, S. 4593–4607

[ide05] identix: *identix FaceIt: G6 Frequently Asked Technical Questions*. Version: 2005. `http://epic.org/privacy/surveillance/spotlight/1105/facefaqs.pdf`, Abruf: 2014-10-06

[IDVR14] Izquierdo-Fuente, Alberto ; Del-Val-Puente, Lara ; Villacorta-Calvo, Juan J. ; Raboso-Mateos, Mariano: Optimization of a biometric system based on acoustic images. In: *ScientificWorldJournal* 2014 (2014)

[[IS10] Norm ISO/IEC CD 24745 2010. *Information technology - Security techniques - Biometric template protection*

[JAS09] Jassim, Sabah ; Al-assam, Hisham ; Sellahewa, Harin: Improving Performance and Security of Biometrics Using Efficient and Stable Random Projection Techniques*. In: *Image and Signal Processing and Analysis (ISPA 2009)*, 2009

[JBP99] Jain, Anil K. (Hrsg.) ; Bolle, Ruud (Hrsg.) ; Pankanti, Sharath (Hrsg.): *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Kluwer Academic Publishers Norwell, MA, USA, 1999

[JG11] Johnson, Emma P. S. ; Guest, Richard M.: The Use of Static Biometric Signature Data from Public Service Forms. In: Vielhauer, C. (Hrsg.) ; Dittmann, J. (Hrsg.) ; Drygajlo, A. (Hrsg.) ; Juul, N. C. (Hrsg.) ; Fairhurst, M. C. (Hrsg.): *Biometrics and ID Management - COST 2101 European Workshop, BioID 2011*, Springer Berlin Heidelberg, 2011 (LNCS 6583), S. 73–82

[JGC02] Jain, Anil K. ; Griess, Friederike D. ; Connell, Scott D.: On-line signature verification. In: *Pattern Recognition* 35 (2002), Nr. 12, S. 2963 – 2972. – Pattern Recognition in Information Systems

[JKP94] John, George H. ; Kohavi, Ron ; Pfleger, Karl: Irrelevant Features and the Subset Selection Problem. In: *Proceedings of the Eleventh International Conference on Machine Learning* Bd. 129, Morgan Kaufmann, 1994, S. 121–129

[JNN08] Jain, Anil K. ; Nandakumar, Karthik ; Nagar, Abhishek: Biometric template security. In: *EURASIP J. Adv. Signal Process* 2008 (2008), S. 1–17

[JW99] Juels, Ari ; Wattenberg, Martin: A Fuzzy Commitment Scheme. In: *Proceedings of the 6th ACM conference on Computer and communications security*, ACM Press, 1999, S. 28–36

[KEM11] Karabat, Cagatay ; Erdogan, Hakan ; Mihcak, Mehmet K.: Information Theoretic Capacity Analysis for Biometric Hashing Methods. In: *Proceedings of the 17th International Conference on Digital Signal Processing (DSP 2011)*, 2011

[KGS10] Kisku, Dakshina R. ; Gupta, Phalguni ; Sing, Jamuna K.: Feature Level Fusion of Face and Palmprint Biometrics by Isomorphic Graph-based Improved K-Medoids Partitioning. In: Kim, T.-h. (Hrsg.) ; Adeli, H. (Hrsg.): *Advances in Computer Science and Information Technology, Joint Proceedings AST/UCMA/ISA/ACN 2010 Conferences*. Springer Berlin Heidelberg, 2010 (LNCS 6059), S. 70–81

[KHH02] Kato, Yosuke ; Hamamoto, Takayuki ; Hangai, Seiichiro: A proposal of writer verification of hand written objects. In: *IEEE International Conference on Multimedia and Expo, ICME 2002* Bd. II IEEE, IEEE, August 2002, 585-588

[KKM+08] Korte, Ulrike ; Krawczak, Michael ; Merkle, Johannes ; Plaga, Rainer ; Niesing, Matthias ; Tiemann, Carsten ; Vinck, Han ; Martini, Ulrich: A Crypto-

graphic Biometric Authentication System based on Genetic Fingerprints. In: *Lecture Notes of Informatics*. Springer, 2008 (LNI P-128), S. 263–276

[KZ05] Kumar, Ajay ; Zhang, David: Biometric Recognition Using Feature Selection and Combination. In: Kanade, T. (Hrsg.) ; Jain, A. K. (Hrsg.) ; Ratha, N.K. (Hrsg.): *Proceedings of 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, Springer Berlin Heidelberg, 2005 (LNCS 3546), S. 813–822

[LGD04] Ly, Van-Bao ; Garcia-Salicetti, Sonia ; Dorizzi, Bernadette: Fusion of HMM's Likelihood and Viterbi Path for On-line Signature Verification. In: *Biometric Authentication, ECCV 2004 International Workshop, BioAW 2004, Prague, Czech Republic, May 15, 2004, Proceedings*, 2004, S. 318–331

[LJ09] Li, Stan Z. (Hrsg.) ; Jain, Anil K. (Hrsg.): *Encyclopedia of Biometrics*. Springer US, 2009

[LLCM07] Lee, Yongjin ; Lee, Yongki ; Chung, Yunsu ; Moon, Kiyoung: One-Time Templates for Face Authentication. In: *Proceedings of the 2007 International Conference on Convergence Information Technology*, IEEE Computer Society Washington, DC, USA, 2007, S. 1818–1823

[MBE04] Monga, Vishal ; Banerjee, Arindam ; Evans, Brian L.: Clustering Algorithms for Perceptual Image Hashing. In: *Proc. IEEE Digital Signal Processing Workshop*, 2004, S. 283–287

[MCN08] Maiorana, Emanuele ; Campisi, Patrizio ; Neri, Alessandro: User adaptive fuzzy commitment for signature template protection and renewability. In: *Journal of Electronic Imaging* (2008)

[ME04] Monga, Vishal ; Evans, Brian L.: Robust Perceptual Image Hashing Using Feature Points. In: *Proceedings of International Conference on Image Processing (ICIP)*, 2004, S. 677–680

[ME06] Modi, Shimon K. ; Elliott, Stephen J.: Impact of image quality on performance: Comparison of young and elderly fingerprints. In: *The 6th International Conference on Recent Advances in Soft Computing (RASC)*, 2006, S. 10–12

[MEKW07] Modi, S. ; Elliott, S.J. ; Kim, H. ; Whetstone, J.: Impact of Age Groups on Fingerprint Recognition Performance. In: *AutoID 2007, IEEE Workshop on Automatic Identification Advanced Technologies*, 2007

[Mic82] Michel, Lothar: *Gerichtliche Schriftvergleichung: Eine Einführung in Grundlagen, Methoden und Praxis*. de Gruyter, 1982

[Min05] Ministry of the Interior & Kingdom Relations - Netherlands: *Evaluation Report Biometrics Trial 2b or not 2b*. Version: 2005. `http://dematerialisedid.com/PDFs/88_630_file.pdf`, Abruf: 2014-10-06

[MKCK01] Mansfield, A.J. ; Kelly, G. ; Chandler, D. ; Kane, J.: *Biometric Product Testing*. Version: March 2001. `http://www.cesg.gov.uk/policy_technologies/biometrics/media/biometrictestreportpt1.pdf`, Abruf: 2012-05-12. (1)

[MRLW01] Monrose, Fabian ; Reiter, Michael K. ; Li, Qi ; Wetzel, Susanne: Using Voice to Generate Cryptographic Keys. In: *In Proc. of Odyssey 2001, The Speaker Verification Workshop*, 2001, S. 237–242

[MSV11a] Makrushin, Andrey ; Scheidat, Tobias ; Vielhauer, Claus: Handwriting Biometrics: Feature Selection Based Improvements in Authentication and Hash Generation Accuracy. In: Vielhauer, C. (Hrsg.) ; Dittmann, J. (Hrsg.) ; Drygajlo, A. (Hrsg.) ; Juul, N. C. (Hrsg.) ; Fairhurst, M. C. (Hrsg.): *Biometrics and ID Management - COST 2101 European Workshop, BioID 2011*, Springer Verlag Berlin Heidelberg, 2011 (LNCS 6583), S. 37–48

[MSV11b] Makrushin, Andrey ; Scheidat, Tobias ; Vielhauer, Claus: Towards Robust Bio-Hash Generation for Dynamic Handwriting Using Feature Selection. In: *Proceedings of 17th Conference on Digital Signal Processing*, 2011

[MSV12] Makrushin, Andrey ; Scheidat, Tobias ; Vielhauer, Claus: Improving Reliability of Biometric Hash Generation through the Selection of Dynamic Handwriting Features. In: *Transactions on Data Hiding and Multimedia Security VIII - Special Issue on Pattern Recognition for IT Security* 7228 (2012), S. 19–41

[MTS+99] Mergl, R. ; Tigges, P. ; Schröter, A. ; Möller, H.J. ; Hegerl, U.: Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. In: *Journal of neuroscience methods* 90 (1999), August, S. 157–169

[MTWZ04] Ma, Li ; Tan, Tieniu ; Wang, Yunhong ; Zhang, Dexin: Efficient Iris Recognition by Characterizing Key Local Variations. In: *IEEE Trans. on Image Processing* 13 (2004), S. 739–750

[MV01] Mihçak, Mehmet K. ; Venkatesan, Ramarathnam: A Perceptual Audio Hashing Algorithm: A Tool for Robust Audio Identification and Information Hiding. In: *Proceedings of the 4th International Workshop on Information Hiding*, Springer-Verlag London, 2001, S. 51–65

[MW02] Mansfield, A. J. ; Wayman, J. L.: Best Practices in Testing and Reporting Performance of Biometric Devices: Version 2.01 / National Physical Laboratory. Centre for Mathematics and Scientific Computing, National Physical Laboratory, 2002. (NPL report). – Forschungsbericht

[NNJ12] Nagar, Abhishek ; Nandakumar, Karthik ; Jain, Anil K.: Multibiometric Cryptosystems Based on Feature-Level Fusion. In: *IEEE Transactions on Information Forensics and Security* 7 (2012), Nr. 1, S. 255–268

[NTG04] Ngo, David Chek L. ; Teoh, Andrew Beng J. ; Goh, Alwyn: Eigenspace-Based Face Hashing. In: Zhang, D. (Hrsg.) ; Jain, A. K. (Hrsg.): *Biometric Authentication, Proceedings of First International Conference on Biometric Authentication (ICBA)*, Springer Berlin Heidelberg, 2004 (LNCS 3072), S. 195–199

[NTG06] Ngo, David Chek L. ; Teoh, Andrew Beng J. ; Goh, Alwyn: Biometric hash: high-confidence face recognition. In: *IEEE Transactions on Circuits and Systems for Video Technology* 16 (2006), Nr. 6, S. 771–775

[NTN02] Nanavati, S. ; Thieme, M. ; Nanavati, R.: *Biometrics: Identity Verification in a Networked World*. Wiley Computer Publishing, 2002

[OPV+99] O'Toole, A.J. ; Price, T. ; Vetter, T. ; Bartlett, J.C. ; Blanz, V.: 3D shape and 2D surface textures of human faces: the role of „ averages“ in attractiveness and age. In: *Image and Vision Computing* 18 (1999), S. 9–19

[PGM+03] Phillips, P. J. ; Grother, Patrick ; Micheals, Ross J. ; Blackburn, Duane M. ; Tabassi, Elham ; Bone, Mike ; Dr, North F. ; Kingdom, United ; Phillips, P. J. ; Grother, Patrick ; Micheals, Ross J. ; Blackburn, Duane M. ; Tabassi, Elham ; Bone, Mike: *Face Recognition Vendor Test 2002: Evaluation Report*. Version: 2003. `http://www.face-rec.org/vendors/FRVT_2002_Evaluation_Report.pdf`, Abruf: 2014-10-06

[PL89] Plamondon, Rejean ; Lorette, Guy: Automatic signature verification and writer identification - the state of the art. In: *Pattern Recognition* 22 (1989), Nr. 2, S. 107 – 131

[PL05] Park, J. ; Lee, C.: Robust iris recognition with region division. In: *Proceedings of SPIE - Image Processing: Algorithms and Systems IV* Bd. 5672, 2005, S. 161–168

[PM13] Parvez, Tanvir ; Mahmoud, Sabri A.: Offline Arabic Handwritten Text Recognition: A Survey. In: *ACM Computing Surveys* 45 (2013), Nr. 2, 23 - 35. `http://search.ebscohost.com/login.aspx?direct=true&db=aci&AN=87522711&site=ehost-live`

[RCB01] Ratha, Nalini K. ; Connell, Jonathan H. ; Bolle, Ruud M.: Enhancing security and privacy in biometrics-based authentication systems. In: *IBM Systems Journal* 40 (2001), Nr. 3, S. 614–634

[RCCB07] Ratha, Nalini K. ; Chikkerur, Sharat ; Connell, Jonathan H. ; Bolle, Ruud M.: Generating Cancelable Fingerprint Templates. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29 (2007), Nr. 4, S. 561–572

[RMR13] Rattani, Ajita ; Marcialis, Gian L. ; Roli, Fabio: Biometric System Adaptation by Self-update and Graph-based techniques. In: *Journal of Visual Languages and Computing, Elsevier* 24 (2013), S. 1–9

[RNJ06] Ross, Arun A. ; Nandakumar, Karthik ; Jain, Anil K.: *Handbook of Multibiometrics (International Series on Biometrics)*. Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2006

[RU10] Rathgeb, Christian ; Uhl, Andreas: Privacy Preserving Key Generation for Iris Biometrics. In: De Decker, B. (Hrsg.) ; Schaumüller-Bichl, I. (Hrsg.): *Communications and Multimedia Security*, Springer, 2010 (LNCS 6109), S. 191–200

[RU11] Rathgeb, Christian ; Uhl, Andreas: A survey on biometric cryptosystems and cancelable biometrics. In: *EURASIP J. Information Security* 2011 (2011), S. 3

[RUW11] Rathgeb, Christian ; Uhl, Andreas ; Wild, Peter: On Combining Selective Best Bits of Iris-Codes. In: Vielhauer, C. (Hrsg.) ; Dittmann, J. (Hrsg.) ; Drygajlo, A. (Hrsg.) ; Juul, N. C. (Hrsg.) ; Fairhurst, M. C. (Hrsg.): *Biometrics and ID Management*. Springer Berlin Heidelberg, 2011 (LNCS 6582), S. 227–237

[RW06] Rosenblum, Sara ; Werner, Perla: Assessing the handwriting process in healthy elderly persons using a computerized system. In: *Aging Clinical and Experimental Research* 18 (2006), Nr. 5, S. 433–439

[Sch99] Schmidt, Christiane: *On-line-Unterschriftenanalyse zur Benutzerverifikation*. Shaker Verlag GmbH, 1999 (Berichte aus der Elektrotechnik). – 173 S.

[SEV06] Scheidat, Tobias ; Engel, Andreas ; Vielhauer, Claus: Parameter Optimization for biometric Fingerprint Recognition using genetic Algorithms. In: *Proceedings of ACM 2006 Multimedia & Security Workshop*, 2006, S. 130 – 134

[Sic05] Sickler, N. C.: An Evaluation of Fingerprint Image Quality of an Elderly Population vis-à- vis an 18- 25 Year Old Baseline. In: *Proceedings Biometric Consortium 2003 Conference*, 2005

[SKV12] Scheidat, Tobias ; Kümmel, Karl ; Vielhauer, Claus: Short Term Template Aging Effects on Biometric Dynamic Handwriting Authentication Performance. In: de Decker, B. (Hrsg.) ; Chadwick, D. W. (Hrsg.): *Communications and Multimedia Security*, Springer Berlin Heidelberg, 2012 (LNCS 7394), 107-116

[SLM07] Sutcu, Yagiz ; Li, Qiming ; Memon, Nasir D.: Protecting Biometric Templates With Sketch: Theory and Practice. In: *IEEE Transactions on Information Forensics and Security* 2 (2007), Nr. 3-2, S. 503–512

[SMV07] Scheidat, Tobias ; Makrushin, Andrey ; Vielhauer, Claus: Automatic Template Update Strategies for Biometrics / Otto-von-Guericke University Magdeburg. 2007. – Forschungsbericht

[SPB96] Slavin, M.J. ; Phillips, J.G. ; Bradshaw, J.L.: Visual cues and the handwriting of older adults: a kinematic analysis. In: *Psychology and aging* 11 (1996), September, S. 521–526

[SRS+99] Soutar, C. ; Roberge, D. ; Stoianov, A. ; Gilroy, R. ; Kumar, B.V.K. V.: Biometric Encryption. In: Nichols, R. K. (Hrsg.): *ICSA Guide to Cryptography*, McGraw Hill, 1999

[SS14] Schneider, Matthias ; Schneider, Thomas: Notes on Non-Interactive Secure Comparison in Ïmage Feature Extraction in the Encrypted Domain with Privacy-Preserving SIFT". In: *Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec*, ACM, June 2014, S. 135–140

[SSM05] Sutcu, Yagiz ; Sencar, Husrev T. ; Memon, Nasir: A secure biometric authentication scheme based on robust hashing. In: *Proceedings of the 7th workshop on Multimedia and security*, ACM New York, NY, USA, 2005, S. 111–116

[Ste14] StepOver GmbH: *StepOver GmbH - Elektronische Signatur mittels Unterschriftenpads, Handys, Tablets oder zertifikatsbasierter Signaturen.* Version: 2014. `http://www.stepover.com/de/`, Abruf: 2014-10-06

[SV06] Scheidat, Tobias ; Vielhauer, Claus: Untersuchung der Möglichkeit eines biometrischen On Pen Matching. In: Horster, P. (Hrsg.): *Proceedings of D-A-CH Security Konferenz 2006*, Syssec, 2006, S. 392 – 404

[SVD05] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Distance-Level Fusion Strategies for Online Signature Verification. In: *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2005

[SVD07a] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Aspekte des Datenschutzes beim Umgang mit multi-modalen biometrischen Daten. In: *Innovationsmotor IT-Sicherheit - 10. Deutscher IT-Sicherheitskongress*, 2007, S. 213–228

[SVD07b] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Single-semantic multi-instance Fusion of Handwriting based biometric Authentication Systems. In: *2007 IEEE International Conference on Image Processing*, IEEE Operations Center, 2007, S. 393–396

[SVD07c] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Study of Possibility of On-pen Matching for Biometric Handwriting Verification. In: *Proceedings of the 15th European Signal Processing Conference (EUSIPCO 2007)*, 2007, S. 184–188

[SVD08a] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Advanced Studies on Reproducibility of Biometric Hashes. In: Schouten, B. (Hrsg.) ; Juul, N. C. (Hrsg.)

; Drygajlo, A. (Hrsg.) ; Tistarelli, M. (Hrsg.): *Biometrics and Identity Management, First European Workshop, BIOID 2008*, Springer Verlag Berlin, Heidelberg, 2008 (LNCS 5372), S. 150–159

[SVD08b] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Performanzmaße biometrischer Hashes am Beispiel Handschrift. In: Horster, P. (Hrsg.): *Proceedings of D-A-CH Security Konferenz 2008*, Syssec, 2008, S. 208–219

[SVD09a] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Biometric Hash Generation and User Authentication based on Handwriting using Secure Sketches. In: *Proceedings of 6th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2009

[SVD09b] Scheidat, Tobias ; Vielhauer, Claus ; Dittmann, Jana: Handwriting verification - comparison of a multi-algorithmic and a multi-semantic approach. In: *Image and vision computing* 27 (2009), February, Nr. 3, S. 269–278

[SVF11] Scheidat, Tobias ; Vielhauer, Claus ; Fischer, Robert: Comparative Study on Fusion Strategies for Biometric Handwriting. In: *Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security*, ACM New York, NY, USA, 2011, 61-68

[SVO07] Scheidat, Tobias ; Vielhauer, Claus ; Oermann, Andrea: Kombination von Sensoren zur biometrischen Handschriftenerkennung. In: Horster, P. (Hrsg.): *Proceedings of D-A-CH Security 2007*, syssec, 2007, S. 438–449

[SW73] Slepian, D. ; Wolf, J.: Noiseless coding of correlated information sources. In: *IEEE Transactions on Information Theory* 19 (1973), Nr. 4, S. 471–480

[TKL08] Teoh, Andrew B. J. ; Kuan, Yip W. ; Lee, Sangyoun: Cancellable biometrics and annotations on BioHash. In: *Pattern Recogn.* 41 (2008), June, S. 2034–2044

[TNG04a] Teoh, Andrew Beng J. ; Ngo, David Chek L. ; Goh, Alwyn: Biohashing: two factor authentication featuring fingerprint data and tokenised random number. In: *Pattern Recognition* 37 (2004), Nr. 11, S. 2245–2255

[TNG04b] Teoh, Andrew Beng J. ; Ngo, David Chek L. ; Goh, Alwyn: Personalised cryptographic key generation based on FaceHashing. In: *Computers & Security* 23 (2004), Nr. 7, S. 606–614

[TWW04] Tabassi, E. ; Wilson, C.L. ; Watson, C.: *NIST Fingerprint Image*. Version: 2004. `ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf`, Abruf: 2014-10-06

[UK 05] UK Passport Service: *Biometrics Enrolment Trial*. Version: 2005. `http://dematerialisedid.com/PDFs/UKPSBiometrics_Enrolment_Trial_Report.pdf`, Abruf: 2014-10-06. (May). – 299 S.

