Data mining in diagnostic charts and treatment outcome prediction for Vision Restoration Therapy

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

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

von Diplominformatiker Tobias Günther,
geboren am 24.11.1979 in Staßfurt

Gutachter:
Prof. Dr. Rudolf Kruse
Prof. Dr. Bernhard A. Sabel
PD Dr. Christian Borgelt

Magdeburg, den 25.03.2008

Für meine Eltern, Elisabeth und Helge

Für Julia

Diese Arbeiten wurden durchgeführt am Institut für Medizinische Psychologie in Magdeburg an der Otto-von-Guericke Universität von Oktober 2004 bis Januar 2008.

Danksagung

Abstract

In medicine, treatment outcome prediction models (TOPM) are used to compare the restoration potential with the costs of a treatment. A TOPM was developed to predict areas where vision restoration is most likely to occur in patients with visual field defects after brain damage. The deficit can be reduced by a vision stimulation treatment but the extent of restoration is highly variable between patients and not homogeneously distributed in the visual field. The TOPM is based on features which were constructed from visual field diagnostic charts and incorporates *a priori* knowledge from the domain of visual field plasticity. The features were evaluated to examine the statistical association between the features and the treatment outcome. Further features are derived from a review of vision restoration literature. A data pool with diagnostic charts from (n = 52) patients with visual field damage was used for model learning. The core of the TOPM is the Self-Organizing-Map (SOM) which is applied in data exploration and prediction. For model evaluation, a method is proposed to calculate the graph of the Receiver-Operating-Characteristic (ROC) for the SOM. Issues relevant for machine learning in medicine are discussed such as appropriateness to the patient sample and the clinical relevance of the prediction model.

Zusammenfassung

Vorhersagemodelle werden in der Medizin verwendet, um das potentiell erreichbare Behandlungsergebnis mit dem Aufwand der Behandlung zu vergleichen. In dieser Arbeit wird ein Therapie-Vorhersagemodell (TOPM) für die visuelle Restorationstherapie vorgestellt. Das Modell prognostiziert den allgemeinen Therapieerfolg und ermöglicht die Lokalisierung visueller Areale mit hohem und geringem Verbesserungspotential. Die Trainingsdaten des Modells wurden aus Diagnostikkarten von Patienten (n = 52) mit visuellen Wahrnehmungsstörungen erzeugt. Dazu wurden Merkmale unter Berücksichtigung von *a priori* Wissen und der aktuellen Plastizitätsliteratur aus den Diagnostikergebnissen extrahiert. Die Selbst-Organisierende-Karte (SOM), welche zur Datenexploration und auch für die Vorhersage genutzt werden kann, ist der Kern des TOPM. Für die TOPM-Evaluierung wird ein Algorithmus vorgeschlagen, mit dem die Funktionskurve der 'Receiver-Operating-Characteristic' (ROC) erstellt werden kann. Die Arbeit diskutiert außerdem die Eignung des Modells für Patienten und relevante Aspekte zur Anwendung von maschinellen Lernverfahren in der Medizin.

Glossary

	Component plane	The visual illustration of SOM unit positions in Φ is used for data exploration
ρ	Correlation	Measure of the statistical linear association between two variables
e	Eccentricity	Angular distance between the central fixation point and the visual stimulus
	Feature	An attribute of the diagnostic chart
Φ	Feature space	The vector space spanned by the features
FCT	Fixation catch trial	Assesses the fixation stability during HRP
	Foveal vision	Center of vision with highest acuity which is surrounded by peripheral vision
	Hemianopia	Visual field defect with visual damage to the left or right hemifield in both eyes
	Hemifield	The visual field, separated into the left and right hemifield enclosing the fixation point
HRP	High Resolution Perimetry	Diagnostic procedure to assess the visual damage by detection of visual stimuli
μ	Mean	Statistical measure describing the central location in a set of observations
φ	Polar angle	Horizontal angle of the visual stimulus