[Uni10] United Nations, Department of Economic and Social Affairs, Population Devision, Population Estimates and Projections Section: *World Population Prospects, the 2010 Revision*. Version: 2010. `http://esa.un.org/unpd/wpp/index.htm`, Abruf: 2012-05-02

[US06] Ueshige, Yoshifumi ; Sakurai, Kouichi: A Proposal of One-Time Biometric Authentication. In: *Security and Management*, 2006, S. 78–83

[Vie06] Vielhauer, Claus: *Biometric User Authentication for IT Security: From Fundamentals to Handwriting (Advances in Information Security)*. Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2006

[VS05] Vielhauer, Claus ; Scheidat, Tobias: Multimodal Biometrics for Voice and Handwriting. In: Dittmann, J. (Hrsg.) ; Katzenbeisser, S. (Hrsg.) ; Uhl, A. (Hrsg.): *Proceedings of 9th IFIP TC-6 TC-11 International Conference on Communications and Multimedia Security (CMS)*, Springer, Berlin, 2005 (LNCS 3677), S. 191–199

[VSM02] Vielhauer, Claus ; Steinmetz, Ralf ; Mayerhoefer, Astrid: Biometric Hash based on Statistical Features of Online Signatures. In: *International Conference on Pattern Recognition* 1 (2002), S. 10123

[WSV06] Wolf, Franziska ; Scheidat, Tobias ; Vielhauer, Claus: Study of Applicability of Virtual Users in Evaluating Multimodal Biometrics. In: *Multimedia Content Representation, Classification and Security* Bd. 4105, Springer Science & Business, New York, 2006, S. 554–561

Abbildungsverzeichnis

Tabellenverzeichnis

A Statistische Merkmale

In Tabelle A.1 sind die statistischen Merkmale aufgelistet, die als Ausgangsbasis zu Beginn dieser Arbeit zur Verfügung standen [Vie06]. Tabelle A.2 stellt die Merkmale dar, die im Rahmen dieser Arbeit entstanden. Die angegebenen Merkmale werden durch die Algorithmen Biometric Hash und Secure Sketch zur Verifikation beziehungsweise Hash-Generierung genutzt. Während die erste Spalte die Nummer eines Merkmals angibt, steht in Spalte *Name* der zugehörige Bezeichner und in Spalte *Beschreibung* eine kurze Erklärung des Merkmals.

Tabelle A.1: Liste der statistischen Merkmale zu Beginn der Arbeit [Vie06], die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
1	TTotal	Total writing time in ms
2	SampleCount	Total number of event pixels
3	AspectRatio	Image Width * 1000 DIV Height
4	VxAbsolute	Average velocity in x direction in 1000 * pixels / ms
5	VyAbsolute	Average velocity in y direction in 1000 * pixels / ms
6	SegmentCount	Number of consecutive pen-down segments
7	VxMin	Minimum absolute x-velocity during sample
8	VxMax	Maximum absolute x-velocity during sample
9	VyMin	Minimum absolute y-velocity during sample
10	VyMax	Maximum absolute y-velocity during sample
11	CentroidX	Centroid of horizontal pen position in bounding box
12	CentroidY	Centroid of vertical pen position in bounding box
13	CentroidDist	Distance of Centroid from origin
14	MaxPressure	Maximum absolute pressure occured during writing
15	CentroidX_SN	Centroid of horizontal pen position normalised to bounding box width * 1000
16	CentroidY_SN	Centroid of vertical pen position normalised to bounding box height * 1000
17	CentroidDist_SN	Distance of Centroid from origin normalised to bounding box diameter * 1000
18	CentroidAzimuth_SN	Horizontal azimuth of centroid from origin normalised to PI/2 * 1000
19	MaxAltitude	Maximum absolute altitude of pen occured during writing
20	MinAltitude	Minimum absolute altitude of pen occured during writing
21	MaxAzimuth	Maximum absolute azimuth of pen occured during writing
22	MinAzimuth	Minimum absolute azimuth of pen occured during writing
23	AvgPressure	Average Writing Pressure relative to MaxPressure * 1000
24	AverageAzimuth	Average Azimuth of pen projected on writing plane
25	AverageAltitude	Average Altitude of pen above the writing plane

Tabelle A.1: Liste der statistischen Merkmale zu Beginn der Arbeit [Vie06], die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
26	Vx_TN	Normalised Average velocity in x direction in pixels / VxMax * 1000
27	Vy_TN	Normalised Average velocity in y direction in pixels / VyMax * 1000
28	TPenUp	Absolute cummulated Pen-up time in ms
29	RatioTPenUpPEnDown	Ratio of TPenUp by TTotal * 1000
30	NoSamples	Total Number of Sample Values
31	PathLength	Total absolute Path Length in Pixels
32	PixelCountR1C1	Number of pixels in first row; first column
33	PixelCountR1C2	Number of pixels in first row; second column
34	PixelCountR1C3	Number of pixels in first row; third column
35	PixelCountR1C4	Number of pixels in first row; fourth column
36	PixelCountR2C1	Number of pixels in second row; first column
37	PixelCountR2C2	Number of pixels in second row; second column
38	PixelCountR2C3	Number of pixels in second row; third column
39	PixelCountR2C4	Number of pixels in second row; fourth column
40	PixelCountR3C1	Number of pixels in third row; first column
41	PixelCountR3C2	Number of pixels in third row; second column
42	PixelCountR3C3	Number of pixels in third row; third column
43	PixelCountR3C4	Number of pixels in third row; fourth column
44	IntegralX	Numeric Integration of normalised X values
45	IntegralY	Numeric Integration of normalised Y values
46	AreaX1	Numeric Integration of X values for 1st one-fifth time period
47	AreaX2	Numeric Integration of X values for 2nd one-fifth time period
48	AreaX3	Numeric Integration of X values for 3rd one-fifth time period
49	AreaX4	Numeric Integration of X values for 4th one-fifth time period
50	AreaX5	Numeric Integration of X values for 4th one-fifth time period
51	AreaY1	Numeric Integration of Y values for 1st one-fifth time period
52	AreaY2	Numeric Integration of Y values for 2nd one-fifth time period
53	AreaY3	Numeric Integration of Y values for 3rd one-fifth time period
54	AreaY4	Numeric Integration of Y values for 4th one-fifth time period
55	AreaY5	Numeric Integration of Y values for 4th one-fifth time period
56	PenDPress	Average PenDown Pressure normalized to 1 * 1000
57	PenUPress	Average PenUp Pressure normalized to 1 * 1000
58	BaselineAngle	Baseline Angle of the Sample

Tabelle A.1: Liste der statistischen Merkmale zu Beginn der Arbeit [Vie06], die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
59	HistYZone1	Histogram of Y for Zone 1 in % * 100
60	HistYZone2	Histogram of Y for Zone 2 in % * 100
61	HistYZone3	Histogram of Y for Zone 3 in %* 100
62	AreaRatio1	Area(ConvexHull) vs Area(BoundingBox) * 1000
63	AreaRatio2	Area(ConvexHull(Segments)) vs Area(ConvexHull(Sample)) * 1000
64	AreaRatio3	Area(ConvexHull(Segments)) vs Area(BoundingBox) * 1000
65	PathRatio1	PathLength(ConvexHull) vs PathLength(BoundingBox) * 1000
66	PathRatio2	PathLength(ConvexHull(Segments)) vs PathLength(ConvexHull(Sample)) * 1000
67	PathRatio3	PathLength(ConvexHull(Segments)) vs PathLength(BoundingBox) * 1000
68	HistXLeft	Histogram of X for left in % * 100
69	HistXRight	Histogram of X for right in % * 100

Tabelle A.2: Liste der im Rahmen dieser Arbeit entstandenen statistischen Merkmale, die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
70	NoMaximums_X	Number of maximum points in the X signal
71	NoMinimums_X	Number of minimum points in the X signal
72	NoMaximums_Y	Number of maximum points in the Y signal
73	NoMinimums_Y	Number of minimum points in the Y signal
74	NoMaximumsRatio	NoMaximums_X * 1000 DIV NoMaximums_Y
75	NoMinimumsRatio	NoMinimums_X * 1000 DIV NoMinimums_Y
76	NoIntersections	Number of intersections within the sample itself
77	NoIntersections_X1	Number of intersections of the vertical line X1 with the sample
78	NoIntersections_X2	Number of intersections of the vertical line X2 with the sample
79	NoIntersections_X3	Number of intersections of the vertical line X3 with the sample
80	NoIntersections_X4	Number of intersections of the vertical line X4 with the sample
81	NoIntersections_Y1	Number of intersections of the horizontal line Y1 with the sample
82	NoIntersections_Y2	Number of intersections of the horizontal line Y2 with the sample
83	NoIntersections_Y3	Number of intersections of the horizontal line Y3 with the sample
84	NoIntersections_D1	Number of intersections of the diagonal line D1 with the sample

Tabelle A.2: Liste der im Rahmen dieser Arbeit entstandenen statistischen Merkmale, die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
85	NoIntersections_D2	Number of intersections of the diagonal line D2 with the sample
86	StartEndRatio	Distance (Start Point; End Point) * 1000 DIV PathLength
87	XMaxXMinRatio	Distance (Maximum X Point; Minimum X Point) * 1000 DIV PathLength
88	YMaxYMinRatio	Distance (Maximum Y Point; Minimum Y Point) * 1000 DIV PathLength
89	StartCentroidEndRatio	Distance (Start Point; Centroid) * 1000 DIV Distance (End Point; Centroid)
90	fmapX	mapped Function of X-Extrema
91	fmapY	mapped Function of Y-Extrema
92	fmapP	mapped Function of Pressure-Extrema
93	fmapa	mapped Function of Acceleration-Extrema
94	intRange	average distance between segments * 1000
95	dotcnt_min	minima Number of neighbour Points
96	dotcmt_max	maxima Number of neighbour Points
97	dotcnt_avg	average Number of neighbour Points
98	angle30	average angle between 0-30 degrees
99	angle60	average angle between 30-60 degrees
100	angle90	average angle between 60-90 degrees
101	count30	count of angles between 0-30 degrees
102	count60	count of angles between 30-60 degrees
103	count90	count of angles between 60-90 degrees
104	MaxPenUpDist	maximum distance between two strokes
105	MinPenUpDist	minimum distance between two strokes
106	AvgPenUpDist	average distance between two strokes
107	areas	fmapped size of enclosed areas
108	Cluster0X	normalized X-coordinate of cluster 0
109	Cluster0Y	normalized Y-coordinate of cluster 0
110	Cluster1X	normalized X-coordinate of cluster 1
111	Cluster1Y	normalized Y-coordinate of cluster 1
112	Cluster2X	normalized X-coordinate of cluster 2
113	Cluster3Y	normalized Y-coordinate of cluster 2
114	Cluster4X	normalized X-coordinate of cluster 3
115	Cluster4Y	normalized Y-coordinate of cluster 3
116	Cluster5X	normalized X-coordinate of cluster 4
117	Cluster5Y	normalized Y-coordinate of cluster 4
118	Cluster6X	normalized X-coordinate of cluster 5
119	Cluster6Y	normalized Y-coordinate of cluster 5
120	Cluster0XYAvgPrs	average pressure of xy cluster 0
121	Cluster1XYAvgPrs	average pressure of xy cluster 1
122	Cluster2XYAvgPrs	average pressure of xy cluster 2
123	Cluster3XYAvgPrs	average pressure of xy cluster 3
124	Cluster4XYAvgPrs	average pressure of xy cluster 4
125	Cluster5XYAvgPrs	average pressure of xy cluster 5

Tabelle A.2: Liste der im Rahmen dieser Arbeit entstandenen statistischen Merkmale, die für die Bestimmung des Biometric Hash und der Secure Sketch verwendet wurden

Nr.	Name	Beschreibung
126	Cluster0Alti	average altitude of pressure cluster 0
127	Cluster1Alti	average altitude of pressure cluster 1
128	Cluster2Alti	average altitude of pressure cluster 2
129	Cluster3Alti	average altitude of pressure cluster 3
130	InflectionPointSpeed	fmapped speed at the inflection point of the sample
131	PressureDeviation	standard deviation of the Pressure

B Verwendete Software und Hardware

Zu Beginn der vorliegenden Arbeit lagen die damals aktuellen Versionen der Algorithmen, wie auch die Akquirierungs- und Evaluierungsumgebung PlataSign in Delphi 5.0 Code vor. Aus hauptsächlich zwei Gründen wurden sowohl die Algorithmen und Evaluierungsmechanismen in MatLab portiert als auch die statistischen Verfahren zur Analyse der Merkmale mittels dieser Software umgesetzt. Zum einen ermöglicht eine Neuimplementierung des Codes durch Konzentration auf die wesentlichen Bestandteile der Algorithmen und Verfahren eine bessere Übersichtlichkeit und Pflege des Codes. Das erleichtert beispielsweise die schnelle Anpassung der Vielzahl von Einstellungen bezüglich Parametrisierung, Merkmalsselektion, Fusion und Alterung. Zusätzlich wurden auch Funktionen und Prozeduren in MatLab geschrieben, die die reproduzierbare Durchführung von Testabläufen für die verschiedenen Szenarien (Verifikation, Hash-Generierung für Optimierung, Alterung) erlauben. Zweitens haben wir uns für die Programmierung unter MatLab entschieden, weil dieses vielfältige Möglichkeiten zur statistischen Analyse und Auswertung von Daten und Ergebnissen bietet. Genutzt wurden diese vor allem bei der Umsetzung und Auswertung der verschiedenen Strategien zur Merkmalsselektion und zur grafischen Aufbereitung der Evaluierungsergebnisse.

Die biometrischen Handschriftendaten wurden mit der Erfassungs- und Evaluierungssoftware PlataSign aufgenommen und in eine MySQL-Datenbank gespeichert. Diese Daten wurden im späteren Verlauf in CSV-Dateien exportiert, welche die Sequenzen mit den physikalischen Informationen der einzelnen Samples enthalten. Über ein MatLab-Programm wurden daraus nach der Implementierung der zusätzlichen Merkmale die statistischen Merkmalsvektoren generiert, die die Grundlage für die beiden Referenzalgorithmen darstellen. Basierend auf diesen Daten wurden die einzelnen Evaluierungen durchgeführt.

Für die Evaluierung der unterschiedlichen Konstellationen für Optimierung und Alterung wurde ein Desktop PC mit der folgenden Konfiguration verwendet:

Tabelle B.1: Ausstattung des Evaluierungssystems

Komponente	Ausstattung
Prozessor	Intel Quad Core i5-2320 3.00 GHz
Speicher	4 GB DDR3, 1333 GHz
Festplatte	Crucial SSD, 128 GB
Betriebssystem	Microsoft Windows 7 Professional (64bit)
Evaluierungssoftware	MathWorks MATLAB R2010b

C Evaluierungsergebnisse

In diesem Teil des Anhangs werden neben den im Hauptteil der Arbeit angegebenen Ergebnisse auch die dargestellt, die aus Platzgründen dort nicht aufgeführt werden konnten. Dies bezieht sich auf die Tabellen zur Auswertung der Ergebnisse der Merkmalsselektion in Abschnitt 7.3. Die Tabellen enthalten die folgenden Informationen:

Ergebnisse der Merkmalsselektion für den Biometric Hash-Algorithmus. In den Tabellen C.1 bis C.10 werden jeweils die Ergebnisse für die Szenarien *bestEER*, *bestCRR* und *fix60* für die fünf Parameter-Sets und die fünf Parameter-Sets dargestellt.

Ergebnisse der Merkmalsselektion für den Secure Sketch-Algorithmus. In den Tabellen C.11 bis C.20 werden jeweils die Ergebnisse für die Szenarien *bestEER*, *bestCRR* und *fix60* für die fünf Parameter-Sets und die fünf Parameter-Sets dargestellt.

Wrapper-basierte Rankings für den Biometric Hash-Algorithmus. In den Tabellen C.21 bis C.25 werden jeweils die Ergebnisse für die Szenarien *bestEER*, *bestCRR* und *fix60* für die fünf Parameter-Sets und die fünf Parameter-Sets dargestellt.

Wrapper-basierte Rankings für den Secure Sketch-Algorithmus. In den Tabellen C.26 bis C.30 werden jeweils die Ergebnisse für die Szenarien *bestEER*, *bestCRR* und *fix60* für die fünf Parameter-Sets und die fünf Parameter-Sets dargestellt.

Filter-basierte Rankings. Tabelle C.31 zeigt die Ergebnisse für die fünf Semantiken.

Tabelle C.1: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,18536	0,50000	0,00000	0,00000	131	*0,18536*	0,50000	0,00000	0,00000
anova	103	0,18389	0,50000	0,00000	0,00000	3	**0,35484**	0,35229	0,50000	0,20457
anova-2class	64	**0,17913**	0,50000	0,00000	0,00000	1	0,47985	0,46667	0,86038	0,79372
correlation	117	0,18469	0,50000	0,00000	0,00000	1	0,37064	**0,34722**	0,56226	0,25671
entropy	112	0,18270	0,50000	0,00000	0,00000	1	0,47985	0,46667	0,86038	0,79372
joint-entropy	131	0,18536	0,50000	0,00000	0,00000	2	0,36249	0,41147	0,24528	0,06822
simple	66	***0,17042***	0,50000	0,00000	0,00000	1	0,33494	0,31682	0,62830	0,26194
sfs	64	0,18480	***0,49811***	0,00377	0,00000	2	**0,32209**	***0,31626***	0,50566	0,13817
sbs	104	0,18471	0,50000	0,00000	0,00000	2	0,36816	0,35386	0,54906	0,25679
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,13036	0,50000	0,00000	0,00000	131	*0,13036*	0,50000	0,00000	0,00000
anova	101	0,12772	0,50000	0,00000	0,00000	2	0,35723	0,36745	0,33585	0,07076
anova-2class	114	**0,12671**	0,50000	0,00000	0,00000	6	0,41826	0,45459	0,16038	0,06956
correlation	118	0,13036	0,50000	0,00000	0,00000	1	**0,33463**	**0,31818**	0,63208	0,26843
entropy	116	0,12788	0,50000	0,00000	0,00000	2	0,47209	0,47219	0,51132	0,45570
joint-entropy	131	0,13036	0,50000	0,00000	0,00000	1	0,40626	0,37157	0,44340	0,18654
simple	58	***0,11927***	0,50000	0,00000	0,00000	1	**0,26080**	0,25639	0,75283	0,26560
sfs	104	0,12412	***0,49906***	0,00189	0,00000	2	0,26916	***0,25345***	0,70943	0,21633
sbs	105	0,12047	0,50000	0,00000	0,00000	1	0,38521	0,35718	0,52075	0,23512
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,11338	0,50000	0,00000	0,00000	131	*0,11338*	0,50000	0,00000	0,00000
anova	50	0,10778	0,50000	0,00000	0,00000	1	**0,34585**	**0,29900**	0,54906	0,14706
anova-2class	117	0,11203	0,50000	0,00000	0,00000	7	0,41152	0,46800	0,08679	0,02279
correlation	98	**0,10202**	0,50000	0,00000	0,00000	1	0,39623	0,37547	0,52453	0,27547
entropy	117	0,11301	0,50000	0,00000	0,00000	1	0,40768	0,38737	0,72264	0,49739
joint-entropy	105	0,10855	0,50000	0,00000	0,00000	2	0,35199	0,36861	0,33396	0,07119
simple	56	***0,08318***	0,50000	0,00000	0,00000	1	0,29854	0,27464	0,66604	0,21531
sfs	112	0,10910	0,50000	0,00000	0,00000	2	**0,28097**	***0,26321***	0,54906	0,07547
sbs	118	0,11338	0,50000	0,00000	0,00000	1	0,33224	0,29040	0,58491	0,16571
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12512	0,50000	0,00000	0,00000	131	*0,12512*	0,50000	0,00000	0,00000
anova	101	0,11963	0,50000	0,00000	0,00000	2	0,40822	0,37417	0,81132	0,55965
anova-2class	67	**0,09934**	0,50000	0,00000	0,00000	7	0,40773	0,44278	0,20566	0,09122
correlation	85	0,11363	0,50000	0,00000	0,00000	1	0,38765	0,37587	0,57170	0,32344
entropy	115	0,12350	0,50000	0,00000	0,00000	4	0,40741	0,39868	0,48491	0,28226
joint-entropy	126	0,12436	0,50000	0,00000	0,00000	1	**0,35174**	**0,32186**	0,57736	0,22108
simple	22	***0,08850***	***0,49717***	0,00566	0,00000	1	**0,26978**	***0,23755***	0,69245	0,16756
sfs	91	0,10937	0,49906	0,00189	0,00000	1	**0,26978**	***0,23755***	0,69245	0,16756
sbs	105	0,11774	0,50000	0,00000	0,00000	2	0,38542	0,36379	0,52075	0,24833
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,09528	0,50000	0,00000	0,00000	131	*0,09528*	0,50000	0,00000	0,00000
anova	116	0,09447	0,50000	0,00000	0,00000	1	0,34759	**0,29399**	0,53019	0,11818
anova-2class	81	0,09008	0,50000	0,00000	0,00000	5	0,42683	0,46212	0,16226	0,08650
correlation	118	0,09528	0,50000	0,00000	0,00000	1	0,34430	0,33064	0,71321	0,37449
entropy	69	**0,08539**	0,50000	0,00000	0,00000	1	0,39509	0,39329	0,59811	0,38469
joint-entropy	131	0,09528	0,50000	0,00000	0,00000	2	**0,30417**	0,31762	0,43208	0,06731
simple	52	***0,08505***	0,50000	0,00000	0,00000	1	**0,26356**	***0,23257***	0,70189	0,16702
sfs	118	0,09528	0,50000	0,00000	0,00000	1	**0,26356**	***0,23257***	0,70189	0,16702
sbs	114	0,09425	0,50000	0,00000	0,00000	2	0,37025	0,35630	0,57170	0,28429

Tabelle C.2: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für TF für Szenario fix60

selection	gegebene PIN n	EER	CRR	RR	CR	geheime PIN n	EER	CRR	RR	CR
raw	60	0,23081	0,49906	0,00189	0,00000	60	0,18235	0,50000	0,00000	0,00000
anova	60	0,22176	0,50000	0,00000	0,00000	60	**0,14917**	0,50000	0,00000	0,00000
anova-2class	60	**0,18438**	0,50000	0,00000	0,00000	60	0,15176	0,50000	0,00000	0,00000
correlation	60	0,23724	0,50000	0,00000	0,00000	60	0,17219	0,50000	0,00000	0,00000
entropy	60	0,20225	0,50000	0,00000	0,00000	60	0,14931	0,50000	0,00000	0,00000
joint-entropy	60	0,22600	0,50000	0,00000	0,00000	60	0,17486	0,50000	0,00000	0,00000
wrapper	60	***0,17441***	0,50000	0,00000	0,00000	60	***0,12463***	0,50000	0,00000	0,00000
sfs	60	0,19542	***0,49811***	0,00377	0,00000	60	0,16906	***0,49717***	0,00566	0,00000
sbs	60	0,20451	0,50000	0,00000	0,00000	60	0,15035	0,50000	0,00000	0,00000

selection	Pseudonym n	EER	CRR	RR	CR	Symbol n	EER	CRR	RR	CR
raw	60	0,12331	0,50000	0,00000	0,00000	60	0,16666	0,49906	0,00189	0,00000
anova	60	**0,12296**	0,50000	0,00000	0,00000	60	0,14501	0,50000	0,00000	0,00000
anova-2class	60	0,12861	0,50000	0,00000	0,00000	60	**0,10459**	0,50000	0,00000	0,00000
correlation	60	0,12414	0,50000	0,00000	0,00000	60	0,14274	0,50000	0,00000	0,00000
entropy	60	0,12362	0,50000	0,00000	0,00000	60	0,15935	0,50000	0,00000	0,00000
joint-entropy	60	0,14073	0,50000	0,00000	0,00000	60	0,20860	0,49906	0,00189	0,00000
wrapper	60	***0,08839***	0,50000	0,00000	0,00000	60	***0,10216***	0,50000	0,00000	0,00000
sfs	60	0,13048	***0,49528***	0,00943	0,00000	60	0,13898	***0,49434***	0,01132	0,00000
sbs	60	0,15840	0,50000	0,00000	0,00000	60	0,15496	0,50000	0,00000	0,00000

selection	Woher n	EER	CRR	RR	CR
raw	60	0,13232	0,50000	0,00000	0,00000
anova	60	0,11616	0,50000	0,00000	0,00000
anova-2class	60	0,09997	0,50000	0,00000	0,00000
correlation	60	0,12808	0,50000	0,00000	0,00000
entropy	60	**0,09378**	0,50000	0,00000	0,00000
joint-entropy	60	0,14013	0,50000	0,00000	0,00000
wrapper	60	***0,08678***	0,50000	0,00000	0,00000
sfs	60	0,12642	***0,49623***	0,00755	0,00000
sbs	60	0,12412	0,50000	0,00000	0,00000

Tabelle C.3: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,20082	0,49253	0,01509	0,00015	131	*0,20082*	0,49253	0,01509	0,00015
anova	118	0,20082	0,49253	0,01509	0,00015	6	0,29120	0,27778	0,61509	0,17065
anova-2class	63	***0,17395***	0,47395	0,05283	0,00073	9	0,33073	0,32201	0,70377	0,34779
correlation	29	0,20041	**0,44294**	0,12264	0,00853	4	**0,27336**	**0,26693**	0,66981	0,20366
entropy	65	0,17922	0,47854	0,04340	0,00047	13	0,33044	0,33091	0,59245	0,25428
joint-entropy	103	0,19956	0,48507	0,03019	0,00033	8	0,28924	0,29572	0,57170	0,16313
simple	25	0,18745	***0,42975***	0,14528	0,00479	3	0,27172	0,26669	0,70566	0,23904
sfs	88	**0,18514**	0,45675	0,08679	0,00029	7	**0,23043**	***0,21892***	0,72075	0,15860
sbs	118	0,20082	0,49253	0,01509	0,00015	7	0,26928	0,26319	0,72453	0,25091
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12790	0,49906	0,00189	0,00000	131	*0,12790*	0,49906	0,00189	0,00000
anova	115	**0,12613**	0,49906	0,00189	0,00000	4	0,27756	0,30998	0,45472	0,07467
anova-2class	118	0,12790	0,49906	0,00189	0,00000	8	**0,27380**	**0,27005**	0,71321	0,25330
correlation	118	0,12790	0,49906	0,00189	0,00000	3	0,30402	0,28643	0,57925	0,15210
entropy	87	0,12789	**0,49813**	0,00377	0,00004	11	0,29899	0,30261	0,56792	0,17315
joint-entropy	131	0,12790	0,49906	0,00189	0,00000	3	0,29930	0,30778	0,50943	0,12500
simple	115	0,12613	0,49906	0,00189	0,00000	3	0,22961	0,21234	0,75472	0,17939
sfs	116	0,12679	0,49811	0,00377	0,00000	5	**0,22514**	***0,20463***	0,70000	0,10925
sbs	100	***0,12187***	***0,49340***	0,01321	0,00000	4	0,31121	0,28786	0,56415	0,13988
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08943	0,45004	0,10000	0,00007	131	*0,08943*	0,45004	0,10000	0,00007
anova	118	0,08943	0,45004	0,10000	0,00007	11	**0,18880**	**0,22003**	0,62453	0,06459
anova-2class	111	0,08818	0,44534	0,10943	0,00011	17	0,20076	0,22235	0,62453	0,06923
correlation	104	**0,08422**	0,44439	0,11132	0,00011	6	0,23060	0,23015	0,66038	0,12068
entropy	111	0,08892	**0,44343**	0,11321	0,00007	12	0,22532	0,22153	0,72264	0,16571
joint-entropy	124	0,08869	0,44909	0,10189	0,00007	14	0,21824	0,22685	0,66038	0,11408
simple	99	0,08844	0,43687	0,12642	0,00015	5	0,18947	0,16881	0,77925	0,11687
sfs	106	***0,08515***	***0,41513***	0,16981	0,00007	9	0,13690	0,13523	0,78868	0,05914
sbs	112	0,08582	0,42647	0,14717	0,00011	9	**0,11919**	***0,11560***	0,87736	0,10856
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,07366	0,13610	0,75472	0,02692	131	0,07366	0,13610	0,75472	0,02692
anova	98	0,07229	**0,11854**	0,79434	0,03142	35	0,09572	**0,07992**	0,93962	0,09946
anova-2class	113	0,07151	0,13458	0,75849	0,02765	77	0,10521	0,10983	0,85283	0,07250
correlation	118	0,07366	0,13610	0,75472	0,02692	60	0,10276	0,10185	0,85283	0,05653
entropy	118	0,07366	0,13610	0,75472	0,02692	71	0,11724	0,11168	0,85472	0,07808
joint-entropy	111	**0,07150**	0,11978	0,79057	0,03012	86	**0,08145**	0,10952	0,82075	0,03980
simple	76	0,07048	0,10684	0,81887	0,03255	35	0,08150	0,08082	0,91509	0,07674
sfs	79	***0,04339***	***0,04338***	0,95660	0,04336	75	***0,04609***	0,04122	0,96415	0,04659
sbs	83	0,05039	0,04933	0,94906	0,04771	46	0,06382	***0,03672***	0,99434	0,06778
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08559	0,50000	0,00000	0,00000	131	*0,08559*	0,50000	0,00000	0,00000
anova	112	0,08419	0,49906	0,00189	0,00000	1	0,25035	0,22529	0,72453	0,17511
anova-2class	61	0,07236	0,49811	0,00377	0,00000	8	0,24160	0,24508	0,60566	0,09583
correlation	118	0,08559	0,50000	0,00000	0,00000	2	0,34888	0,33423	0,55094	0,21941
entropy	68	**0,06257**	**0,49811**	0,00377	0,00000	9	**0,21633**	**0,22137**	0,62642	0,06916
joint-entropy	131	0,08559	0,50000	0,00000	0,00000	2	0,26480	0,23902	0,65849	0,13654
simple	40	***0,07488***	***0,48213***	0,03585	0,00011	2	0,18211	0,17509	0,77358	0,12377
sfs	117	0,08541	0,50000	0,00000	0,00000	2	**0,17358**	***0,15292***	0,80000	0,10584
sbs	118	0,08559	0,50000	0,00000	0,00000	3	0,21177	0,20867	0,78302	0,20036

Tabelle C.4: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,22926	0,44911	0,10377	0,00200	60	0,18448	0,48396	0,03208	0,00000
anova	60	0,21671	0,47600	0,04906	0,00105	60	0,15359	0,49245	0,01509	0,00000
anova-2class	60	**0,17936**	0,46925	0,06226	0,00076	60	**0,14133**	0,49530	0,00943	0,00004
correlation	60	0,21724	0,47242	0,05660	0,00145	60	0,17521	0,49247	0,01509	0,00004
entropy	60	0,18733	0,47493	0,05094	0,00080	60	0,14514	0,49343	0,01321	0,00007
joint-entropy	60	0,23361	**0,45633**	0,08868	0,00134	60	0,16982	**0,48962**	0,02076	0,00000
simple	60	***0,20598***	0,47763	0,04528	0,00054	60	0,13891	0,48774	0,02453	0,00000
sfs	60	0,21082	***0,39604***	0,21321	0,00530	60	0,15626	***0,44915***	0,10189	0,00018
sbs	60	0,23749	0,40902	0,18491	0,00294	60	***0,13519***	0,48019	0,03962	0,00000

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,12023	0,37277	0,25660	0,00214	60	0,09800	0,09967	0,86415	0,06350
anova	60	0,10563	0,39372	0,21321	0,00065	60	**0,08304**	**0,09570**	0,85094	0,04234
anova-2class	60	0,10558	**0,37944**	0,24151	0,00040	60	0,12007	0,11892	0,86792	0,10577
correlation	60	0,12013	0,40158	0,19811	0,00127	60	0,10276	0,10185	0,85283	0,05653
entropy	60	**0,10018**	0,38229	0,23585	0,00044	60	0,12100	0,11707	0,86038	0,09452
joint-entropy	60	0,12854	0,38440	0,23396	0,00276	60	0,12017	0,11676	0,85094	0,08447
simple	60	***0,09238***	0,39082	0,21887	0,00051	60	0,07358	0,09329	0,85094	0,03752
sfs	60	0,10799	***0,28329***	0,43585	0,00243	60	***0,04843***	***0,04156***	0,96604	0,04917
sbs	60	0,10735	0,30210	0,40000	0,00421	60	0,05975	0,04414	0,97358	0,06187

	Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,12345	0,48868	0,02264	0,00000
anova	60	0,10202	**0,49528**	0,00943	0,00000
anova-2class	60	0,07454	0,49717	0,00566	0,00000
correlation	60	0,11161	0,49717	0,00566	0,00000
entropy	60	**0,06684**	0,49717	0,00566	0,00000
joint-entropy	60	0,13491	0,49623	0,00755	0,00000
simple	60	***0,08287***	0,49340	0,01321	0,00000
sfs	60	0,10239	***0,45849***	0,08302	0,00000
sbs	60	0,11028	0,46511	0,06981	0,00004

Tabelle C.5: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minCRR für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,22124	0,22206	0,73019	0,17431	131	0,22124	0,22206	0,73019	0,17431
anova	100	**0,20792**	**0,20762**	0,79057	0,20581	83	0,21762	**0,20172**	0,82642	0,22986
anova-2class	92	0,21720	0,21030	0,80189	0,22250	86	0,21738	0,20633	0,81321	0,22587
correlation	92	0,21926	0,21912	0,78113	0,21938	81	0,23269	0,21802	0,80943	0,24546
entropy	118	0,22124	0,22206	0,73019	0,17431	78	0,22479	0,21934	0,79057	0,22925
joint-entropy	107	0,21650	0,21326	0,79245	0,21898	107	**0,21650**	0,21326	0,79245	0,21898
simple	72	0,20373	0,20337	0,79434	0,20109	72	0,20373	0,20337	0,79434	0,20109
sfs	106	***0,18835***	0,18703	0,81509	0,18915	74	***0,20207***	0,17030	0,88302	0,22362
sbs	102	0,20109	***0,17456***	0,86981	0,21894	75	0,20759	***0,16232***	0,91509	0,23973
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,20295	0,20758	0,72830	0,14347	131	0,20295	0,20758	0,72830	0,14347
anova	89	**0,18638**	0,18509	0,81698	0,18716	87	**0,19383**	0,18327	0,83396	0,20051
anova-2class	90	0,19276	**0,18331**	0,83208	0,19869	74	0,20348	**0,18111**	0,85660	0,21883
correlation	118	0,20295	0,20758	0,72830	0,14347	109	0,20699	0,20686	0,75283	0,16655
entropy	118	0,20295	0,20758	0,72830	0,14347	110	0,20539	0,20737	0,73396	0,14869
joint-entropy	108	0,20183	0,20254	0,76226	0,16734	102	0,20186	0,20229	0,76415	0,16872
simple	104	0,19500	0,19666	0,75472	0,14804	106	0,19519	0,19634	0,75472	0,14739
sfs	104	***0,16431***	0,16121	0,84340	0,16582	80	***0,17791***	0,14960	0,89434	0,19354
sbs	103	0,17480	***0,15425***	0,87736	0,18585	71	0,19366	***0,13944***	0,94906	0,22794
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,15291	0,17099	0,73962	0,08160	131	0,15291	0,17099	0,73962	0,08160
anova	78	**0,13296**	**0,13168**	0,86226	0,12562	65	**0,13604**	**0,12404**	0,89245	0,14053
anova-2class	74	0,14387	0,14287	0,85283	0,13857	70	0,14681	0,13915	0,87170	0,15000
correlation	111	0,14943	0,16023	0,76226	0,08273	91	0,15293	0,15464	0,78679	0,09608
entropy	118	0,15291	0,17099	0,73962	0,08160	118	0,15291	0,17099	0,73962	0,08160
joint-entropy	112	0,14987	0,15784	0,77358	0,08926	112	0,14987	0,15784	0,77358	0,08926
simple	58	0,14133	0,13902	0,84906	0,12710	65	0,14157	0,13819	0,83774	0,11411
sfs	105	0,11449	0,11393	0,88679	0,11466	72	***0,12626***	0,08436	0,97170	0,14042
sbs	105	***0,11272***	***0,10882***	0,89623	0,11386	60	0,13553	***0,08316***	0,98868	0,15501
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,09259	0,09048	0,89623	0,07718	131	0,09259	0,09048	0,89623	0,07718
anova	105	**0,08105**	**0,08023**	0,92075	0,08121	84	0,08870	**0,06910**	0,95472	0,09293
anova-2class	114	0,09187	0,08971	0,89811	0,07754	114	0,09187	0,08971	0,89811	0,07754
correlation	106	0,09043	0,08882	0,90377	0,08142	84	0,09467	0,08861	0,91887	0,09608
entropy	116	0,09049	0,08853	0,90189	0,07896	116	0,09049	0,08853	0,90189	0,07896
joint-entropy	111	0,08316	0,08224	0,91887	0,08335	111	***0,08316***	0,08224	0,91887	0,08335
simple	90	0,07975	0,07952	0,92075	0,07979	56	0,09466	0,07255	0,95472	0,09982
sfs	114	***0,07831***	***0,07442***	0,93019	0,07903	102	**0,08784**	0,06094	0,97170	0,09358
sbs	114	***0,07831***	***0,07442***	0,93019	0,07903	77	0,08905	***0,05994***	0,97547	0,09536
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12157	0,15688	0,74528	0,05904	131	0,12157	0,15688	0,74528	0,05904
anova	63	**0,11236**	**0,11221**	0,88679	0,11121	63	**0,11236**	**0,11221**	0,88679	0,11121
anova-2class	118	0,12157	0,15688	0,74528	0,05904	91	0,12193	0,12676	0,83019	0,08371
correlation	110	0,12012	0,15181	0,75849	0,06212	57	0,14855	0,14565	0,85849	0,14978
entropy	118	0,12157	0,15688	0,74528	0,05904	47	0,14693	0,12204	0,91321	0,15729
joint-entropy	108	0,11842	0,14187	0,78868	0,07242	69	0,14910	0,13819	0,87736	0,15374
simple	113	0,12157	0,15688	0,74528	0,05904	59	0,12965	0,13467	0,81698	0,08632
sfs	109	0,09222	0,09214	0,90755	0,09184	90	***0,09997***	0,07398	0,95849	0,10646
sbs	105	***0,09215***	***0,08296***	0,92830	0,09423	80	0,10002	***0,07066***	0,96604	0,10737

Tabelle C.6: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung minCRR für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,27542	0,24035	0,83774	0,31843	60	0,25993	0,22580	0,84528	0,29688
anova	60	**0,24325**	**0,21257**	0,84717	0,27231	60	**0,21614**	**0,19526**	0,84151	0,23204
anova-2class	60	0,27704	0,23931	0,84528	0,32391	60	0,26532	0,22397	0,86415	0,31208
correlation	60	0,26319	0,23075	0,83774	0,29924	60	0,23115	0,22620	0,78302	0,23541
entropy	60	0,25613	0,23463	0,80943	0,27870	60	0,24638	0,23839	0,77736	0,25414
joint-entropy	60	0,29974	0,25287	0,86415	0,36988	60	0,26962	0,24171	0,81887	0,30229
simple	60	0,21753	0,21052	0,80189	0,22293	60	0,20086	0,20020	0,79057	0,19097
sfs	60	***0,20377***	0,17148	0,88302	0,22598	60	***0,18513***	0,15151	0,90189	0,20490
sbs	60	0,21385	***0,16348***	0,92453	0,25149	60	0,19767	***0,14167***	0,95094	0,23429

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,18465	0,16724	0,86038	0,19485	60	0,12649	0,09338	0,95094	0,13770
anova	60	**0,14234**	**0,12783**	0,89245	0,14811	60	**0,09827**	**0,07382**	0,95660	0,10425
anova-2class	60	0,18010	0,15695	0,87925	0,19314	60	0,20519	0,14868	0,94717	0,24452
correlation	60	0,19222	0,16571	0,87736	0,20878	60	0,12980	0,09786	0,94528	0,14100
entropy	60	0,18020	0,17788	0,80755	0,16332	60	0,17716	0,12912	0,94528	0,20352
joint-entropy	60	0,23285	0,19316	0,88113	0,26745	60	0,15204	0,11520	0,93774	0,16814
simple	60	0,14580	0,14316	0,83962	0,12594	60	0,09334	0,07602	0,94528	0,09732
sfs	60	***0,12720***	0,08496	0,97170	0,14162	60	0,09066	0,06174	0,97358	0,09706
sbs	60	0,13553	***0,08316***	0,98868	0,15501	60	***0,09050***	***0,06080***	0,97547	0,09706