ROC	Receiver Operating Characteristic	A measure evaluating the prediction error of the TOPM
SOM	Self-Organizing-Map	A non-linear algorithm which is used for data exploration, classification and prediction
	Spot	The smallest unit in the visual field describing the response to a visual stimulus
S.E.	Standard Error	The spread around the statistical mean
	Scotoma	Visual field area without perception of visual stimuli
TOPM	Treatment outcome prediction model	Predicts the treatment outcome for subjects based on a history of treated patients
Γ	Test set	Set of samples used for SOM testing
	Training	The SOM model learning phase
Λ	Training set	Set of samples used for SOM training
	Treatment outcome	Improvement (hot spot) or lack thereof (cold spot) after VRT
VRT	Vision Restoration Therapy	An experimental treatment reducing visual field damage by intense visual stimulation

Contents

Chapter 1

Introduction

In medical applications, treatments were administered by the physician to support the patient's recovery. If the treatment is time consuming and if it does not have the same result for all patients with a specific disease, it is desirable to use a prediction model telling whether the regimen is helpful or not. A simple manual approach to built such a prognostic system would require the following steps:

1. Collect as much data as possible about patients: the therapeutic regimen and the result after treatment.

2. To make the prognosis for a new patient P, just look into the set of collected samples and find a patient P' with identical properties (with respect to age, the history of the patient's diseases, the regimen and some other criteria which were collected).

3. Return the treatment outcome from P' as prognosis for P.

This approach lacks some very important aspects. If the database is huge (containing more than 10,000 cases), it is tedious to go through the database manually and finding the correct match. Furthermore, each patient is unique; no two patients are identical. This example shows that computers are helpful to support the physician who wants to make a prognosis for a new patient. Data Mining tools were designed to efficiently access large databases and can be used to find cases which best match the properties of the patient P'.

1.1 The Aim of Data Mining Algorithms

The discipline of Data Mining is described by Fayyad et al. [1996] as: 'Data Mining' is the step in the KDD[1] process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of patterns (or models) over the data'. Another but equally valid definition was formulated by Thuraisingham [1998]: 'Data mining is the process of posing various queries and extracting useful information patterns from large quantities of data'.

To give a short insight into the diversity of Data Mining algorithms, several concepts are briefly summarized below. For further reading, recent text books give a detailed introduction to the field of Data Mining [Han and Kamber, 2006, Hand et al., 2001, Hastie et al., 2002, Mitchell, 1997, Peter, 1986, Wi et al., 2007, Witten and Frank, 1999].

1.1.1 Properties of Data Mining Problems

The domain or problem to which data mining is applied is different with respect to the available type of data, the representation and quality of data. This section shows a categorization of Data Mining problems. The description of these problems is supported by a formal model in which R is the result of the Data Mining process. M is the model which is learned by the Data Mining algorithm. The model input are n samples $S_1, .., S_n$ which are the training set. The samples are comprised of d observation variables $Var_1, ..., Var_d$.

Known and Unknown Target Variable

A typical goal of data mining is to extract a more abstract representation from a huge number of samples such that the domain expert can interpret the result of the data mining algorithm to gain new hypothesis. An example for this could be that the learning set consists of diagnostic data with the

---

[1]KDD = 'Knowledge discovery in databases' aims on mapping raw data into a more compact and more useful representation

observed variables, blood pressure (Var_1), body temperature (Var_2) and pulse rate (Var_3). M is the model which extracts general rules R from the n patient samples. The k-th sample is described by the observations $S_k = [Var_1, Var_2, Var_3]$ and the formal description of the problem is:

$$R = M(S_1, S_2, S_3, ..., S_n) \tag{1.1}$$

An interpretation of R could have the following form: "In many patients, the pulse raises and decreases together with the blood pressure, but the body temperature is independent of blood pressure and the pulse rate". Another form of R could represent clusters in the learning set, for example: "A large group of patients has a large blood pressure, a low temperature and a high pulse rate but another and different group of patients has a high blood pressure, high temperature and a normal pulse rate".

The presented problem is an unsupervised learning problem because the samples in the learning set are described by observations without additional information and the output is a more abstract description of the learning samples. In supervised learning, the learning samples are described additionally by a class variable (also termed target variable) and the k-th sample is now described by $S_k = [Var_1, ..., Var_d, class]$. The formal description is identical to 1.1 but, R can be used now for predicting the class of a novel sample S' which is comprised of the observations $S' = [Var_1, ..., Var_d]$ according to:

$$class = R(S') \tag{1.2}$$

For example, R could be used to predict whether patients have a disease or not, on the basis of the diagnostic test results (blood pressure, body temperature and pulse rate). If the target variable represents a class, the variable is nominal or ordinal scaled. Another type of algorithms are required to find a model M for which the target variable is interval scaled.

Framework Requirements

In the example presented above, the model M was learned during model training to approximate the learning samples. This means that any model is an acceptable result that can sufficiently approximate the learning samples irrespective of the specific model parameters. In some problems the high degree of freedom of resulting models is payed by a high computational effort. In other problems, it is important that the model parameters satisfy further requirements, for example, if machine learning is used to learn a motor controller, the controller should never overload the engine. Thus, background knowledge D (also termed a priori knowledge or domain knowledge) limits the number of acceptable models to only those, which satisfy the domain knowledge or the framework requirements:

$$R_D = M(S_1, S_2, S_3, ..., S_n, D) \tag{1.3}$$

Problems with Uncertain Data

Data in real world problems are often gathered by sensor systems (e.g. medical diagnostic systems). Those systems are not without error. The measurements often contain a source of noise. If the uncertainty of the model input (N) is known to some degree or can be extracted from inconsistent learning samples, the uncertainty (U) which is contained in the results can be estimated and equ. 1.1 has now the following form :

$$[R, U] = M(S_1 + N, S_2 + N, S_3 + N, ..., S_n + N) \tag{1.4}$$

Vague Data Representation

The description of data is often made with precise numbers (e.g. "the temperature today is 30.5 °"). To humans, this precise description is not very intuitive and linguistic terms are preferably used (e.g. "it is hot today"). Thus, precise numbers are substituted by intervals such that 'hot' is equivalent to a subject specific interval (e.g. between 27 ° and 35 °). In problems, where the model M shall be learned to substitute human decision making, the

variables Var_i are substituted by linguistic terms which are more compatible to the human language.

1.1.2 Properties of Data Mining Algorithms

Lazy versus Eager Learning

In lazy learning (also termed 'Case based reasoning' or 'Prototype learning') the data processing of the learning samples is delayed until a novel sample is classified (e.g. k-means [Jain and Dubes, 1988]). In eager learning, the process of model learning is separate from the process of applying the model. The data approximation model is created at first on the basis of a training set and then used later for predicting the target variable (or class) of novel samples (e.g. Association learning [Agrawal et al., 1993] for ordinal target variables or linear regression [Karlsson, 2007] for metric target variables).

As eager learning algorithms create an approximation model, a priori knowledge is often used to specify the bounds of the resulting model (e.g. Prolog-EGB [Wusteman, 1992]). This example uses a strongly defined domain theory in the form of logical expressions. However, in most real world problems, an explicit domain theory is either hard to define or often, the domain theory is not an input source but, contrastingly, the final goal.

Interpretability of the Model

In some problems, not only the resulting class or pattern (R) are interesting for the user but so is the model M which generates R. Algorithms are then preferred which generate an interpretable model (e.g. in the form of rules in the decision tree algorithm C4.5 [Quinlan, 1993]). Contrarily, other algorithms produce a black box model which is not interpretable (e.g. in neural networks [Nürnberger et al., 2002]. In these algorithms, often no interpretation is found for weights between neurons or the structure of the network [Schwarzer et al., 2000]).