	Woher			
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,15008	0,14949	0,84151	0,14049
anova	60	**0,12422**	**0,11869**	0,88868	0,12605
anova-2class	60	0,14880	0,14735	0,85472	0,14942
correlation	60	0,15944	0,16154	0,79057	0,11364
entropy	60	0,14661	0,15192	0,79811	0,10196
joint-entropy	60	0,17868	0,15085	0,89245	0,19416
simple	60	0,13119	0,13728	0,81132	0,08589
sfs	60	0,10670	0,07551	0,96415	0,11517
sbs	60	***0,10233***	***0,07121***	0,96792	0,11034

Tabelle C.7: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung $CR \leq 5\%$ für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER gegebene PIN					(b) bestCRR gegebene PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,19783	0,28759	0,46604	0,04122	131	0,19783	0,28759	0,46604	0,04122
anova	101	0,19624	0,27103	0,50566	0,04771	45	0,21482	0,22213	0,67736	0,12163
anova-2class	67	**0,17646**	**0,24156**	0,60000	0,08313	49	**0,20529**	**0,20657**	0,73208	0,14521
correlation	118	0,19783	0,28759	0,46604	0,04122	21	0,23674	0,20780	0,84717	0,26277
entropy	72	0,18485	0,25285	0,57358	0,07928	49	0,21575	0,22747	0,67736	0,13229
joint-entropy	127	0,19751	0,28904	0,46604	0,04412	55	0,24154	0,23979	0,67358	0,15316
simple	103	0,18970	0,27544	0,49245	0,04332	11	0,23185	0,21283	0,82264	0,24829
sfs	48	0,13668	0,13619	0,86226	0,13465	43	0,14143	0,13516	0,87358	0,14390
sbs	47	***0,13388***	***0,13382***	0,86604	0,13367	37	***0,14008***	***0,13278***	0,87736	0,14292

	geheime PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,16168	0,23169	0,58868	0,05207	131	0,16168	0,23169	0,58868	0,05207
anova	95	**0,15437**	**0,20176**	0,66038	0,06390	85	**0,16605**	**0,17409**	0,73774	0,08592
anova-2class	96	0,16015	0,20430	0,65849	0,06709	43	0,21381	0,18474	0,86604	0,23552
correlation	118	0,16168	0,23169	0,58868	0,05207	53	0,21201	0,21245	0,72264	0,14753
entropy	118	0,16168	0,23169	0,58868	0,05207	77	0,18163	0,21464	0,65660	0,08589
joint-entropy	131	0,16168	0,23169	0,58868	0,05207	86	0,18900	0,20900	0,68302	0,10102
simple	93	0,15280	0,20971	0,64151	0,06092	36	0,19045	0,19198	0,75849	0,14245
sfs	62	***0,12263***	***0,12262***	0,87736	0,12261	52	***0,12943***	0,12092	0,89057	0,13240
sbs	73	0,12423	0,12397	0,87547	0,12340	54	0,14036	***0,11558***	0,91887	0,15004

	Pseudonym					Pseudonym				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12517	0,18101	0,68302	0,04503	131	0,12517	0,18101	0,68302	0,04503
anova	103	**0,11840**	0,16607	0,71698	0,04913	65	**0,12117**	**0,11651**	0,85283	0,08585
anova-2class	113	0,12362	0,18204	0,68302	0,04710	69	0,13470	0,12874	0,83396	0,09144
correlation	101	0,12155	**0,16522**	0,71887	0,04931	45	0,16897	0,14450	0,89245	0,18146
entropy	115	0,12517	0,18224	0,68302	0,04750	40	0,16384	0,15942	0,84717	0,16600
joint-entropy	113	0,12276	0,16880	0,71321	0,05080	69	0,15583	0,15294	0,83208	0,13795
simple	91	0,11872	0,16422	0,72264	0,05109	53	0,12879	0,12850	0,82642	0,08342
sfs	85	0,08693	0,08601	0,91509	0,08712	65	0,09778	***0,07522***	0,95283	0,10327
sbs	82	***0,08065***	***0,08001***	0,92075	0,08077	51	***0,08926***	0,07538	0,94151	0,09227

	Symbol					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08866	0,10825	0,82830	0,04481	131	0,08866	0,10825	0,82830	0,04481
anova	85	**0,07534**	**0,07476**	0,90566	0,05519	55	**0,07599**	**0,07451**	0,91887	0,06789
anova-2class	118	0,08866	0,10825	0,82830	0,04481	89	0,09912	0,09862	0,89623	0,09347
correlation	118	0,08866	0,10825	0,82830	0,04481	71	0,09445	0,09314	0,88113	0,06742
entropy	118	0,08866	0,10825	0,82830	0,04481	84	0,10418	0,10154	0,89057	0,09365
joint-entropy	108	0,08538	0,09361	0,86415	0,05138	106	0,08539	0,09361	0,86415	0,05138
simple	62	0,07695	0,07678	0,90377	0,05733	52	0,07792	0,07636	0,91132	0,06404
sfs	102	***0,05864***	***0,05591***	0,94717	0,05900	66	***0,06533***	***0,04372***	0,98113	0,06858
sbs	104	0,05873	0,05862	0,94151	0,05875	65	0,06552	0,04383	0,98113	0,06880

	Woher					Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,10910	0,16272	0,72264	0,04808	131	0,10910	0,16272	0,72264	0,04808
anova	102	0,10787	0,15664	0,73774	0,05102	63	**0,10984**	**0,10983**	0,87358	0,09325
anova-2class	118	0,10910	0,16272	0,72264	0,04808	89	0,11550	0,13130	0,80755	0,07014
correlation	111	0,10823	0,15949	0,73208	0,05105	57	0,14788	0,14768	0,83396	0,12932
entropy	118	0,10910	0,16272	0,72264	0,04808	47	0,13137	0,11566	0,90566	0,13697
joint-entropy	108	**0,10658**	**0,14227**	0,77547	0,06002	69	0,13100	0,13082	0,86792	0,12957
simple	114	0,10910	0,16272	0,72264	0,04808	56	0,12416	0,13469	0,80566	0,07504
sfs	101	***0,07926***	0,07667	0,92642	0,07975	83	***0,09207***	0,06851	0,96038	0,09739
sbs	100	0,07961	***0,07255***	0,93585	0,08095	67	0,09379	***0,06613***	0,96792	0,10018

Tabelle C.8: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung $CR \leq 5\%$ für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,24441	0,24755	0,64151	0,13661	60	0,20905	0,20865	0,77170	0,18901
anova	60	0,19808	0,23509	0,62642	0,09659	60	**0,18294**	**0,18503**	0,75094	0,12101
anova-2class	60	**0,19356**	**0,23324**	0,63396	0,10044	60	0,19972	0,20078	0,74340	0,14496
correlation	60	0,21371	0,24684	0,62075	0,11444	60	0,20869	0,21763	0,69434	0,12961
entropy	60	0,19692	0,24262	0,61509	0,10033	60	0,21190	0,22625	0,67358	0,12609
joint-entropy	60	0,24417	0,24332	0,65660	0,14325	60	0,21592	0,21517	0,74528	0,17562
wrapper	60	0,20311	0,24015	0,59811	0,07841	60	0,18027	0,19750	0,70377	0,09877
sfs	60	***0,14851***	***0,14429***	0,82830	0,11687	60	***0,12386***	***0,12172***	0,88113	0,12456
sbs	60	0,15836	0,14608	0,82264	0,11480	60	0,13810	0,11577	0,91509	0,14663

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,15971	0,15816	0,82830	0,14463	60	0,09321	0,09033	0,91321	0,09387
anova	60	**0,12438**	**0,11967**	0,85283	0,09216	60	**0,08010**	**0,07845**	0,90755	0,06444
anova-2class	60	0,14332	0,13982	0,84340	0,12304	60	0,15259	0,12329	0,91887	0,16546
correlation	60	0,14886	0,14815	0,84717	0,14347	60	0,09719	0,09492	0,89811	0,08795
entropy	60	0,16691	0,16809	0,76981	0,10599	60	0,13585	0,11359	0,91698	0,14416
joint-entropy	60	0,18955	0,17837	0,83962	0,19637	60	0,11715	0,11462	0,88868	0,11792
simple	60	0,13438	0,14133	0,79623	0,07888	60	0,07710	0,07691	0,90377	0,05758
sfs	60	0,09797	***0,07533***	0,95283	0,10348	60	***0,06559***	***0,04387***	0,98113	0,06887
sbs	60	***0,08444***	0,07694	0,93208	0,08596	60	0,06572	0,04394	0,98113	0,06901

	Woher				
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,14227	0,14234	0,83396	0,11865
anova	60	**0,11588**	**0,11562**	0,87547	0,10671
anova-2class	60	0,15019	0,14996	0,82830	0,12823
correlation	60	0,15432	0,16644	0,76604	0,09891
entropy	60	0,13940	0,15127	0,78302	0,08556
joint-entropy	60	0,15725	0,14087	0,88302	0,16477
wrapper	60	0,12367	0,13674	0,80000	0,07348
sfs	60	0,09504	0,06939	0,96226	0,10105
sbs	60	***0,09478***	***0,06671***	0,96792	0,10134

Tabelle C.9: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung $RR \geq 80\%$ für TF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER gegebene PIN					(b) bestCRR gegebene PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,27398	0,25212	0,79623	0,30047	131	0,27398	0,25212	0,79623	0,30047
anova	114	**0,27333**	**0,24869**	0,80566	0,30305	114	**0,27333**	0,24869	0,80566	0,30305
anova-2class	117	0,27398	0,25212	0,79623	0,30047	100	0,27892	0,25125	0,81132	0,31382
correlation	118	0,27398	0,25212	0,79623	0,30047	98	0,28639	**0,24641**	0,84717	0,33999
entropy	118	0,27398	0,25212	0,79623	0,30047	118	0,27398	0,25212	0,79623	0,30047
joint-entropy	131	0,27398	0,25212	0,79623	0,30047	107	0,28858	0,24768	0,84906	0,34441
simple	111	0,27100	0,24514	0,81132	0,30160	111	***0,27100***	0,24514	0,81132	0,30160
sfs	108	***0,26532***	***0,22397***	0,86415	0,31208	86	0,27324	0,21978	0,89811	0,33766
sbs	108	***0,26532***	***0,22397***	0,86415	0,31208	81	0,29327	***0,21852***	0,96226	0,39931
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
	geheime PIN					**geheime PIN**				
raw	131	0,20620	0,20475	0,77736	0,18687	131	0,20620	0,20475	0,77736	0,18687
anova	102	**0,19725**	**0,19615**	0,80566	0,19797	88	0,21799	0,19245	0,85283	0,23774
anova-2class	100	0,20341	0,19933	0,80755	0,20620	90	0,22358	**0,19234**	0,86415	0,24884
correlation	118	0,20620	0,20475	0,77736	0,18687	118	**0,20620**	0,20475	0,77736	0,18687
entropy	118	0,20620	0,20475	0,77736	0,18687	118	0,20620	0,20475	0,77736	0,18687
joint-entropy	131	0,20620	0,20475	0,77736	0,18687	107	0,20716	0,19964	0,81321	0,21248
simple	107	0,19933	0,19831	0,79245	0,18908	107	0,19933	0,19831	0,79245	0,18908
sfs	116	***0,19448***	***0,19156***	0,81321	0,19634	92	***0,19857***	0,16362	0,89434	0,22159
sbs	117	0,20348	0,19797	0,81132	0,20726	83	0,21518	***0,15969***	0,93774	0,25711
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
	Pseudonym					**Pseudonym**				
raw	131	0,16874	0,17698	0,75849	0,11245	131	0,16874	0,17698	0,75849	0,11245
anova	94	**0,14435**	**0,14160**	0,86226	0,14546	65	0,16632	**0,13826**	0,90377	0,18030
anova-2class	89	0,15898	0,15419	0,85283	0,16121	70	0,18072	0,15147	0,89434	0,19728
correlation	106	0,16247	0,16393	0,78868	0,11655	91	0,16490	0,16344	0,80189	0,12877
entropy	118	0,16874	0,17698	0,75849	0,11245	93	0,17493	0,17527	0,79245	0,14300
joint-entropy	112	0,16446	0,16704	0,78679	0,12086	112	**0,16446**	0,16704	0,78679	0,12086
simple	79	0,15245	0,15078	0,82830	0,12986	79	0,15245	0,15078	0,82830	0,12986
sfs	108	0,14255	0,13425	0,87736	0,14586	76	***0,15216***	0,09670	0,98302	0,17642
sbs	108	***0,13950***	***0,12848***	0,88679	0,14376	65	0,15734	***0,09566***	0,99434	0,18567
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
	Symbol					**Symbol**				
raw	131	0,08866	0,10825	0,82830	0,04481	131	0,08866	0,10825	0,82830	0,04481
anova	85	**0,07534**	**0,07476**	0,90566	0,05519	55	**0,07599**	**0,07451**	0,91887	0,06789
anova-2class	118	0,08866	0,10825	0,82830	0,04481	89	0,09912	0,09862	0,89623	0,09347
correlation	118	0,08866	0,10825	0,82830	0,04481	71	0,09445	0,09314	0,88113	0,06742
entropy	118	0,08866	0,10825	0,82830	0,04481	84	0,10418	0,10154	0,89057	0,09365
joint-entropy	108	0,08538	0,09361	0,86415	0,05138	106	0,08539	0,09361	0,86415	0,05138
simple	62	0,07695	0,07678	0,90377	0,05733	52	0,07792	0,07636	0,91132	0,06404
sfs	102	***0,05864***	***0,05591***	0,94717	0,05900	66	***0,06533***	***0,04372***	0,98113	0,06858
sbs	104	0,05873	0,05862	0,94151	0,05875	65	0,06552	0,04383	0,98113	0,06880
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
	Woher					**Woher**				
raw	131	0,18363	0,16950	0,85283	0,19184	131	0,18363	0,16950	0,85283	0,19184
anova	115	0,18363	0,16950	0,85283	0,19184	98	0,21222	**0,15882**	0,93396	0,25160
anova-2class	118	0,18363	0,16950	0,85283	0,19184	91	0,22297	0,16702	0,93396	0,26800
correlation	117	0,18363	0,16950	0,85283	0,19184	113	0,18576	0,16941	0,85660	0,19543
entropy	118	0,18363	0,16950	0,85283	0,19184	118	***0,18363***	0,16950	0,85283	0,19184
joint-entropy	131	0,18363	0,16950	0,85283	0,19184	124	0,18474	0,16364	0,86981	0,19710
simple	102	0,18363	0,16950	0,85283	0,19184	91	**0,18437**	0,16121	0,87547	0,19790
sfs	117	***0,18128***	***0,16139***	0,86981	0,19260	82	0,19107	0,15269	0,90943	0,21480
sbs	118	0,18363	0,16950	0,85283	0,19184	74	0,21613	***0,14332***	0,98491	0,27155

Tabelle C.10: Biometric Hash: Ergebnisse der Merkmalsselektion bei Parametrisierung $RR \geq 80\%$ für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,33566	0,28110	0,88491	0,44710	60	0,28610	0,23656	0,87925	0,35236
anova	60	**0,32590**	**0,27048**	0,88868	0,42964	60	**0,24739**	**0,20742**	0,87170	0,28654
anova-2class	60	0,36277	0,30619	0,90000	0,51237	60	0,29613	0,24151	0,89245	0,37547
correlation	60	0,33243	0,28024	0,87547	0,43596	60	0,25940	0,23520	0,81509	0,28549
entropy	60	0,32881	0,28489	0,84340	0,41317	60	0,27316	0,24621	0,81321	0,30562
joint-entropy	60	0,36411	0,30323	0,92075	0,52721	60	0,29399	0,24619	0,86981	0,36219
simple	60	0,29657	0,25718	0,83962	0,35399	60	0,22499	0,20742	0,82453	0,23937
sfs	60	***0,27811***	0,22221	0,90377	0,34819	60	***0,20712***	0,16685	0,90189	0,23560
sbs	60	0,29595	***0,22112***	0,96226	0,40450	60	0,22441	***0,16277***	0,94906	0,27460

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,20575	0,17915	0,86604	0,22435	60	0,09321	0,09033	0,91321	0,09387
anova	60	**0,17191**	**0,14193**	0,90377	0,18763	60	**0,08010**	**0,07845**	0,90755	0,06444
anova-2class	60	0,21665	0,17446	0,90000	0,24891	60	0,15259	0,12329	0,91887	0,16546
correlation	60	0,22107	0,17839	0,89811	0,25490	60	0,09719	0,09492	0,89811	0,08795
entropy	60	0,19922	0,18817	0,83019	0,20653	60	0,13585	0,11359	0,91698	0,14416
joint-entropy	60	0,25944	0,21194	0,88679	0,31067	60	0,11715	0,11462	0,88868	0,11792
simple	60	0,15858	0,15548	0,84906	0,16001	60	0,07710	0,07691	0,90377	0,05758
sfs	60	***0,15364***	0,09771	0,98302	0,17845	60	***0,06559***	***0,04387***	0,98113	0,06887
sbs	60	0,15820	***0,09626***	0,99434	0,18687	60	0,06572	0,04394	0,98113	0,06901

	Woher			
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,26052	0,21526	0,87925	0,30976
anova	60	0,25962	**0,18636**	0,96604	0,33875
anova-2class	60	0,30062	0,22783	0,95472	0,41038
correlation	60	0,24285	0,20330	0,87358	0,28019
entropy	60	**0,23952**	0,20207	0,86981	0,27395
joint-entropy	60	0,30785	0,23338	0,96038	0,42714
simple	60	0,20477	0,17288	0,88113	0,22689
sfs	60	***0,19426***	0,15492	0,90943	0,21927
sbs	60	0,21982	***0,14630***	0,98491	0,27750