Quantification of Model Uncertainty

An appropriate method that considers the uncertainty in model generation are probabilistic models. These introduce the concept of belief. It is interpreted as confidence of the correctness of assumptions and is based on the Bayesian theory. Applications of Bayesian theory are graphical models [Borgelt and Kruse, 1997, 2002, Kruse and Borgelt, 1998] which represent the observations as nodes, and the conditional dependence between the observations as links between the nodes. The advantage of such a graphical representation is the propagation of belief. The confidence about observations which are not known is computed from observations from known variables. Probabilistic models consider uncertainty, and their output is the predicted class together with a confidence that the class is correct. Other algorithms, for example, the k-Nearest-Neighbor classifier in its unmodified form, does not support the quantification of uncertainty in its prediction results.

Algorithms Processing Vagueness

One purpose of 'Fuzzy Theory' is to provide linguistic terms for describing vague data. The user specifies the meaning of linguistic terms and the 'Fuzzy' algorithm uses this definition to compute a model. During model learning, the fuzzy algorithms manipulate the definition of the linguistic terms such that the linguistic model describes the raw data approximately and in a form which is pleasant for human interpretation. Concepts from Fuzzy Theory were also combined with algorithms from other domains, e.g. neuro-fuzzy nets [D. Nauck and Kruse, 1997] and fuzzy decision trees [Wang et al., 2004]. Another theory which also incorporates vagueness is Rough Sets [Skowron et al., 2002].

1.1.3 Requirements of a TOPM

In this Thesis, Data Mining and machine learning are primarily used to discover general rules from a medical dataset. The samples in the learning set are patient data and are comprised of observations derived from diagnostic

charts of the patient's visual field. The samples are supervised, hence, for each sample it is known whether the treatment was successfully or not (according to equ. 1.1). Variables are metric or ordinal scaled. The output variable is nominal scaled. How much the observations are confounded by noise is not known. After the relations between observed variables and the class label are approximated by a model, it is used to predict the treatment outcome of a novel patient (according to equ. 1.2). The domain knowledge is not an explicit input to the model learner (as in equ. 1.3) but is used to construct the variables from the diagnostic charts (see chapter 3). I have decided to use the Kohonen Self-Organizing-Map (in it's supervised mode) as model learner because the model results are interpretable and can be used for prediction and for data visualization as well (see chapter 4.3). The visualization of data and the interpretability of the model is explicitly relevant in clinical problems (see chapter 4.6). For prediction, a k-Nearest-Neighbor component [Tan et al., 2006] is added to the model. The prediction model supports the assessment of uncertainty about the correct prediction in the form of outlier detection and computation of the degree of concordance between sample and learning data (see chapter 4.2).

A more detailed introduction and concrete applications of machine learning in medicine are presented below.

1.2 Machine Learning in Medicine

In the medical domain and in other areas as well, where the 'curse of dimensionality' makes effective human interpretation of collected data practically impossible [Ubaudi, 2005], machine learning provides tools that help with the interpretation of high-dimensional and complex sets of data [Sajda, 2006]. In this Thesis, the fundamentals were developed to use machine learning for (i) prediction of diseases on the basis of available test results and (ii) prediction of the course of diseases based on test results and the history of patients (who have been in a similar situation and finished the treatment). The cardinal problem for both applications is to transform the knowledge of the domain expert into a machine-accessible form. As a major topic, the Thesis shows,

how diagnostic data are processed by algorithms from image processing and spatial data-mining such that the data complexity is reduced and how machine accessible features are constructed with the help of expert knowledge.

The benefits of diagnosis prediction models and TOPMs are briefly reviewed.