Tabelle C.11: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,20630	0,50000	0,00000	0,00000	131	*0,20630*	0,50000	0,00000	0,00000
anova	102	0,20233	0,50000	0,00000	0,00000	1	0,45255	0,42264	0,26226	0,10755
anova-2class	112	0,20461	0,50000	0,00000	0,00000	1	**0,42214**	**0,39526**	0,77736	0,56789
correlation	108	0,19971	0,50000	0,00000	0,00000	1	0,46967	0,44614	0,16604	0,05831
entropy	113	**0,19893**	0,50000	0,00000	0,00000	1	**0,42214**	**0,39526**	0,77736	0,56789
joint-entropy	131	0,20630	0,50000	0,00000	0,00000	1	0,48353	0,46987	0,11509	0,05483
simple	93	***0,19463***	0,50000	0,00000	0,00000	1	0,42665	***0,38841***	0,35094	0,12776
sfs	118	0,20630	0,50000	0,00000	0,00000	1	**0,42214**	0,39526	0,77736	0,56789
sbs	113	0,20135	0,50000	0,00000	0,00000	1	0,42665	***0,38841***	0,35094	0,12776
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,16013	0,50000	0,00000	0,00000	131	*0,16013*	0,50000	0,00000	0,00000
anova	102	0,15687	0,50000	0,00000	0,00000	1	0,48122	0,46548	0,11509	0,04605
anova-2class	114	0,15693	0,50000	0,00000	0,00000	1	**0,28776**	***0,26499***	0,78868	0,31865
correlation	118	0,16013	0,50000	0,00000	0,00000	1	0,44591	0,41303	0,28302	0,10907
entropy	109	**0,15557**	0,50000	0,00000	0,00000	1	**0,28776**	***0,26499***	0,78868	0,31865
joint-entropy	131	0,16013	0,50000	0,00000	0,00000	1	0,48581	0,47360	0,09623	0,04343
simple	66	***0,14249***	0,50000	0,00000	0,00000	1	**0,28776**	***0,26499***	0,78868	0,31865
sfs	97	0,15452	0,50000	0,00000	0,00000	1	**0,28776**	***0,26499***	0,78868	0,31865
sbs	113	0,15788	0,50000	0,00000	0,00000	1	0,44591	0,41303	0,28302	0,10907
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,15539	0,50000	0,00000	0,00000	131	*0,15539*	0,50000	0,00000	0,00000
anova	102	0,14456	0,50000	0,00000	0,00000	1	0,45855	0,42700	0,19245	0,04644
anova-2class	117	0,15536	0,50000	0,00000	0,00000	1	0,45171	0,43687	0,40943	0,28316
correlation	101	**0,14226**	0,50000	0,00000	0,00000	1	0,47734	0,45943	0,14528	0,06415
entropy	108	0,14629	0,50000	0,00000	0,00000	1	**0,33214**	***0,27400***	0,55283	0,10083
joint-entropy	100	0,15426	0,50000	0,00000	0,00000	1	0,49073	0,48182	0,03774	0,00138
simple	82	***0,14052***	0,50000	0,00000	0,00000	1	**0,33214**	***0,27400***	0,55283	0,10083
sfs	95	0,14750	0,50000	0,00000	0,00000	1	**0,33214**	***0,27400***	0,55283	0,10083
sbs	113	0,15391	0,50000	0,00000	0,00000	1	0,43097	0,39060	0,31698	0,09819
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,16158	0,50000	0,00000	0,00000	131	*0,16158*	0,50000	0,00000	0,00000
anova	96	0,15373	0,50000	0,00000	0,00000	1	**0,25583**	***0,18427***	0,96226	0,33081
anova-2class	74	***0,11999***	0,50000	0,00000	0,00000	1	0,40442	0,39610	0,56038	0,35258
correlation	85	0,12635	0,50000	0,00000	0,00000	1	0,47789	0,46150	0,16792	0,09093
entropy	118	0,16158	0,50000	0,00000	0,00000	1	**0,25583**	***0,18427***	0,96226	0,33081
joint-entropy	127	0,16053	0,50000	0,00000	0,00000	1	0,48070	0,46364	0,09434	0,02163
simple	97	**0,15480**	0,50000	0,00000	0,00000	1	**0,25583**	***0,18427***	0,96226	0,33081
sfs	116	0,16019	0,50000	0,00000	0,00000	1	**0,25583**	***0,18427***	0,96226	0,33081
sbs	113	0,16073	0,50000	0,00000	0,00000	1	0,44462	0,40882	0,26792	0,08556
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12147	0,50000	0,00000	0,00000	131	*0,12147*	0,50000	0,00000	0,00000
anova	105	**0,10342**	0,50000	0,00000	0,00000	1	0,46568	0,43757	0,15283	0,02798
anova-2class	118	0,12147	0,50000	0,00000	0,00000	1	0,48408	0,47723	0,30755	0,26201
correlation	117	0,12116	0,50000	0,00000	0,00000	1	0,45623	0,42378	0,20566	0,05323
entropy	118	0,12147	0,50000	0,00000	0,00000	1	**0,39526**	***0,34401***	0,41132	0,09935
joint-entropy	131	0,12147	0,50000	0,00000	0,00000	1	0,49069	0,48175	0,03774	0,00123
simple	66	***0,09924***	0,50000	0,00000	0,00000	1	**0,39526**	***0,34401***	0,41132	0,09935
sfs	92	0,11905	0,50000	0,00000	0,00000	1	**0,39526**	***0,34401***	0,41132	0,09935
sbs	103	0,11857	0,50000	0,00000	0,00000	1	0,48891	0,48104	0,16415	0,12623

Tabelle C.12: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung *Null* für TF für Szenario fix60

selection	gegebene PIN n	EER	CRR	RR	CR	geheime PIN n	EER	CRR	RR	CR
raw	60	0,27675	0,50000	0,00000	0,00000	60	0,21920	0,50000	0,00000	0,00000
anova	60	0,23480	0,50000	0,00000	0,00000	60	0,19264	0,50000	0,00000	0,00000
anova-2class	60	**0,23345**	0,50000	0,00000	0,00000	60	**0,17881**	0,50000	0,00000	0,00000
correlation	60	0,25411	0,50000	0,00000	0,00000	60	0,22253	0,50000	0,00000	0,00000
entropy	60	0,24074	0,50000	0,00000	0,00000	60	0,19460	0,50000	0,00000	0,00000
joint-entropy	60	0,27105	0,50000	0,00000	0,00000	60	0,22326	0,50000	0,00000	0,00000
simple	60	***0,20495***	0,50000	0,00000	0,00000	60	***0,14611***	0,50000	0,00000	0,00000
sfs	60	0,23580	0,50000	0,00000	0,00000	60	0,18948	0,50000	0,00000	0,00000
sbs	60	0,24428	0,50000	0,00000	0,00000	60	0,22203	0,50000	0,00000	0,00000

selection	Pseudonym n	EER	CRR	RR	CR	Symbol n	EER	CRR	RR	CR
raw	60	0,17571	0,50000	0,00000	0,00000	60	0,17055	0,50000	0,00000	0,00000
anova	60	**0,17170**	0,50000	0,00000	0,00000	60	0,25824	0,50000	0,00000	0,00000
anova-2class	60	0,17486	0,50000	0,00000	0,00000	60	**0,12210**	0,50000	0,00000	0,00000
correlation	60	0,17595	0,50000	0,00000	0,00000	60	0,16107	0,50000	0,00000	0,00000
entropy	60	0,18690	0,50000	0,00000	0,00000	60	0,22351	0,50000	0,00000	0,00000
joint-entropy	60	0,20130	0,50000	0,00000	0,00000	60	0,27975	0,50000	0,00000	0,00000
simple	60	***0,14702***	0,50000	0,00000	0,00000	60	***0,17504***	0,50000	0,00000	0,00000
sfs	60	0,17838	0,50000	0,00000	0,00000	60	0,20955	0,50000	0,00000	0,00000
sbs	60	0,22111	0,50000	0,00000	0,00000	60	0,20713	0,50000	0,00000	0,00000

selection	Woher n	EER	CRR	RR	CR
raw	60	0,17702	0,50000	0,00000	0,00000
anova	60	**0,12909**	0,50000	0,00000	0,00000
anova-2class	60	0,15753	0,50000	0,00000	0,00000
correlation	60	0,16189	0,50000	0,00000	0,00000
entropy	60	0,16025	0,50000	0,00000	0,00000
joint-entropy	60	0,20219	0,50000	0,00000	0,00000
simple	60	***0,09958***	0,50000	0,00000	0,00000
sfs	60	0,14387	0,50000	0,00000	0,00000
sbs	60	0,18669	0,50000	0,00000	0,00000

Tabelle C.13: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,16211	0,50000	0,00000	0,00000	131	*0,16211*	0,50000	0,00000	0,00000
anova	116	0,15996	0,50000	0,00000	0,00000	1	0,34135	**0,32139**	0,74151	0,38429
anova-2class	41	***0,15312***	***0,49528***	0,00943	0,00000	6	0,33016	0,32903	0,54717	0,20522
correlation	118	0,16211	0,50000	0,00000	0,00000	2	**0,32633**	0,33387	0,57736	0,24510
entropy	65	0,15592	0,50000	0,00000	0,00000	2	0,41604	0,37440	0,87358	0,62239
joint-entropy	131	0,16211	0,50000	0,00000	0,00000	2	0,36197	0,35981	0,40377	0,12340
simple	45	0,15989	**0,49906**	0,00189	0,00000	2	0,26976	0,24243	0,67547	0,16034
sfs	67	0,15508	**0,49906**	0,00189	0,00000	3	**0,26736**	***0,24205***	0,63774	0,12184
sbs	105	**0,15474**	0,50000	0,00000	0,00000	4	0,32971	0,31404	0,57170	0,19978
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,10927	0,50000	0,00000	0,00000	131	*0,10927*	0,50000	0,00000	0,00000
anova	109	0,10587	0,50000	0,00000	0,00000	2	0,35898	0,35731	0,35472	0,06934
anova-2class	68	***0,09452***	0,50000	0,00000	0,00000	1	0,33302	0,25296	0,98679	0,49271
correlation	118	0,10927	0,50000	0,00000	0,00000	1	0,30827	0,29238	0,74906	0,33382
entropy	115	0,10640	0,50000	0,00000	0,00000	2	**0,26835**	**0,22580**	0,86604	0,31763
joint-entropy	101	0,10388	0,50000	0,00000	0,00000	2	0,33250	0,36785	0,33962	0,07533
simple	84	0,10565	0,50000	0,00000	0,00000	2	0,23945	0,21733	0,69057	0,12522
sfs	87	**0,09945**	0,50000	0,00000	0,00000	3	**0,16674**	***0,16566***	0,83208	0,16339
sbs	105	0,10524	0,50000	0,00000	0,00000	2	0,29596	0,28282	0,56226	0,12790
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08757	0,47170	0,05660	0,00000	131	*0,08757*	0,47170	0,05660	0,00000
anova	118	0,08757	0,47170	0,05660	0,00000	5	0,17106	**0,15071**	0,79623	0,09764
anova-2class	33	***0,08424***	0,34071	0,32075	0,00218	6	0,21186	0,19537	0,71698	0,10773
correlation	95	0,08512	0,46228	0,07547	0,00004	7	0,17693	0,23748	0,57170	0,04666
entropy	118	0,08757	0,47170	0,05660	0,00000	3	0,19172	0,18552	0,82453	0,19557
joint-entropy	126	0,08679	0,47170	0,05660	0,00000	12	**0,15695**	0,19866	0,65283	0,05015
simple	118	0,08757	0,47170	0,05660	0,00000	2	0,16261	0,14673	0,82264	0,11611
sfs	60	**0,08432**	0,36515	0,26981	0,00011	5	0,13566	0,13409	0,78113	0,04931
sbs	64	0,08577	0,34067	0,31887	0,00022	9	**0,10743**	***0,10250***	0,90377	0,10878
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,07259	0,26392	0,47547	0,00330	131	0,07259	0,26392	0,47547	0,00330
anova	98	0,06132	0,23456	0,53585	0,00497	24	**0,07377**	**0,07132**	0,91698	0,05962
anova-2class	63	0,06212	**0,18790**	0,63208	0,00787	7	0,11395	0,11292	0,88491	0,11074
correlation	115	0,06857	0,26116	0,48113	0,00345	23	0,12350	0,11451	0,84906	0,07808
entropy	118	0,07259	0,26392	0,47547	0,00330	8	0,12035	0,11348	0,86415	0,09111
joint-entropy	107	**0,05755**	0,25490	0,49434	0,00414	16	0,14228	0,12779	0,89245	0,14804
simple	63	0,05430	0,20171	0,60189	0,00530	15	0,07907	0,07640	0,91698	0,06978
sfs	76	0,04845	0,15492	0,69811	0,00795	22	0,05273	0,06054	0,90189	0,02297
sbs	26	***0,03286***	***0,03142***	0,96604	0,02888	22	***0,04078***	***0,02848***	0,98491	0,04187
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,06817	0,50000	0,00000	0,00000	131	*0,06817*	0,50000	0,00000	0,00000
anova	118	0,06817	0,50000	0,00000	0,00000	1	**0,22336**	**0,20555**	0,76226	0,17337
anova-2class	80	0,06098	**0,49906**	0,00189	0,00000	5	0,23786	0,22941	0,65094	0,10976
correlation	118	0,06817	0,50000	0,00000	0,00000	1	0,31189	0,28336	0,79245	0,35918
entropy	70	***0,04823***	0,50000	0,00000	0,00000	1	0,25254	0,21578	0,85849	0,29006
joint-entropy	131	0,06817	0,50000	0,00000	0,00000	2	0,26954	0,28151	0,51509	0,07812
simple	61	**0,06321**	0,49906	0,00189	0,00000	1	0,21356	0,18044	0,87736	0,23824
sfs	69	0,06602	***0,49340***	0,01321	0,00000	2	**0,18757**	***0,16736***	0,77925	0,11397
sbs	103	0,06651	0,49811	0,00377	0,00000	2	0,24033	0,22074	0,81698	0,25845

Tabelle C.14: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minEER für TF für Szenario fix60

selection	gegebene PIN n	EER	CRR	RR	CR	geheime PIN n	EER	CRR	RR	CR
raw	60	0,20977	0,50000	0,00000	0,00000	60	0,15475	0,50000	0,00000	0,00000
anova	60	0,20325	0,50000	0,00000	0,00000	60	0,14728	0,50000	0,00000	0,00000
anova-2class	60	***0,16522***	0,50000	0,00000	0,00000	60	***0,10782***	0,50000	0,00000	0,00000
correlation	60	0,21502	0,50000	0,00000	0,00000	60	0,17324	0,50000	0,00000	0,00000
entropy	60	0,16584	0,50000	0,00000	0,00000	60	0,12116	0,50000	0,00000	0,00000
joint-entropy	60	0,19374	0,50000	0,00000	0,00000	60	0,14111	0,50000	0,00000	0,00000
simple	60	**0,16533**	0,50000	0,00000	0,00000	60	0,12227	***0,49906***	0,00189	0,00000
sfs	60	0,16749	0,49906	0,00189	0,00000	60	**0,11727**	0,50000	0,00000	0,00000
sbs	60	0,17423	**0,49623**	0,00755	0,00000	60	0,13614	0,50000	0,00000	0,00000

selection	Pseudonym n	EER	CRR	RR	CR	Symbol n	EER	CRR	RR	CR
raw	60	0,09732	0,41894	0,16226	0,00015	60	0,07751	0,17462	0,66981	0,01905
anova	60	0,10315	**0,42077**	0,15849	0,00004	60	0,06761	**0,13545**	0,74340	0,01430
anova-2class	60	0,09403	0,42266	0,15472	0,00004	60	**0,06356**	0,17861	0,65094	0,00816
correlation	60	0,09703	0,43022	0,13962	0,00007	60	0,07685	0,16646	0,67736	0,01027
entropy	60	**0,08889**	0,43115	0,13774	0,00004	60	0,08472	0,23197	0,54340	0,00733
joint-entropy	60	0,09680	0,42549	0,14906	0,00004	60	0,07808	0,18168	0,65660	0,01996
simple	60	0,09990	0,43304	0,13396	0,00004	60	0,05781	0,19118	0,62453	0,00689
sfs	60	***0,08432***	0,36515	0,26981	0,00011	60	***0,05313***	0,11593	0,77925	0,01110
sbs	60	0,09143	***0,32377***	0,35283	0,00036	60	0,06918	***0,06475***	0,88868	0,01818

selection	Woher n	EER	CRR	RR	CR
raw	60	0,09297	0,50000	0,00000	0,00000
anova	60	0,08291	**0,49811**	0,00377	0,00000
anova-2class	60	0,06689	0,49906	0,00189	0,00000
correlation	60	0,09588	0,49906	0,00189	0,00000
entropy	60	***0,05074***	0,50000	0,00000	0,00000
joint-entropy	60	0,09977	**0,49811**	0,00377	0,00000
simple	60	**0,06453**	0,49906	0,00189	0,00000
sfs	60	0,06785	0,49245	0,01509	0,00000
sbs	60	0,07926	***0,49151***	0,01698	0,00000

Tabelle C.15: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minCRR für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,22595	0,25789	0,60566	0,12144	131	0,22595	0,25789	0,60566	0,12144
anova	83	**0,20967**	**0,20889**	0,79245	0,21023	79	**0,21254**	**0,20314**	0,81321	0,21948
anova-2class	118	0,22595	0,25789	0,60566	0,12144	50	0,25252	0,23888	0,78868	0,26644
correlation	118	0,22595	0,25789	0,60566	0,12144	50	0,26656	0,23997	0,81698	0,29692
entropy	118	0,22595	0,25789	0,60566	0,12144	92	0,23599	0,25356	0,63585	0,14296
joint-entropy	131	0,22595	0,25789	0,60566	0,12144	65	0,24864	0,23552	0,79057	0,26161
simple	42	0,22267	0,22251	0,77547	0,22050	26	0,24011	0,22123	0,81509	0,25755
sfs	90	0,20481	0,20198	0,76604	0,16999	90	0,20481	0,20198	0,76604	0,16999
sbs	103	***0,18688***	***0,18612***	0,81509	0,18734	78	***0,20008***	***0,16538***	0,89245	0,22322
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,19485	0,19726	0,70943	0,10396	131	0,19485	0,19726	0,70943	0,10396
anova	89	**0,16332**	**0,16214**	0,83962	0,16390	86	**0,17525**	**0,15454**	0,87736	0,18643
anova-2class	100	0,19234	0,19176	0,72642	0,10994	96	0,19244	0,19073	0,73396	0,11542
correlation	76	0,17672	0,17464	0,81887	0,16814	76	0,17672	0,17464	0,81887	0,16814
entropy	111	0,19396	0,19541	0,71509	0,10591	110	0,19407	0,19541	0,71509	0,10591
joint-entropy	82	0,18202	0,17997	0,81321	0,17315	66	0,19346	0,16941	0,86981	0,20864
simple	31	0,17524	0,17513	0,82453	0,17478	53	0,17759	0,16999	0,79623	0,13621
sfs	93	0,14855	0,14487	0,86038	0,15011	75	***0,15085***	0,14390	0,86604	0,15385
sbs	109	***0,14830***	***0,14394***	0,86226	0,15015	86	0,17450	***0,12280***	0,95660	0,20221
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,18061	0,20708	0,67547	0,08962	131	0,18061	0,20708	0,67547	0,08962
anova	94	**0,17060**	**0,16985**	0,79811	0,13781	94	**0,17060**	**0,16985**	0,79811	0,13781
anova-2class	118	0,18061	0,20708	0,67547	0,08962	74	0,19014	0,18884	0,75094	0,12863
correlation	109	0,17705	0,19403	0,70566	0,09372	77	0,18178	0,17910	0,79434	0,15254
entropy	118	0,18061	0,20708	0,67547	0,08962	27	0,20087	0,19623	0,78302	0,17547
joint-entropy	113	0,17728	0,19953	0,69245	0,09151	88	0,18408	0,18320	0,76792	0,13433
simple	41	0,17171	0,17025	0,77925	0,11974	32	0,17260	0,16959	0,78679	0,12598
sfs	51	0,15321	0,15165	0,84151	0,14481	28	0,15593	0,15005	0,85849	0,15860
sbs	98	***0,13571***	***0,13338***	0,86981	0,13657	73	***0,13769***	***0,13137***	0,87736	0,14009
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,10021	0,17393	0,69057	0,03843	131	0,10021	0,17393	0,69057	0,03843
anova	114	**0,09911**	**0,17208**	0,69434	0,03850	68	0,14023	0,11471	0,92075	0,15018
anova-2class	118	0,10021	0,17393	0,69057	0,03843	7	**0,11486**	**0,11464**	0,88491	0,11419
correlation	117	0,10021	0,17393	0,69057	0,03843	61	0,12498	0,12155	0,88302	0,12612
entropy	118	0,10021	0,17393	0,69057	0,03843	8	0,12169	0,11484	0,86415	0,09383
joint-entropy	105	0,09996	0,17438	0,69057	0,03933	92	0,12974	0,13585	0,79811	0,06981
simple	94	0,09990	0,17221	0,69434	0,03875	11	0,14395	0,10762	0,94340	0,15864
sfs	86	0,07426	0,07299	0,92453	0,07050	81	0,07427	0,07299	0,92453	0,07050
sbs	76	***0,07309***	***0,07186***	0,92642	0,07014	55	***0,07368***	***0,06669***	0,94151	0,07489
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,15270	0,19282	0,68868	0,07431	131	0,15270	0,19282	0,68868	0,07431
anova	66	**0,14566**	**0,14394**	0,85849	0,14637	65	**0,14622**	**0,14349**	0,86038	0,14735
anova-2class	118	0,15270	0,19282	0,68868	0,07431	91	0,15790	0,17081	0,75660	0,09822
correlation	118	0,15270	0,19282	0,68868	0,07431	59	0,18523	0,16907	0,85660	0,19474
entropy	118	0,15270	0,19282	0,68868	0,07431	57	0,17902	0,18235	0,75660	0,12130
joint-entropy	124	0,15042	0,19213	0,69245	0,07671	42	0,21406	0,16226	0,92830	0,25283
simple	87	0,14702	0,17905	0,72075	0,07885	24	0,16424	0,16310	0,82264	0,14884
sfs	100	***0,12109***	***0,11851***	0,88491	0,12192	51	***0,12962***	0,11782	0,89811	0,13374
sbs	108	0,12177	0,12159	0,87736	0,12054	74	0,13872	***0,10586***	0,93962	0,15134