1.2.1 Prediction of Disease

Diagnosis prediction models are used to improve the detection, diagnosis, and therapeutic monitoring of a disease. Well designed prediction models can improve the efficiency and quality of medical care significantly [Magoulas and Prentza, 2001] because such predictions are independent of subjective judgement. They are used for novelty detection of diseases [Marsland and Buchan, 2004] and support the recognition of rare diagnostic types. In clinical practice, prediction models can reduce errors caused by habituation and human fatigue. They supplement physician judgement as a 'second opinion', based on the scheme of double reading; it was shown that the scheme of double reading increases the detection rate of breast cancer by 5 % - 15 % [Sajda, 2006]. However, double reading by two human experts is expensive. When a verified diagnosis prediction model is used as a second 'expert', the double reading scheme is implemented at significantly reduced costs. Furthermore, by using an objective diagnosis prediction model, the concordance of experts improves as examined in the field of reproductive medicine [Steeg et al., 2006]. Several experts who have received the same medical data varied largely in the proposed therapeutic regimens. Invasive treatments of childless women were less frequent when prediction models were consulted. For the visual domain, diagnosis prediction models were constructed to support the identification of visual disorders [Ruseckaite, 1999].

1.2.2 Prediction of Treatment Outcome

In contrast to predicting the type of diseases, treatment outcome prediction considers the time component [Abu-Hanna and Lucas, 2001]. Such a model helps to make a prediction of recovery on the basis of baseline test results

of the subject. This requires a history of pre and post treatment results such that appropriate rules are found describing the post-treatment results as a function of pre-treatment diagnostic data [Pepe, 2005]. In general, the TOPM is used by clinicians as a recommendation tool to select the treatment with the highest expected benefit to the patient. As in diagnosis prediction, the prediction model extends the knowledge of the physician by the knowledge base of the model and offers the benefit of a second opinion. To achieve this goal, methods of machine learning are combined with *a priori* knowledge of a specific disease to be able to forecast specific treatment effects. If successful, it will aid the clinician to estimate efficacy in an individual patient and perhaps provide new hints at improving efficacy. The construction of such models requires expensive studies for data collection, but, after completion and validation, the prediction model is applied to many patients [Alexopoulos et al., 1999].

1.3 The Clinical Application

1.3.1 The Visual System

For a proper understanding of the domain in which the present TOPM is used, some fundamental aspects of the visual system are presented in Fig. 1.1.

The visual system information pathway begins at the retina (I) where light is transformed to electrophysiological signals which are transmitted by the retinal ganglion cells (leaving the eye via the optic disc II) through the optic nerve (III) to other brain regions. The primary and most important stream of information travels through the thalamus, where information is relayed onto geniculate neurons (VI) which, in turn, send their axons (VII) directly to the primary visual cortex (VIII) located in the occipital lobe.

A neuron n in the primary visual cortex (also called 'V1') receives input from many receptors in the retina. This area of the visual field that is able to excite the neuron is called 'receptive field' of n (see Fig. 1.2). The primary visual cortex analyzes abstract features of visual images and scenes and works
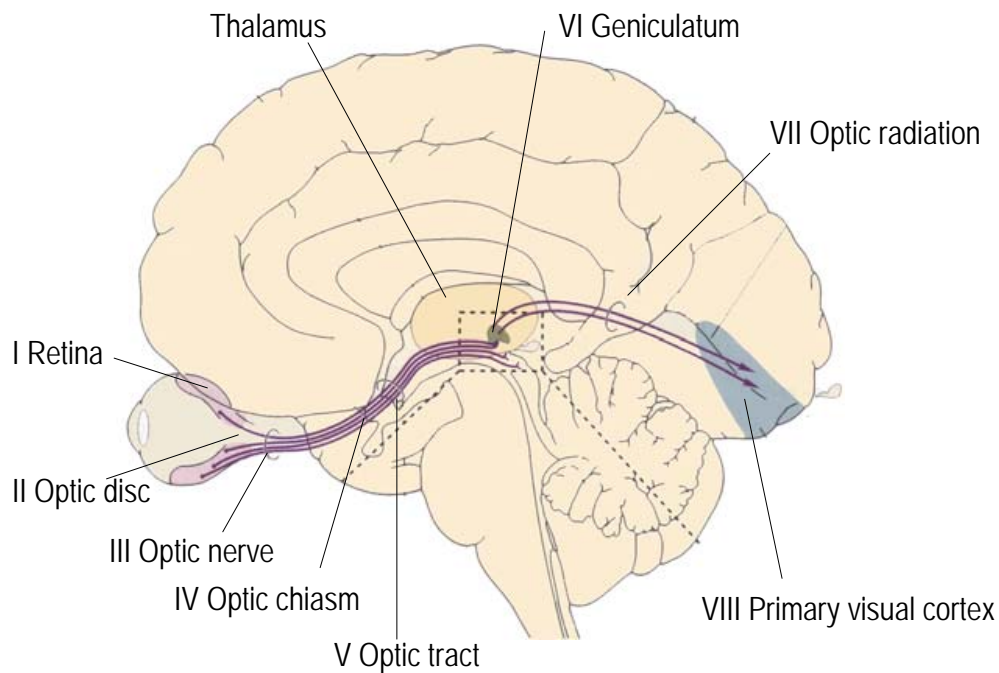
Figure 1.1: The visual information pathway. Adapted figure from [Kandel and Schartz, 1996].

in concert with many 'higher' brain regions with which it is reciprocally connected [Chalupa and Werner, 2004]. The visual system is the largest sensory processing system in the brain. About 30 % - 50 % of the cerebral cortex is committed to process visual information in comparison to only 8 % for the auditory and 7 % for the somatosensory system [Van Essen, 2004]. For this reason, brain damage due to stroke or head injury is often accompanied by loss of visual functions, leading to partial or total blindness.

1.3.2 Damage to the Visual System

Retinal (I) and/or optic nerve (III) lesion damage which is located in the early visual pathway (pre-chiasmatic lesions, e.g., after glaucoma or optic nerve trauma) cause different types of visual field loss in both eyes (heteronymous field defect) or monocular damage if only one eye or one optic nerve is affected (see Fig. 1.3, left). Damage to the mid-section of the visual pathway (e.g.,