Tabelle C.16: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung minCRR für TF für Szenario fix60

selection	gegebene PIN n	EER	CRR	RR	CR	geheime PIN n	EER	CRR	RR	CR
raw	60	0,24212	0,24126	0,74906	0,23157	60	0,19206	0,17926	0,84151	0,20004
anova	60	**0,22801**	**0,21283**	0,81509	0,24075	60	0,18997	**0,15987**	0,88868	0,20842
anova-2class	60	0,25917	0,25836	0,71132	0,22805	60	0,21573	0,21540	0,78302	0,21382
correlation	60	0,24911	0,24904	0,72830	0,22638	60	**0,18741**	0,17919	0,83396	0,19234
entropy	60	0,26065	0,26595	0,65094	0,18284	60	0,22139	0,21515	0,73208	0,16237
joint-entropy	60	0,25486	0,23930	0,79245	0,27104	60	0,19808	0,17181	0,87170	0,21531
simple	60	0,23481	0,24842	0,65849	0,15533	60	0,17837	0,17095	0,79245	0,13436
sfs	60	0,21088	0,20749	0,76792	0,18291	60	***0,15296***	0,14517	0,86604	0,15639
sbs	60	***0,20262***	***0,16716***	0,89245	0,22678	60	0,17781	***0,12440***	0,95849	0,20729

selection	Pseudonym n	EER	CRR	RR	CR	Symbol n	EER	CRR	RR	CR
raw	60	0,19280	**0,19080**	0,79811	0,17972	60	0,16741	0,14953	0,87736	0,17642
anova	60	0,21755	0,18465	0,87358	0,24289	60	0,16595	0,12291	0,94151	0,18734
anova-2class	60	0,20170	0,19768	0,78113	0,17649	60	**0,12204**	0,14102	0,76792	0,04996
correlation	60	**0,19278**	0,19116	0,79811	0,18044	60	0,12523	**0,12170**	0,88302	0,12642
entropy	60	0,20722	0,20980	0,70189	0,12148	60	0,12389	0,17286	0,70189	0,04761
joint-entropy	60	0,24009	0,20831	0,85283	0,26945	60	0,15922	0,14975	0,86415	0,16364
simple	60	0,17665	0,18166	0,74528	0,10860	60	0,10603	0,17516	0,69434	0,04467
sfs	60	0,15986	0,15724	0,82830	0,14278	60	0,07435	0,07311	0,92453	0,07076
sbs	60	***0,13824***	***0,13169***	0,87736	0,14075	60	***0,07368***	***0,06669***	0,94151	0,07489

selection	Woher n	EER	CRR	RR	CR
raw	60	0,18279	0,18311	0,77547	0,14169
anova	60	**0,16903**	**0,15506**	0,86604	0,17616
anova-2class	60	0,19571	0,19459	0,78302	0,17221
correlation	60	0,18568	0,18530	0,78491	0,15552
entropy	60	0,17875	0,19748	0,70189	0,09684
joint-entropy	60	0,20262	0,17983	0,85849	0,21814
simple	60	0,15486	0,17930	0,73019	0,08879
sfs	60	***0,12893***	0,12063	0,89057	0,13182
sbs	60	0,14111	***0,10659***	0,94151	0,15468

Tabelle C.17: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung $CR \leq 5\%$ für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,18792	0,29915	0,44717	0,04546	131	0,18792	0,29915	0,44717	0,04546
anova	83	0,18537	**0,23039**	0,63019	0,09097	24	**0,20701**	**0,20608**	0,78113	0,19329
anova-2class	80	**0,18359**	0,26524	0,54717	0,07765	49	0,22193	0,22854	0,68679	0,14387
correlation	118	0,18792	0,29915	0,44717	0,04546	36	0,21237	0,21235	0,78679	0,21150
entropy	72	0,18691	0,27977	0,51321	0,07275	30	0,26835	0,26952	0,64528	0,18433
joint-entropy	131	0,18792	0,29915	0,44717	0,04546	55	0,22982	0,23529	0,67547	0,14604
simple	112	0,18252	0,29926	0,44717	0,04568	21	0,21124	0,21079	0,78491	0,20649
sfs	61	0,17571	0,18062	0,76604	0,12729	38	0,18000	0,17850	0,80000	0,15700
sbs	68	***0,14049***	***0,13986***	0,85849	0,13821	45	***0,17313***	***0,13452***	0,92453	0,19358
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,15617	0,20905	0,63396	0,05207	131	0,15617	0,20905	0,63396	0,05207
anova	49	**0,13760**	**0,13752**	0,86226	0,13730	32	**0,15781**	**0,12589**	0,92075	0,17253
anova-2class	118	0,15617	0,20905	0,63396	0,05207	44	0,19706	0,18037	0,84717	0,20791
correlation	117	0,15617	0,20905	0,63396	0,05207	56	0,17825	0,17130	0,78491	0,12750
entropy	118	0,15617	0,20905	0,63396	0,05207	107	0,16142	0,20791	0,63962	0,05544
joint-entropy	131	0,15617	0,20905	0,63396	0,05207	65	0,16248	0,15454	0,81887	0,12794
simple	110	0,15397	0,20559	0,64151	0,05269	16	0,18584	0,15490	0,89434	0,20414
sfs	88	0,11519	0,11515	0,88491	0,11520	40	0,13078	0,11290	0,91132	0,13712
sbs	103	***0,10595***	***0,10582***	0,89434	0,10599	72	***0,11740***	***0,08614***	0,95472	0,12700
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,13610	0,21179	0,62642	0,05000	131	0,13610	0,21179	0,62642	0,05000
anova	116	0,13551	0,20530	0,63962	0,05022	48	**0,16195**	**0,14153**	0,88868	0,17173
anova-2class	118	0,13610	0,21179	0,62642	0,05000	70	0,16905	0,17349	0,74906	0,09605
correlation	111	**0,13505**	**0,20071**	0,64906	0,05047	58	0,16773	0,16709	0,79434	0,12852
entropy	118	0,13610	0,21179	0,62642	0,05000	27	0,19081	0,18456	0,76226	0,13139
joint-entropy	113	0,13564	0,20428	0,64340	0,05196	69	0,16785	0,16504	0,80943	0,13951
simple	97	0,13352	0,20089	0,64906	0,05084	41	0,14453	0,15510	0,76604	0,07623
sfs	76	0,13511	0,13422	0,83962	0,10806	55	***0,13570***	0,13380	0,84340	0,11099
sbs	93	***0,11052***	***0,11007***	0,88868	0,10882	68	0,13719	***0,09775***	0,95660	0,15210
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12516	0,17311	0,70566	0,05189	131	0,12516	0,17311	0,70566	0,05189
anova	114	0,12382	0,17219	0,70755	0,05192	74	0,14221	0,12854	0,89057	0,14764
anova-2class	7	**0,11486**	**0,11464**	0,88491	0,11419	7	**0,11486**	**0,11464**	0,88491	0,11419
correlation	117	0,12516	0,17311	0,70566	0,05189	79	0,13596	0,13418	0,86038	0,12874
entropy	8	0,12169	0,11484	0,86415	0,09383	8	0,12169	0,11484	0,86415	0,09383
joint-entropy	102	0,12412	0,17357	0,70566	0,05279	92	0,14299	0,13823	0,81698	0,09343
simple	14	0,11571	0,11324	0,87925	0,10573	8	0,15114	0,10996	0,94906	0,16898
sfs	92	0,08963	***0,08665***	0,91698	0,09028	71	***0,09163***	0,08521	0,92264	0,09307
sbs	87	***0,08678***	0,08674	0,91321	0,08668	42	0,09837	***0,08061***	0,94151	0,10272
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,12367	0,20468	0,63962	0,04898	131	0,12367	0,20468	0,63962	0,04898
anova	107	**0,12239**	**0,17910**	0,69623	0,05443	61	**0,12803**	**0,12716**	0,85094	0,10526
anova-2class	118	0,12367	0,20468	0,63962	0,04898	40	0,17818	0,16899	0,84528	0,18327
correlation	118	0,12367	0,20468	0,63962	0,04898	52	0,15812	0,15776	0,83774	0,15327
entropy	118	0,12367	0,20468	0,63962	0,04898	47	0,15736	0,17504	0,74151	0,09158
joint-entropy	124	0,12319	0,20201	0,64717	0,05120	42	0,17485	0,14089	0,91132	0,19311
simple	87	0,12232	0,18427	0,68679	0,05533	34	0,13838	0,14055	0,81698	0,09808
sfs	96	0,11153	0,11368	0,85849	0,08585	44	***0,11234***	0,11081	0,88113	0,10276
sbs	100	***0,10080***	***0,10054***	0,89811	0,09920	62	0,12756	***0,08599***	0,96981	0,14180

Tabelle C.18: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung $CR \leq 5\%$ für TF für Szenario fix60

selection	gegebene PIN n	EER	CRR	RR	CR	geheime PIN n	EER	CRR	RR	CR
raw	60	0,22183	0,23752	0,64906	0,12409	60	0,16999	0,16094	0,80566	0,12754
anova	60	**0,19471**	**0,22505**	0,66981	0,11992	60	**0,14479**	**0,14066**	0,84717	0,12848
anova-2class	60	0,21855	0,26034	0,59057	0,11125	60	0,21360	0,21406	0,71509	0,14321
correlation	60	0,20404	0,25488	0,59811	0,10787	60	0,17967	0,17271	0,77925	0,12467
entropy	60	0,20374	0,27816	0,53208	0,08839	60	0,21032	0,22237	0,66038	0,10512
joint-entropy	60	0,22828	0,23843	0,66226	0,13911	60	0,16386	0,15604	0,81887	0,13095
simple	60	0,18597	0,24136	0,60755	0,09028	60	0,16431	0,17319	0,73396	0,08033
sfs	60	0,17625	0,18139	0,76604	0,12881	60	0,12365	0,11511	0,89623	0,12645
sbs	60	***0,15665***	***0,13743***	0,89057	0,16542	60	***0,11988***	***0,08766***	0,95472	0,13004

selection	Pseudonym n	EER	CRR	RR	CR	Symbol n	EER	CRR	RR	CR
raw	60	0,17088	0,17054	0,77925	0,12032	60	0,20401	0,16533	0,90000	0,23066
anova	60	**0,15086**	**0,14966**	0,84151	0,14082	60	0,21347	0,15288	0,95283	0,25860
anova-2class	60	0,18120	0,18193	0,75094	0,11480	60	**0,13145**	0,14490	0,76792	0,05773
correlation	60	0,17404	0,17741	0,76604	0,12086	60	0,15885	**0,13804**	0,89245	0,16854
entropy	60	0,17845	0,20528	0,66792	0,07848	60	0,13322	0,17288	0,70943	0,05519
joint-entropy	60	0,17864	0,17812	0,81887	0,17511	60	0,19228	0,16216	0,88679	0,21110
simple	60	0,14323	0,17163	0,72075	0,06401	60	0,12845	0,17386	0,70755	0,05526
sfs	60	***0,13639***	0,13507	0,83962	0,10976	60	***0,09170***	0,08525	0,92264	0,09314
sbs	60	0,13770	***0,09808***	0,95660	0,15276	60	0,09837	***0,08061***	0,94151	0,10272

selection	Woher n	EER	CRR	RR	CR
raw	60	0,16217	0,17271	0,75660	0,10203
anova	60	**0,13571**	**0,13560**	0,85094	0,12213
anova-2class	60	0,17948	0,18837	0,74528	0,12202
correlation	60	0,16959	0,17881	0,75472	0,11234
entropy	60	0,15258	0,19559	0,67547	0,06666
joint-entropy	60	0,16362	0,16274	0,83208	0,15755
simple	60	0,13093	0,17611	0,71132	0,06353
sfs	60	***0,11455***	0,11323	0,87358	0,10004
sbs	60	0,12782	***0,08616***	0,96981	0,14213

Tabelle C.19: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung $RR \geq 70\%$ für EF für Szenarien (a) bestEER, (b) bestCRR

	(a) bestEER					(b) bestCRR				
	gegebene PIN					**gegebene PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,32181	0,30517	0,74151	0,35185	131	*0,32181*	0,30517	0,74151	0,35185
anova	114	***0,31765***	***0,29528***	0,76604	0,35660	107	0,33373	**0,29517**	0,82075	0,41110
anova-2class	118	0,32181	0,30517	0,74151	0,35185	100	0,32241	0,30368	0,74906	0,35642
correlation	116	0,32181	0,30517	0,74151	0,35185	116	***0,32181***	0,30517	0,74151	0,35185
entropy	118	0,32181	0,30517	0,74151	0,35185	118	***0,32181***	0,30517	0,74151	0,35185
joint-entropy	131	0,32181	0,30517	0,74151	0,35185	107	0,32372	0,30475	0,74906	0,35856
simple	80	0,31828	0,29731	0,76038	0,35501	53	**0,32019**	0,29688	0,76792	0,36168
sfs	118	0,32181	0,30517	0,74151	0,35185	33	0,33306	0,29078	0,83585	0,41742
sbs	116	**0,31795**	**0,29704**	0,76038	0,35446	65	0,32904	***0,28511***	0,84340	0,41361
	geheime PIN					**geheime PIN**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,19485	0,19726	0,70943	0,10396	131	0,19485	0,19726	0,70943	0,10396
anova	89	**0,16332**	**0,16214**	0,83962	0,16390	86	**0,17525**	**0,15454**	0,87736	0,18643
anova-2class	100	0,19234	0,19176	0,72642	0,10994	96	0,19244	0,19073	0,73396	0,11542
correlation	76	0,17672	0,17464	0,81887	0,16814	76	0,17672	0,17464	0,81887	0,16814
entropy	111	0,19396	0,19541	0,71509	0,10591	110	0,19407	0,19541	0,71509	0,10591
joint-entropy	82	0,18202	0,17997	0,81321	0,17315	66	0,19346	0,16941	0,86981	0,20864
simple	31	0,17524	0,17513	0,82453	0,17478	53	0,17759	0,16999	0,79623	0,13621
sfs	93	0,14855	0,14487	0,86038	0,15011	75	***0,15085***	0,14390	0,86604	0,15385
sbs	109	***0,14830***	***0,14394***	0,86226	0,15015	86	0,17450	***0,12280***	0,95660	0,20221
	Pseudonym					**Pseudonym**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,20777	0,21214	0,70943	0,13371	131	0,20777	0,21214	0,70943	0,13371
anova	101	**0,18226**	**0,18202**	0,81698	0,18102	94	**0,19064**	**0,18052**	0,83585	0,19688
anova-2class	118	0,20777	0,21214	0,70943	0,13371	85	0,20832	0,20582	0,75472	0,16636
correlation	109	0,20333	0,20419	0,73019	0,13857	77	0,21014	0,19880	0,82075	0,21836
entropy	118	0,20777	0,21214	0,70943	0,13371	118	0,20777	0,21214	0,70943	0,13371
joint-entropy	88	0,19590	0,19436	0,79811	0,18683	88	0,19590	0,19436	0,79811	0,18683
simple	25	0,18343	0,18287	0,81509	0,18084	25	0,18343	0,18287	0,81509	0,18084
sfs	31	0,18741	0,18692	0,81132	0,18516	89	0,18837	0,18592	0,80189	0,17373
sbs	109	***0,17010***	***0,16849***	0,83396	0,17094	68	***0,17303***	***0,15758***	0,86604	0,18120
	Symbol					**Symbol**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,08165	0,17866	0,66226	0,01959	131	0,08165	0,17866	0,66226	0,01959
anova	95	**0,07444**	**0,15876**	0,70566	0,02319	51	**0,10412**	**0,07896**	0,95283	0,11074
anova-2class	118	0,08165	0,17866	0,66226	0,01959	7	0,11480	0,11451	0,88491	0,11393
correlation	111	0,07881	0,16502	0,69057	0,02061	61	0,10536	0,09967	0,86604	0,06539
entropy	118	0,08165	0,17866	0,66226	0,01959	8	0,12169	0,11484	0,86415	0,09383
joint-entropy	131	0,08165	0,17866	0,66226	0,01959	76	0,11895	0,12144	0,83396	0,07685
simple	50	0,07725	0,16401	0,70189	0,02990	11	0,13563	0,11027	0,92453	0,14507
sfs	29	0,05722	0,05615	0,94151	0,05381	40	0,05826	0,05610	0,93774	0,04993
sbs	36	***0,05264***	***0,05229***	0,94717	0,05174	36	***0,05264***	***0,05229***	0,94717	0,05174
	Woher					**Woher**				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	131	0,19986	0,21190	0,71132	0,13512	131	0,19986	0,21190	0,71132	0,13512
anova	107	**0,19502**	**0,19964**	0,75094	0,15022	98	**0,19646**	**0,18924**	0,82264	0,20112
anova-2class	118	0,19986	0,21190	0,71132	0,13512	91	0,20013	0,20067	0,76981	0,17115
correlation	118	0,19986	0,21190	0,71132	0,13512	90	0,20812	0,20815	0,76038	0,17667
entropy	118	0,19986	0,21190	0,71132	0,13512	118	0,19986	0,21190	0,71132	0,13512
joint-entropy	124	0,19879	0,21252	0,71132	0,13636	68	0,23360	0,20356	0,85283	0,25994
simple	32	0,19001	0,18928	0,80566	0,18422	32	0,19001	0,18928	0,80566	0,18422
sfs	105	0,18366	***0,15490***	0,89057	0,20036	104	***0,18381***	0,15426	0,89245	0,20098
sbs	115	***0,18141***	0,18115	0,81509	0,17739	78	0,19210	***0,14837***	0,92264	0,21938

Tabelle C.20: Secure Sketch: Ergebnisse der Merkmalsselektion bei Parametrisierung $RR \geq 70\%$ für TF für Szenario fix60

	gegebene PIN					geheime PIN				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,36704	0,31887	0,86226	0,50000	60	0,19206	0,17926	0,84151	0,20004
anova	60	0,38592	0,32614	0,93585	0,58813	60	0,18997	**0,15987**	0,88868	0,20842
anova-2class	60	0,37480	0,33828	0,80755	0,48411	60	0,21573	0,21540	0,78302	0,21382
correlation	60	0,38649	0,35094	0,80566	0,50755	60	**0,18741**	0,17919	0,83396	0,19234
entropy	60	**0,34419**	**0,32159**	0,75094	0,39412	60	0,22139	0,21515	0,73208	0,16237
joint-entropy	60	0,39603	0,34280	0,91321	0,59880	60	0,19808	0,17181	0,87170	0,21531
simple	60	***0,32056***	0,29918	0,76038	0,35874	60	0,17837	0,17095	0,79245	0,13436
sfs	60	0,33428	0,29615	0,81887	0,41118	60	***0,15296***	0,14517	0,86604	0,15639
sbs	60	0,32914	***0,28520***	0,84340	0,41379	60	0,17781	***0,12440***	0,95849	0,20729

	Pseudonym					Symbol				
selection	**n**	**EER**	**CRR**	**RR**	**CR**	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,22144	0,20702	0,81887	0,23291	60	0,11928	0,11909	0,85472	0,09289
anova	60	0,27502	0,21954	0,90377	0,34285	60	**0,09131**	**0,08503**	0,92264	0,09271
anova-2class	60	0,23516	0,22495	0,79434	0,24423	60	0,09664	0,13534	0,76226	0,03295
correlation	60	**0,22526**	**0,20894**	0,82075	0,23864	60	0,10581	0,10015	0,86604	0,06633
entropy	60	0,22899	0,22640	0,71698	0,16978	60	0,09448	0,17870	0,67358	0,03099
joint-entropy	60	0,28377	0,23228	0,88679	0,35134	60	0,12997	0,12986	0,83585	0,09557
simple	60	0,20028	0,19829	0,74906	0,14565	60	0,07852	0,16268	0,69811	0,02348
sfs	60	0,18908	0,18663	0,80189	0,17515	60	0,06644	0,06303	0,92264	0,04869
sbs	60	***0,17337***	***0,15780***	0,86604	0,18164	60	***0,06075***	***0,05599***	0,93585	0,04782

	Woher			
selection	**n**	**EER**	**CRR**	**RR**	**CR**
raw	60	0,21758	0,21600	0,78679	0,21880
anova	60	0,25432	**0,21156**	0,87547	0,29858
anova-2class	60	0,26004	0,24302	0,79245	0,27848
correlation	60	0,24140	0,22729	0,80000	0,25457
entropy	60	**0,21938**	0,22184	0,71698	0,16067
joint-entropy	60	0,27442	0,22723	0,87736	0,33182
simple	60	0,19719	0,20528	0,73396	0,14452
sfs	60	***0,18707***	0,15646	0,89245	0,20537
sbs	60	0,19327	***0,14920***	0,92264	0,22104