Primary visual cortex

Cortical neuron

Lateral connection

Visual information pathway

Eccentricity

Retina

Receptive field
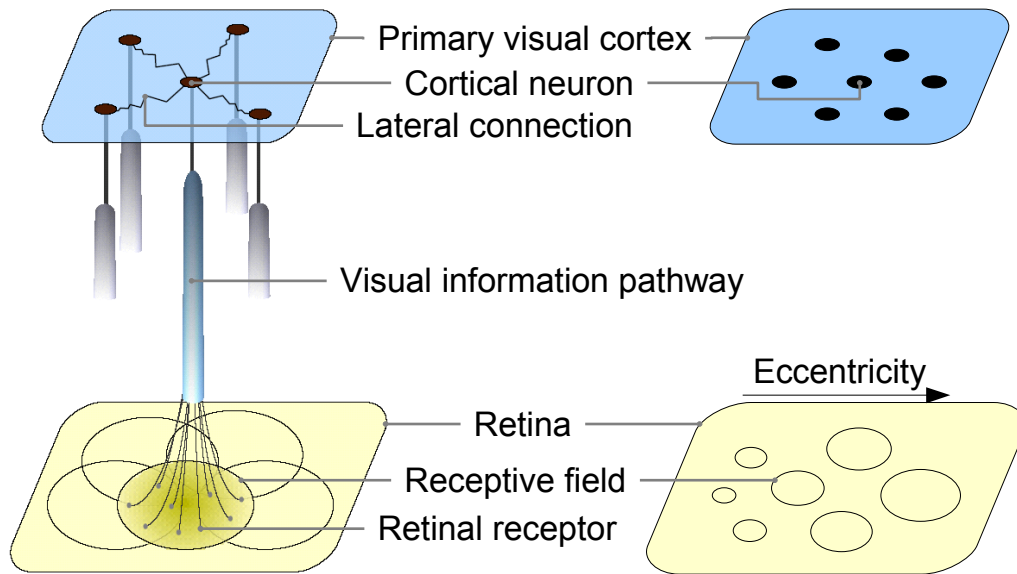
Retinal receptor

Figure 1.2: Neurons in the primary visual cortex are connected via the visual pathway with receptors located in the retina (left). All receptors which cause activation of the cortical neuron are termed 'receptive field'. Receptive fields are not of equal size but depend on the eccentricity from which the cortical neurons receive visual information (right). Cortical neurons from the periphery (high eccentricity) have larger receptive fields than cortical neurons from the foveal area (low eccentricity).

lesions of the optic chiasm, IV, or optic tract, V) typically produces visual field defects which are inverted in both eyes and are restricted to one side of the visual field, i.e., the deficit does not cross the zero vertical meridian (see Fig. 1.3, middle). Damage higher up in the visual pathway (e.g., optic radiation, VII or visual cortex, VIII) typically produces visual field loss which is more or less symmetrical in both eyes (homonymous field defect, see Fig. 1.3, right) and is usually limited to one side of the visual field [Purves et al., 2004]. Typically, the deficit does not cross the zero vertical meridian. Such 'post-chiasmatic' deficits are observed in about 20 % of stroke survivors [Gray et al., 1989] and are the most frequent diagnosis (75 %) of visual impairment [Zihl, 1995].

Optic nerve damage Chiasma damage Post-chiasmatic damage

L   R   L   R   L   R
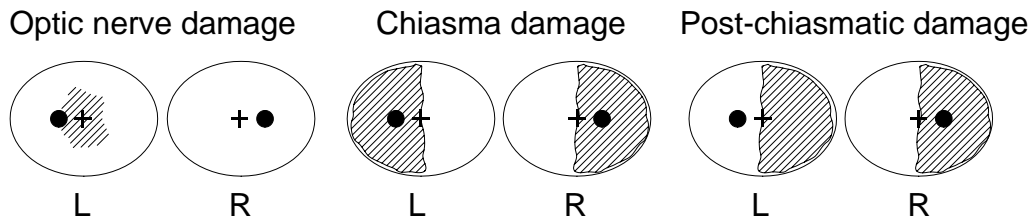
Figure 1.3: This graph displays three different types of visual defects. The solid black circle represents the blind spot in the left (L) and right (R) eye, the cross indicates the center of the visual field. Left: After damage to the left optic nerve (III), a heteronymous scotoma (hatched area) results in the visual field of the left eye. Middle: Damage to the optic chiasm (IV) causes heteronymous damage (not identical in both eyes) to the temporal visual field in each eye. Right: In cases of post-chiasmatic damage (damage of the optic radiation, VII and visual cortex, VIII) the damage is homonymous, i.e., the damage is identical in both eyes). The case presented here suffers a complete loss of the right visual field (termed hemianopia). In most subjects the visual field loss is not in the 'ideal' form as presented here but is a mixture of different lesion types with different extent of impairment.

1.3.3 Visual System Plasticity

Plasticity of the visual system is not just limited to early development (pre- and postnatal) but continues throughout life. It is the basis for spontaneous recovery in animals with visual system damage [Chino et al., 1995, Heinen and Skavenski, 1991, Sober et al., 1997]. Restoration (improvements of sensory functions) was observed after retinal lesions [Das and Gilbert, 1995, Giannikopoulos and Eysel, 2006] and deprivation (the visual input is suppressed) [Gilbert and Wiesel, 1992] or after nerve segregation (the nerves are chirurgically cut) [Buonomano and Merzenich, 1998]. Visual plasticity is observed in humans, as well. Patients with damage to the visual pathway typically experience visual field enlargements within the first few weeks or months after brain trauma or stroke [Zhang et al., 2006]. Visual system plasticity is also a feature of the intact brain [Pascual-Leone et al., 2005] and is believed to be the basis for the learning of perceptual tasks [Gilbert, 1994, Teyler et al., 2005, Watanabe et al., 2002]. In patients with damage to the post-chiasmatic visual system, repetitive visual stimulation (the treatment)