Tabelle C.21: Biometric Hash: Wrapper basiertes Ranking für Parameter-Set *Null*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	106	102	77	83	63
2	42	54	74	85	50
3	107	79	37	61	76
4	66	114	113	113	112
5	65	72	98	111	81
6	93	116	117	97	116
7	118	105	118	116	118
8	101	109	101	102	106
9	109	117	111	118	115
10	70	90	72	112	109
11	84	88	84	114	103
12	116	118	111	106	101
13	87	96	89	115	111
14	53	92	95	99	117
15	91	17	54	69	45
16	45	105	64	53	29
17	82	19	66	59	78
18	74	67	29	41	96
19	119	119	119	119	119
20	120	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	56	41	59	34	90
24	123	123	123	123	123
25	124	124	124	124	124
26	102	50	55	53	81
27	62	74	44	64	45
28	113	115	103	70	58
29	110	109	48	57	60
30	25	39	52	65	36
31	96	97	109	107	88
32	117	113	91	90	99
33	66	86	31	80	113
34	70	91	64	107	98
35	48	76	104	80	93
36	108	34	40	97	79
37	93	64	48	62	69
38	52	32	55	49	29
39	73	72	99	51	41
40	34	83	99	36	72
41	84	89	115	84	103
42	47	47	72	110	59
43	45	56	69	117	75
44	12	9	47	46	5
45	9	70	33	68	15
46	68	61	42	28	24
47	15	37	80	31	86
48	48	50	87	17	87
49	92	6	94	12	10
50	51	27	85	62	8
51	41	85	10	44	14
52	8	16	5	36	28
53	61	112	27	55	79
54	24	38	87	25	9
55	27	25	74	35	7
56	31	24	31	22	74
57	125	125	125	125	125
58	115	68	86	89	53
59	89	54	33	90	53
60	59	81	60	85	31
61	23	23	107	93	40
62	42	50	53	22	26
63	77	61	7	11	57
64	78	79	1	6	45
65	100	41	48	9	34
66	81	65	35	10	72
67	82	61	30	2	65
68	76	21	44	90	83
69	69	30	80	77	49
70	56	39	62	67	33
71	78	60	57	105	38
72	44	53	7	51	16
73	98	95	61	107	1
74	60	78	106	100	107
75	39	58	109	104	88
76	48	70	40	71	50
77	114	111	96	94	84
78	103	58	108	43	94
79	112	81	114	101	114
80	86	98	116	71	96
81	99	108	17	87	91
82	110	44	26	46	68
83	88	56	91	88	67
84	105	99	35	78	102
85	93	92	57	39	70
86	27	44	66	7	42
87	29	49	38	5	38
88	55	101	101	3	17
89	16	1	68	1	3
90	10	8	9	33	62
91	97	48	79	16	2
92	20	74	16	42	105
93	56	84	77	58	36
94	80	15	90	19	56
95	6	2	3	4	42
96	3	6	4	24	21
97	2	3	6	20	32
98	33	19	76	28	108
99	53	25	23	94	110
100	103	34	28	26	99
101	17	44	14	38	12
102	14	104	13	96	21
103	20	32	25	59	25
104	70	4	82	15	4
105	32	18	82	12	10
106	74	13	91	8	34
107	20	28	39	40	44
108	126	126	126	126	126
109	127	127	127	127	127
110	63	65	22	55	84
111	19	68	97	50	19
112	37	34	17	14	45
113	64	105	51	75	21
114	40	92	62	44	55
115	11	102	43	79	76
116	37	43	15	17	50
117	29	31	24	74	6
118	35	86	104	65	95
119	4	28	20	80	18
120	128	128	128	128	128
121	12	9	20	20	20
122	7	76	12	76	70
123	35	21	46	32	63
124	1	11	19	28	65
125	4	13	11	26	60
126	17	100	70	103	92
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	90	12	70	71	13
131	26	5	2	46	27

Tabelle C.22: Biometric Hash: Wrapper basiertes Ranking für Parameter-Set *minEER*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	113	98	28	37	60
2	2	16	4	21	54
3	55	102	5	15	84
4	49	93	99	27	113
5	78	94	106	1	99
6	78	89	66	86	36
7	105	105	88	111	117
8	75	103	56	78	116
9	108	111	88	110	92
10	39	83	101	85	112
11	25	86	79	79	98
12	74	85	22	31	114
13	23	86	85	24	107
14	99	100	69	102	111
15	94	15	17	94	70
16	20	64	38	40	49
17	109	12	15	51	84
18	51	108	47	23	53
19	119	119	119	119	119
20	120	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	16	54	53	73	84
24	123	123	123	123	123
25	124	124	124	124	124
26	101	101	114	8	99
27	101	97	44	16	64
28	73	114	29	60	69
29	63	110	27	29	70
30	1	7	3	33	41
31	88	79	107	95	90
32	86	71	78	32	43
33	103	95	39	49	67
34	18	54	48	76	78
35	80	76	24	41	78
36	23	45	92	57	99
37	103	71	73	86	68
38	5	58	104	35	42
39	45	50	103	84	83
40	51	57	20	30	31
41	33	39	56	43	45
42	114	70	71	26	72
43	106	63	118	71	34
44	6	6	18	51	14
45	12	39	40	39	25
46	46	74	63	25	10
47	69	62	14	74	3
48	19	79	22	28	7
49	9	36	26	10	9
50	14	9	82	12	4
51	75	42	21	12	7
52	15	68	30	7	13
53	83	73	33	44	16
54	56	115	12	3	17
55	59	24	13	2	6
56	20	19	54	63	81
57	125	125	125	125	125
58	112	88	110	108	50
59	37	21	61	70	39
60	39	74	59	37	38
61	36	8	19	54	47
62	85	59	11	53	19
63	100	103	10	4	2
64	97	112	1	22	1
65	60	83	6	41	17
66	80	33	41	9	24
67	75	25	61	6	30
68	56	23	94	18	107
69	44	64	75	68	57
70	92	10	111	55	31
71	56	69	43	76	43
72	10	59	45	91	11
73	82	18	2	72	5
74	96	13	116	89	90
75	32	47	113	116	110
76	61	95	72	61	33
77	17	19	65	82	29
78	51	115	115	106	22
79	89	42	80	69	51
80	117	81	94	75	26
81	94	109	66	80	20
82	118	118	86	59	47
83	115	66	35	66	15
84	111	37	49	93	58
85	26	28	52	100	109
86	67	46	9	5	21
87	39	41	6	20	93
88	89	22	101	64	96
89	70	2	50	17	36
90	22	1	108	114	118
91	35	78	8	107	55
92	70	51	117	115	55
93	4	91	94	104	46
94	37	33	34	96	75
95	7	5	25	64	26
96	7	4	16	36	11
97	3	3	31	11	28
98	30	27	54	46	105
99	33	99	45	103	74
100	65	82	42	56	59
101	42	106	32	62	22
102	29	54	56	90	35
103	49	107	87	80	62
104	110	47	99	96	60
105	91	28	68	112	102
106	43	28	35	91	78
107	92	61	73	66	95
108	126	126	126	126	126
109	127	127	127	127	127
110	67	91	109	105	106
111	12	32	51	117	97
112	62	38	70	109	89
113	27	49	104	88	93
114	64	31	82	48	75
115	116	112	98	99	115
116	107	25	92	57	62
117	27	76	80	83	82
118	47	10	75	50	64
119	48	89	60	118	51
120	128	128	128	128	128
121	30	51	91	19	88
122	51	42	90	44	72
123	65	51	84	47	102
124	83	67	97	100	102
125	98	33	63	33	84
126	70	117	75	96	39
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	87	17	112	113	64
131	11	14	37	14	77

Tabelle C.23: Biometric Hash: Wrapper basiertes Ranking für Parameter-Set *minCRR*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	82	9	71	12	19
2	50	58	51	61	51
3	24	73	15	20	27
4	1	82	95	4	111
5	26	4	3	17	14
6	101	60	19	103	5
7	113	96	103	100	105
8	74	8	37	83	92
9	119	120	98	89	76
10	12	39	48	42	69
11	14	49	73	72	57
12	39	87	74	25	64
13	59	34	67	69	55
14	2	27	6	93	90
15	113	53	87	85	56
16	51	36	54	49	30
17	116	84	66	74	94
18	15	50	25	44	45
19	117	117	118	118	119
20	118	118	119	119	120
21	120	119	120	120	121
22	121	121	121	121	122
23	17	12	25	37	77
24	122	122	122	122	123
25	123	123	123	123	124
26	23	31	30	14	54
27	27	61	12	16	22
28	100	81	3	35	34
29	36	93	50	1	89
30	6	2	1	53	1
31	84	1	5	15	2
32	10	44	83	52	86
33	96	74	46	64	45
34	42	24	29	58	9
35	37	100	55	61	111
36	20	18	13	63	16
37	60	23	59	87	17
38	61	40	43	39	43
39	40	52	105	33	26
40	76	91	87	38	73
41	92	57	89	59	88
42	32	100	93	57	47
43	89	103	110	79	67
44	7	18	17	47	11
45	18	43	42	41	24
46	72	98	99	30	34
47	65	68	45	40	49
48	108	84	18	3	66
49	57	105	19	9	52
50	85	20	14	11	50
51	38	22	35	19	74
52	58	78	85	7	65
53	90	94	39	20	63
54	93	64	61	6	81
55	55	72	53	2	40
56	68	66	70	64	39
57	124	125	125	124	125
58	35	35	89	28	86
59	34	47	11	8	17
60	8	5	76	20	31
61	83	69	8	29	42
62	47	55	81	68	85
63	78	25	32	5	3
64	91	51	8	23	29
65	30	17	64	60	72
66	13	15	58	13	25
67	32	14	2	17	23
68	29	86	40	30	36
69	25	64	24	45	6
70	45	42	31	80	79
71	72	82	34	91	84
72	30	48	28	94	7
73	62	88	32	75	4
74	22	29	76	43	58
75	86	38	57	117	59
76	4	13	101	78	28
77	105	71	84	103	102
78	93	69	60	96	71
79	80	66	107	97	77
80	110	115	65	82	99
81	41	41	21	101	8
82	77	28	47	91	82
83	44	16	96	99	37
84	101	75	69	102	40
85	87	112	85	67	69
86	53	6	51	10	12
87	19	25	16	33	32
88	74	62	110	46	67
89	47	3	25	24	21
90	54	45	35	110	117
91	45	88	41	55	105
92	107	114	115	111	116
93	63	109	117	115	47
94	68	59	48	86	61
95	93	54	61	73	44
96	43	76	22	54	37
97	66	20	7	26	13
98	98	106	93	49	94
99	106	110	97	107	107
100	81	113	102	108	83
101	104	104	106	109	109
102	88	106	99	98	104
103	78	97	80	89	97
104	9	32	91	84	15
105	11	6	23	81	53
106	70	37	43	76	59
107	124	124	113	71	80
108	126	126	126	126	126
109	127	127	127	127	127
110	99	91	74	27	97
111	52	111	78	125	93
112	97	99	112	70	99
113	55	46	116	106	110
114	66	108	108	66	101
115	112	79	114	116	96
116	71	79	79	87	75
117	21	30	103	77	90
118	101	90	81	56	113
119	16	32	38	105	62
120	128	128	128	128	128
121	47	94	56	35	114
122	64	63	68	48	107
123	115	55	61	112	33
124	28	102	91	94	115
125	5	11	72	51	10
126	111	77	108	113	102
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	109	116	123	114	118
131	3	10	10	32	19

Tabelle C.24: Biometric Hash: Wrapper basiertes Ranking für Parameter-Set $RR \geq 80\%$

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	75	14	72	16	5
2	32	62	52	46	52
3	18	78	16	19	40
4	1	81	93	6	101
5	31	4	3	7	19
6	105	59	31	105	15
7	108	106	106	111	101
8	84	6	38	71	51
9	109	103	100	91	76
10	15	30	48	76	86
11	5	17	75	82	71
12	51	73	71	23	62
13	45	20	68	43	67
14	3	1	6	97	95
15	115	72	70	79	90
16	85	46	55	54	64
17	119	68	79	57	99
18	14	61	30	35	60
19	116	118	118	119	117
20	117	119	119	120	118
21	119	120	120	121	120
22	121	121	121	122	122
23	36	16	31	68	85
24	122	122	122	123	123
25	123	124	123	124	124
26	44	37	28	15	66
27	25	76	11	14	31
28	86	83	2	65	84
29	58	94	45	1	3
30	35	2	1	39	1
31	82	67	5	94	2
32	8	9	90	33	88
33	53	85	47	51	68
34	50	28	26	54	10
35	38	91	61	48	81
36	20	25	13	61	8
37	70	25	65	78	26
38	70	49	53	40	37
39	57	48	104	74	27
40	78	97	84	36	52
41	45	68	84	49	16
42	39	99	97	37	46
43	90	101	107	58	23
44	4	13	8	59	7
45	25	47	29	21	39
46	78	53	95	30	54
47	65	102	44	74	69
48	100	99	21	4	92
49	41	39	25	13	75
50	21	35	13	16	36
51	56	24	36	19	91
52	61	50	81	8	86
53	99	97	38	18	82
54	94	80	62	4	96
55	66	81	54	2	56
56	76	86	77	32	62
57	124	125	124	125	125
58	30	33	82	92	33
59	37	55	10	11	17
60	7	18	40	25	59
61	90	78	7	64	18
62	66	58	59	61	94
63	48	29	31	3	6
64	62	55	9	27	40
65	47	21	69	49	99
66	19	5	64	10	20
67	49	22	3	12	43
68	12	19	41	34	73
69	9	75	21	38	3
70	41	35	42	73	27
71	83	84	48	87	12
72	29	45	31	89	14
73	59	90	50	68	23
74	24	32	74	44	93
75	104	43	58	115	83
76	2	10	94	81	47
77	95	62	104	104	105
78	88	66	88	94	106
79	72	68	98	96	107
80	105	115	102	97	108
81	41	57	23	101	9
82	74	27	13	87	29
83	22	12	102	100	32
84	103	73	77	102	72
85	93	117	96	68	76
86	52	6	51	9	13
87	25	38	20	26	48
88	73	60	108	31	42
89	60	3	19	23	21
90	68	64	35	112	49
91	39	41	46	109	21
92	117	111	115	115	114
93	110	114	117	109	57
94	33	39	60	84	64
95	101	71	66	63	57
96	80	50	23	53	69
97	13	11	17	22	25
98	89	111	87	45	108
99	113	113	108	106	110
100	80	107	92	89	111
101	107	107	110	76	112
102	87	110	101	72	113
103	90	104	79	82	101
104	16	43	91	108	30
105	11	8	27	84	35
106	55	31	18	86	50
107	114	123	114	65	101
108	126	126	126	126	126
109	127	127	127	127	127
110	96	93	88	102	11
111	34	109	76	118	43
112	54	89	110	99	78
113	63	53	116	107	96
114	98	96	84	42	74
115	96	77	113	117	45
116	63	87	83	60	98
117	28	22	99	67	60
118	112	92	63	56	80
119	17	42	43	114	54
120	128	128	128	128	128
121	77	65	57	27	115
122	69	95	73	47	89
123	125	50	67	52	78
124	22	105	55	80	118
125	6	34	37	41	34
126	102	88	112	93	116
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	111	116	125	113	121
131	10	14	12	29	37

Tabelle C.25: Biometric Hash: Wrapper basiertes Ranking für Parameter-Set $CR \leq 5\%$

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	82	19	57	16	20
2	19	23	11	46	39
3	102	35	12	19	25
4	87	69	105	6	107
5	8	22	2	7	12
6	77	49	19	105	76
7	116	73	99	111	103
8	61	27	40	71	95
9	106	121	77	91	73
10	4	16	88	76	74
11	73	54	69	82	58
12	43	75	72	23	51
13	69	51	49	43	55
14	1	29	3	97	94
15	99	25	80	79	48
16	85	85	52	54	28
17	100	55	63	57	86
18	28	101	21	35	39
19	119	117	119	119	119
20	120	118	120	120	120
21	121	119	121	121	121
22	122	120	122	122	122
23	32	6	35	68	59
24	123	122	123	123	123
25	124	123	124	124	124
26	14	26	28	15	50
27	15	62	6	14	19
28	101	48	36	65	32
29	108	46	17	1	88
30	2	1	22	39	1
31	87	3	28	94	27
32	55	38	77	33	91
33	89	83	51	51	69
34	25	58	33	54	7
35	47	58	60	48	109
36	38	41	18	61	10
37	52	11	77	78	18
38	23	71	46	40	45
39	75	74	107	74	70
40	102	98	84	36	67
41	86	28	74	49	88
42	30	95	94	37	49
43	65	97	111	58	77
44	16	24	13	59	4
45	6	21	33	21	23
46	75	64	75	30	34
47	41	66	39	74	54
48	117	109	16	4	71
49	34	108	14	13	55
50	79	41	9	16	45
51	83	14	42	19	64
52	55	70	64	8	66
53	37	65	41	18	62
54	57	34	53	4	92
55	20	67	54	2	31
56	5	44	61	32	42
57	125	124	125	125	125
58	104	96	91	92	88
59	63	19	7	11	16
60	91	47	94	25	29
61	79	56	8	64	38
62	98	40	67	61	77
63	31	15	37	3	2
64	46	33	5	27	25
65	11	45	62	49	64
66	29	10	20	10	20
67	10	13	1	12	15
68	36	81	24	34	36
69	12	53	26	38	5
70	61	17	26	73	72
71	32	93	31	87	83
72	23	39	37	89	6
73	109	77	25	68	3
74	18	31	65	44	59
75	54	103	101	115	44
76	38	8	108	81	24
77	66	116	80	104	101
78	107	112	65	94	67
79	70	49	103	96	75
80	118	103	71	97	100
81	93	81	28	101	11
82	111	105	43	87	82
83	105	12	90	100	37
84	53	102	100	102	41
85	57	94	82	68	55
86	17	5	47	9	9
87	70	18	10	26	32
88	97	110	112	31	103
89	45	2	32	23	22
90	35	30	112	112	117
91	38	106	49	109	111
92	77	113	114	115	114
93	22	88	115	109	47
94	59	37	45	84	59
95	41	31	57	63	35
96	51	83	55	53	43
97	84	9	14	22	13
98	50	79	86	45	97
99	73	86	96	106	108
100	43	98	83	89	84
101	66	72	97	76	109
102	66	78	92	72	102
103	64	68	72	82	87
104	26	56	93	108	14
105	27	41	22	84	52
106	60	35	59	86	53
107	115	124	117	65	81
108	126	126	126	126	126
109	127	127	127	127	127
110	89	92	48	102	99
111	79	111	102	118	96
112	95	76	97	99	80
113	48	98	115	107	113
114	91	90	106	42	79
115	114	58	110	117	106
116	96	91	104	60	85
117	72	63	108	67	105
118	113	107	70	56	93
119	7	61	43	114	63
120	128	128	128	128	128
121	9	80	56	27	115
122	13	89	67	47	112
123	112	52	87	52	29
124	49	87	76	80	116
125	20	7	89	41	7
126	110	115	84	93	98
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	93	114	118	113	118
131	3	4	4	29	16

Tabelle C.26: Secure Sketch: Wrapper basiertes Ranking für Parameter-Set *Null*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	108	73	104	88	76
2	57	24	58	100	51
3	59	96	40	68	34
4	106	56	102	94	93
5	62	106	114	105	78
6	17	13	10	23	15
7	118	117	114	114	113
8	100	113	94	76	80
9	116	115	118	112	106
10	62	68	86	107	71
11	45	109	50	81	98
12	95	92	53	97	95
13	91	85	40	109	105
14	113	115	117	72	115
15	39	11	7	43	26
16	19	45	18	40	36
17	53	18	12	38	38
18	90	40	20	55	17
19	119	119	119	119	119
20	120	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	42	61	56	71	39
24	123	123	123	123	123
25	124	124	124	124	124
26	45	27	16	59	52
27	54	19	14	58	55
28	114	104	106	35	110
29	112	88	109	75	114
30	22	12	43	94	28
31	103	66	81	101	101
32	74	90	111	113	110
33	76	65	114	108	118
34	69	118	108	106	90
35	114	113	113	81	99
36	19	40	76	111	96
37	30	95	68	48	45
38	23	29	86	67	66
39	26	99	99	92	117
40	117	112	110	118	112
41	69	76	111	110	87
42	50	73	85	115	75
43	111	57	64	89	69
44	13	26	7	86	14
45	10	33	27	81	27
46	81	79	45	102	80
47	33	54	58	27	43
48	57	67	30	116	19
49	78	49	50	104	17
50	76	35	54	62	58
51	18	34	45	30	25
52	39	63	24	61	66
53	107	80	49	116	20
54	73	87	68	64	53
55	43	55	34	43	16
56	37	61	30	53	35
57	125	125	125	125	125
58	108	108	28	93	13
59	72	40	25	35	109
60	36	100	72	50	60
61	59	51	83	76	54
62	38	40	47	26	42
63	12	50	103	19	108
64	54	31	77	21	22
65	27	13	33	69	7
66	39	47	98	16	100
67	34	35	57	29	50
68	50	77	84	64	63
69	104	15	79	91	60
70	32	29	38	11	33
71	86	25	36	17	9
72	11	10	4	13	6
73	15	4	25	18	23
74	66	57	50	50	40
75	83	19	90	97	47
76	4	6	71	49	5
77	83	17	5	12	90
78	91	105	92	5	93
79	104	96	89	30	115
80	13	81	6	10	84
81	96	83	38	8	30
82	93	5	32	8	73
83	45	71	37	13	10
84	44	53	9	15	63
85	52	70	62	4	12
86	45	39	73	47	2
87	64	9	29	32	2
88	100	107	96	22	80
89	6	23	2	39	29
90	9	60	74	24	60
91	100	100	63	103	30
92	65	109	90	79	103
93	28	86	17	78	71
94	87	22	99	5	97
95	2	3	23	56	8
96	3	7	48	64	4
97	5	2	54	46	76
98	99	57	19	73	32
99	29	98	20	37	41
100	93	51	107	25	106
101	66	94	64	74	43
102	80	100	12	41	101
103	21	45	11	33	11
104	75	40	97	2	87
105	56	75	94	3	86
106	66	16	92	1	89
107	110	28	101	52	70
108	126	126	126	126	126
109	127	127	127	127	127
110	96	93	77	79	103
111	69	100	43	86	74
112	89	89	61	94	83
113	31	48	86	99	57
114	82	81	70	42	68
115	61	90	67	90	65
116	24	77	60	62	85
117	78	38	15	81	48
118	87	68	80	59	56
119	83	83	64	85	36
120	128	128	128	128	128
121	8	32	40	56	45
122	6	1	35	53	49
123	1	72	22	33	78
124	16	8	75	43	1
125	35	37	3	20	58
126	49	64	81	27	90
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	98	111	104	70	21
131	24	21	1	7	24

Tabelle C.27: Secure Sketch: Wrapper basiertes Ranking für Parameter-Set *minEER*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	82	9	88	27	5
2	10	11	3	22	21
3	41	68	33	42	37
4	4	27	68	13	64
5	23	13	6	7	30
6	2	1	1	1	7
7	116	116	113	37	117
8	48	50	49	41	69
9	112	114	116	108	114
10	33	8	55	66	81
11	27	31	76	82	28
12	17	38	61	43	84
13	42	3	14	75	44
14	3	33	10	48	101
15	104	107	108	114	91
16	95	92	81	108	47
17	101	95	107	111	106
18	15	30	30	57	96
19	119	118	119	119	119
20	120	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	69	42	70	85	92
24	123	123	123	123	123
25	124	124	124	124	124
26	45	56	55	28	86
27	32	54	19	24	31
28	68	66	8	81	83
29	44	38	34	11	87
30	20	35	37	39	1
31	78	2	9	32	2
32	34	4	69	65	13
33	79	98	65	82	56
34	40	9	7	54	3
35	29	40	53	95	12
36	61	16	12	25	24
37	55	33	47	70	40
38	82	53	31	107	63
39	42	22	109	104	19
40	86	19	58	117	65
41	20	69	67	38	40
42	48	85	5	32	4
43	18	93	2	70	34
44	8	36	35	60	10
45	66	59	18	35	23
46	88	46	66	52	26
47	64	81	78	9	74
48	118	67	41	69	105
49	87	101	53	12	78
50	19	57	22	21	45
51	63	24	43	28	80
52	91	88	16	15	75
53	51	64	60	87	71
54	105	78	86	6	60
55	75	100	74	8	66
56	36	80	93	62	62
57	125	125	125	125	125
58	35	32	82	96	87
59	47	47	45	23	55
60	30	14	64	55	58
61	13	16	24	31	42
62	56	71	90	112	102
63	70	89	4	9	6
64	80	76	17	16	29
65	46	72	98	106	103
66	24	51	61	14	27
67	64	42	20	19	71
68	37	26	14	43	33
69	1	6	29	43	14
70	72	64	115	93	18
71	85	87	63	90	11
72	30	74	23	5	84
73	115	94	73	78	8
74	28	63	79	50	95
75	98	54	85	115	68
76	6	15	52	19	49
77	110	78	38	73	38
78	107	115	101	2	98
79	90	109	11	87	38
80	57	37	28	4	93
81	107	27	35	90	35
82	97	21	51	32	46
83	25	42	88	78	24
84	73	98	103	36	112
85	89	110	97	103	99
86	67	25	57	16	16
87	25	23	24	51	57
88	77	70	112	49	81
89	81	6	31	40	22
90	53	84	58	90	115
91	58	40	46	56	31
92	94	102	80	53	118
93	113	106	98	112	93
94	11	5	24	47	43
95	99	48	40	100	35
96	22	62	50	60	52
97	16	18	13	30	9
98	101	112	110	66	108
99	59	73	70	96	75
100	14	77	101	59	107
101	103	111	114	64	109
102	107	113	111	66	110
103	100	107	106	80	111
104	7	20	77	16	17
105	12	12	20	43	19
106	48	59	43	3	50
107	117	119	96	108	104
108	126	126	126	126	126
109	127	127	127	127	127
110	93	90	38	26	15
111	52	82	42	116	78
112	54	75	95	89	73
113	74	49	118	77	77
114	96	83	100	86	87
115	110	86	90	75	50
116	60	90	103	72	99
117	71	57	87	96	48
118	92	105	83	63	67
119	9	51	72	93	69
120	128	128	128	128	128
121	61	95	83	74	90
122	75	104	105	58	53
123	84	95	75	102	61
124	39	59	92	96	113
125	37	29	48	84	58
126	106	102	94	104	97
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	113	117	117	118	115
131	5	45	27	101	54

Tabelle C.28: Secure Sketch: Wrapper basiertes Ranking für Parameter-Set *minCRR*

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	8	9	81	12	8
2	15	11	13	58	37
3	31	68	25	23	56
4	4	27	30	28	87
5	32	13	9	7	31
6	1	1	1	1	26
7	76	116	106	57	113
8	61	50	43	56	36
9	104	114	113	113	115
10	20	8	60	50	85
11	11	31	68	73	66
12	36	38	44	42	72
13	29	3	17	86	22
14	5	33	11	83	97
15	93	107	108	112	102
16	100	92	77	80	74
17	77	95	110	110	106
18	43	30	35	59	85
19	119	118	118	119	119
20	120	120	119	120	120
21	121	121	120	121	121
22	122	122	121	122	122
23	81	42	90	67	95
24	123	123	122	123	123
25	124	124	123	124	124
26	95	56	63	20	94
27	34	54	27	19	46
28	29	66	10	77	51
29	103	38	26	6	58
30	41	35	11	25	1
31	3	2	6	11	7
32	10	4	56	31	13
33	59	98	58	81	65
34	72	9	5	64	3
35	48	40	76	106	10
36	34	16	13	32	28
37	59	33	35	83	38
38	65	53	38	113	55
39	57	22	67	107	28
40	23	19	52	115	68
41	62	69	69	44	5
42	67	85	3	25	4
43	18	93	2	111	2
44	22	36	58	53	11
45	58	59	30	51	20
46	74	46	81	29	56
47	64	81	75	10	77
48	104	67	79	36	88
49	25	101	72	16	83
50	33	57	38	12	54
51	80	24	54	24	76
52	98	88	19	14	40
53	87	64	86	54	89
54	84	78	88	4	79
55	70	100	85	4	63
56	85	80	90	60	66
57	125	125	124	125	125
58	2	32	32	100	83
59	53	47	70	36	49
60	12	14	45	35	61
61	43	16	29	45	49
62	63	71	98	108	102
63	93	89	4	7	9
64	95	76	22	17	39
65	72	72	102	102	108
66	102	51	33	9	32
67	78	42	55	51	68
68	25	26	18	32	26
69	7	6	41	22	15
70	74	64	116	109	16
71	95	87	56	94	16
72	47	74	47	38	93
73	100	94	66	72	41
74	38	63	81	62	73
75	99	54	93	116	64
76	9	15	16	48	44
77	83	78	37	75	12
78	81	115	95	2	23
79	106	109	7	90	53
80	6	37	28	3	89
81	20	27	8	90	6
82	12	21	49	27	44
83	19	42	89	83	21
84	65	98	101	81	110
85	107	110	97	99	99
86	90	25	78	15	19
87	56	23	40	30	74
88	89	70	70	66	78
89	108	6	21	34	24
90	109	84	65	102	71
91	52	40	53	61	47
92	110	102	87	87	117
93	111	106	102	73	92
94	39	5	24	45	33
95	112	48	48	93	62
96	14	62	62	55	47
97	45	18	15	38	14
98	113	112	90	95	107
99	39	73	73	97	108
100	16	77	51	68	70
101	114	111	111	92	110
102	91	113	106	87	112
103	115	107	105	95	113
104	46	20	20	38	25
105	28	12	23	49	28
106	25	59	46	18	34
107	24	119	112	104	102
108	126	126	126	126	126
109	127	127	127	127	127
110	53	90	64	21	18
111	41	82	42	117	80
112	50	75	95	79	91
113	55	49	125	71	43
114	78	83	102	77	96
115	85	86	84	62	58
116	87	90	98	97	101
117	70	57	109	89	42
118	69	105	94	75	82
119	16	51	61	101	81
120	128	128	128	128	128
121	68	95	114	42	100
122	91	104	115	45	60
123	116	95	98	70	97
124	51	59	74	69	116
125	36	29	50	65	35
126	117	102	80	104	105
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	118	117	117	118	118
131	49	45	34	41	52

Tabelle C.29: Secure Sketch: Wrapper basiertes Ranking für Parameter-Set $RR \geq 70\%$

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	115	80	14	58	70
2	40	22	7	46	66
3	95	86	2	24	29
4	91	99	78	10	115
5	92	90	51	22	80
6	57	13	11	1	32
7	119	117	91	116	109
8	105	114	19	61	112
9	117	118	117	117	100
10	64	77	79	68	110
11	81	101	103	64	107
12	103	112	47	81	102
13	85	110	87	65	111
14	102	116	108	52	118
15	85	14	29	72	81
16	43	64	22	49	47
17	94	26	26	56	92
18	83	81	23	24	42
19	118	119	119	119	119
20	119	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	63	55	46	37	63
24	123	123	123	123	123
25	124	124	124	124	124
26	103	81	66	11	90
27	110	92	84	50	102
28	116	94	32	70	76
29	111	95	33	47	86
30	23	11	1	31	51
31	93	97	64	103	108
32	99	61	89	80	76
33	78	97	51	67	117
34	61	111	82	85	61
35	112	115	93	105	114
36	78	41	28	26	84
37	67	89	45	84	89
38	83	60	88	115	61
39	88	75	93	97	94
40	109	108	93	92	113
41	85	48	71	39	84
42	37	66	107	65	98
43	108	55	97	95	90
44	14	8	3	35	14
45	7	8	4	76	38
46	74	71	20	28	32
47	29	48	23	13	20
48	60	50	17	43	26
49	96	15	13	8	18
50	101	15	55	7	30
51	18	33	10	9	22
52	8	30	6	21	57
53	89	68	8	51	31
54	12	35	17	3	45
55	11	19	15	2	43
56	36	18	75	54	34
57	125	125	125	125	125
58	114	113	115	113	81
59	47	68	49	63	58
60	45	52	57	75	64
61	39	47	44	44	56
62	70	70	38	76	12
63	26	84	63	4	53
64	55	44	9	27	4
65	51	66	30	53	11
66	64	27	62	6	21
67	98	7	65	19	10
68	25	54	60	33	74
69	58	27	41	54	73
70	12	24	102	71	8
71	22	91	90	105	35
72	20	59	21	69	3
73	51	24	16	86	6
74	99	85	98	114	60
75	75	102	112	110	55
76	9	10	80	14	1
77	45	38	109	87	39
78	72	55	118	28	45
79	17	62	109	98	50
80	28	43	101	16	48
81	53	104	85	99	15
82	68	52	71	73	97
83	20	62	71	92	39
84	14	72	83	101	79
85	37	83	43	104	87
86	6	12	76	5	2
87	10	34	5	12	5
88	105	104	53	22	102
89	26	1	30	19	7
90	54	22	85	88	99
91	89	79	25	109	8
92	66	109	109	118	78
93	75	27	98	82	16
94	69	73	41	41	69
95	4	4	37	32	28
96	5	3	49	28	13
97	2	2	12	18	54
98	49	106	116	76	24
99	113	106	39	73	27
100	107	102	47	96	37
101	29	31	27	42	19
102	19	31	35	94	22
103	47	76	54	76	17
104	41	35	61	15	72
105	58	21	36	44	96
106	70	50	40	39	74
107	34	37	56	57	25
108	126	126	126	126	126
109	127	127	127	127	127
110	72	58	114	100	116
111	43	86	104	107	67
112	42	78	77	82	70
113	82	74	100	89	65
114	35	95	67	36	87
115	97	39	92	91	39
116	24	92	57	60	51
117	78	40	59	59	36
118	62	44	81	48	101
119	50	19	69	111	49
120	128	128	128	128	128
121	32	5	70	102	58
122	14	46	68	90	95
123	32	64	113	34	102
124	1	42	96	38	102
125	31	17	106	61	81
126	55	88	71	108	92
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	75	100	105	112	43
131	3	6	33	17	68

Tabelle C.30: Secure Sketch: Wrapper basiertes Ranking für Parameter-Set $CR \leq 5\%$

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	62	6	85	55	17
2	3	30	1	60	28
3	27	81	37	38	28
4	74	56	88	12	110
5	5	12	2	9	20
6	9	1	22	1	23
7	118	115	116	26	118
8	29	31	51	29	75
9	113	116	117	110	114
10	18	15	64	76	54
11	12	14	71	72	82
12	7	39	79	43	70
13	51	19	48	71	15
14	17	90	5	38	107
15	94	106	91	100	77
16	67	63	73	101	40
17	106	100	90	102	102
18	22	19	45	52	81
19	119	119	118	119	119
20	120	120	119	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	34	27	61	80	84
24	123	123	123	123	123
25	124	124	124	124	124
26	34	64	36	22	67
27	29	42	13	32	24
28	99	62	3	88	84
29	63	35	17	8	83
30	2	4	16	21	1
31	78	2	4	61	2
32	32	8	51	49	8
33	94	80	66	98	62
34	38	13	7	37	3
35	68	58	50	99	12
36	72	22	8	16	22
37	52	51	57	76	28
38	72	59	21	115	44
39	56	28	107	105	18
40	92	47	65	118	59
41	60	69	68	29	38
42	27	73	29	27	7
43	46	95	35	74	31
44	13	36	12	67	9
45	24	70	32	56	19
46	93	50	68	68	14
47	83	103	60	18	60
48	113	83	55	34	103
49	87	108	46	6	73
50	55	77	15	12	27
51	8	26	27	43	53
52	75	84	48	15	63
53	98	67	55	42	63
54	101	57	75	10	58
55	50	79	70	11	51
56	24	86	77	82	51
57	125	125	125	125	125
58	96	34	88	48	94
59	19	41	17	14	49
60	31	15	33	79	39
61	40	17	24	24	26
62	59	49	71	103	97
63	45	68	22	16	4
64	69	94	10	57	34
65	11	32	104	107	98
66	6	53	38	41	16
67	49	24	6	46	47
68	14	3	11	47	50
69	15	5	44	40	11
70	54	36	110	88	25
71	39	65	51	95	9
72	34	65	14	5	87
73	108	103	59	88	5
74	48	75	86	36	71
75	78	53	92	107	61
76	70	45	62	23	42
77	111	87	42	64	89
78	109	114	111	2	57
79	61	113	58	95	63
80	115	92	38	4	105
81	102	28	76	97	33
82	104	40	81	25	41
83	53	47	96	82	46
84	107	97	107	28	111
85	76	97	95	103	92
86	40	17	25	19	13
87	47	23	33	62	45
88	103	76	112	45	63
89	57	7	38	31	48
90	22	70	103	76	114
91	78	90	42	86	94
92	115	112	114	50	117
93	91	99	109	110	100
94	10	46	20	34	56
95	88	44	31	85	21
96	42	38	30	65	37
97	44	33	9	62	6
98	76	110	100	53	105
99	82	105	79	69	67
100	58	89	106	58	116
101	84	106	102	54	107
102	86	110	104	50	109
103	81	109	101	59	104
104	21	24	78	7	36
105	19	9	19	33	43
106	34	10	27	3	32
107	117	118	96	88	90
108	126	126	126	126	126
109	127	127	127	127	127
110	85	92	67	20	79
111	64	81	51	80	80
112	88	78	63	74	71
113	65	43	115	66	75
114	66	95	96	72	96
115	105	88	112	106	93
116	71	72	99	107	90
117	96	51	84	110	67
118	90	85	86	70	88
119	4	53	47	86	73
120	128	128	128	128	128
121	24	101	74	92	84
122	43	61	83	93	78
123	112	74	82	113	34
124	15	60	92	94	112
125	33	10	38	82	101
126	100	102	94	114	98
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	110	117	120	117	112
131	1	21	26	116	55

Tabelle C.31: Ranking der Merkmale basierend auf den Filtern in Abhängigkeit von der Semantik

n	Geg. PIN	Geh. PIN	Pseu-donym	Symbol	Woher
1	61	52	9	84	29
2	49	12	8	13	6
3	25	31	29	46	38
4	34	13	55	34	47
5	4	5	19	52	69
6	2	1	1	1	2
7	56	81	79	100	58
8	10	8	68	5	63
9	39	18	34	102	75
10	5	29	30	7	20
11	42	23	56	35	21
12	17	11	35	72	41
13	35	28	74	45	19
14	1	2	3	3	3
15	58	39	11	42	32
16	73	89	77	105	15
17	80	39	16	39	81
18	26	3	25	29	8
19	119	119	119	119	119
20	120	120	120	120	120
21	121	121	121	121	121
22	122	122	122	122	122
23	6	52	40	8	47
24	123	123	123	123	123
25	124	124	124	124	124
26	57	65	15	74	53
27	14	30	13	16	9
28	81	93	38	52	91
29	73	69	57	4	44
30	68	25	12	55	10
31	20	16	74	24	11
32	40	18	48	76	88
33	66	71	77	44	71
34	60	36	47	21	27
35	47	55	62	76	101
36	78	58	87	48	35
37	55	45	74	12	28
38	76	32	81	23	35
39	77	58	87	70	18
40	50	79	84	19	43
41	83	35	58	70	31
42	35	34	73	97	25
43	38	36	44	66	21
44	99	65	23	93	26
45	101	69	25	89	57
46	96	108	107	81	82
47	107	105	68	62	84
48	115	117	102	27	106
49	90	78	41	100	56
50	95	83	37	49	24
51	58	79	48	42	66
52	93	110	65	67	88
53	117	107	59	25	107
54	96	71	33	25	113
55	71	49	31	59	40
56	8	68	65	18	60
57	125	125	125	125	125
58	79	54	103	117	103
59	21	46	46	89	91
60	43	27	72	33	21
61	19	41	65	98	45
62	86	91	82	98	93
63	12	76	18	15	30
64	54	71	51	68	61
65	50	47	21	35	47
66	37	20	43	6	69
67	46	21	59	76	63
68	7	4	39	19	75
69	18	15	44	110	41
70	16	17	7	60	7
71	8	36	5	31	4
72	30	26	2	2	1
73	3	22	4	28	5
74	94	99	106	69	94
75	105	97	95	89	96
76	13	6	6	16	37
77	26	33	23	74	50
78	30	51	35	56	74
79	61	63	28	80	38
80	32	47	48	8	51
81	24	65	54	8	33
82	71	50	10	29	13
83	40	63	51	49	17
84	14	10	64	41	12
85	70	9	22	57	54
86	29	77	82	38	72
87	26	24	42	112	68
88	106	113	99	84	97
89	108	112	86	72	111
90	102	100	109	88	102
91	113	86	85	81	85
92	92	105	71	102	109
93	108	57	90	106	115
94	33	44	31	57	66
95	73	13	13	22	16
96	87	71	20	51	34
97	84	43	17	11	14
98	87	93	103	109	99
99	50	103	92	83	111
100	110	85	105	64	113
101	66	91	91	104	83
102	50	97	108	107	108
103	87	100	62	92	78
104	104	89	97	14	55
105	96	118	80	63	105
106	64	87	61	39	98
107	115	116	111	95	109
108	126	126	126	126	126
109	127	127	127	127	127
110	100	41	111	31	46
111	84	115	116	113	95
112	91	95	113	47	62
113	102	111	89	108	117
114	111	108	117	35	51
115	114	88	93	118	103
116	112	104	113	54	87
117	64	61	97	111	79
118	82	102	96	65	58
119	43	58	68	114	72
120	128	128	128	128	128
121	43	114	110	60	116
122	47	96	94	87	88
123	68	71	100	76	86
124	23	56	101	96	75
125	22	61	53	94	79
126	61	82	115	115	100
127	129	129	129	129	129
128	130	130	130	130	130
129	131	131	131	131	131
130	118	84	118	116	118
131	10	7	27	86	63