also helps to improve visual performance such as detection in perimetric tasks [Kasten et al., 1998a, Mueller et al., 2003, Sabel et al., 2004, Zihl and von Cramon, 1985] and improves vision parameters confirmed by electrophysiological recordings [Julkunen et al., 2003] or functional magnetic resonance imaging [Henriksson et al., 2007, Julkunen et al., 2006, Marshall et al., 2007, Pleger et al., 2003]. In this present Thesis, only subjects with different types of homonymous damage (the damage is identical in both eyes) are examined.

Whereas diagnosis prediction models are not considered further, the Thesis focuses on the development of a TOPM for VRT.

1.3.4 Vision Restoration Therapy

The principle of Vision Restoration Therapy (VRT) is visual stimulation of the visual field border in order to repetitively activate partially damaged regions located between areas of the intact and defective visual fields [Kasten et al., 1998a]. The treatment is typically carried out in a step-wise fashion. Regions of unreliable vision (also termed 'residual') are identified using the visual field diagnostic charts (see chapter 2). Depending on the type of stimulation, static or dynamic visual stimuli of increasing brightness thresholds are then presented in these treatment regions. Depending on the progress of the subject, the treatment area is adjusted monthly. When hemianopic subjects carry out this treatment for a period of six months (one hour per day), significant improvements (enlargements) of the visual field have been observed in 30 % to 70 % of all subjects [Kasten et al., 1998a, 1999, 2000, Mueller et al., 2003, Sabel and Kasten, 2000, Sabel et al., 2004].

1.4 Outline of this Thesis

After a brief overview on data mining and machine learning in medicine was presented in the first chapter, the second chapter describes the visual field diagnostics and the resulting diagnostic charts which have a fundamental role in the Thesis. The third chapter describes the features which are extracted from the diagnostic charts [Guenther et al., 2007a]. Some of these features

make use of a transformation model which allows bridging the visual field and the visual cortex, offering a variety of new features [Guenther et al., 2007b]. The newly designed features were evaluated in a retrospective study [Guenther et al., subm.c]. In the fourth chapter, I describe the treatment outcome prediction model (TOPM) which was built on top of the features [Guenther et al., subm.b]. The fifth chapter derives quality guidelines and examines eye movements during the diagnostic testing [Guenther et al., subm.a]. The Thesis closes with a summary and an outlook.

Chapter 2

Assessing the Visual Field

This chapter describes the background of assessing the visual field by diagnostic testing. The algorithm of the diagnostic test procedure is described and different types of diagnostic charts are presented. These form the basis of the subsequent chapters. Generally, locations in the diagnostic charts are measured in visual field degree and are localized by using two different coordinate systems (see Fig. 2.1). Hereby, the visual field is the complete visual area which is perceived when the eye fixates upon a static fixation point.

2.1 Visual Field Testing

Visual field diagnostic testing is needed to detect and localize defective areas and is required to determine the specific pattern, size and depth of visual field loss. The observed visual field defect (also called scotoma) is indicative of specifically where the damage may be located in the visual pathway. The analysis of the visual field by perimetric and campimetrical methods has therefore a high diagnostic value and is an important part of the neurological and ophthalmologic examination [Barton and Benatar, 2003]. Visual field charts drawn from such diagnostic testing are used as well to document the dynamics (deterioration or improvements) or stability of the visual field damage [Spry et al., 2001].

There are various types of visual field test methods which are designed to
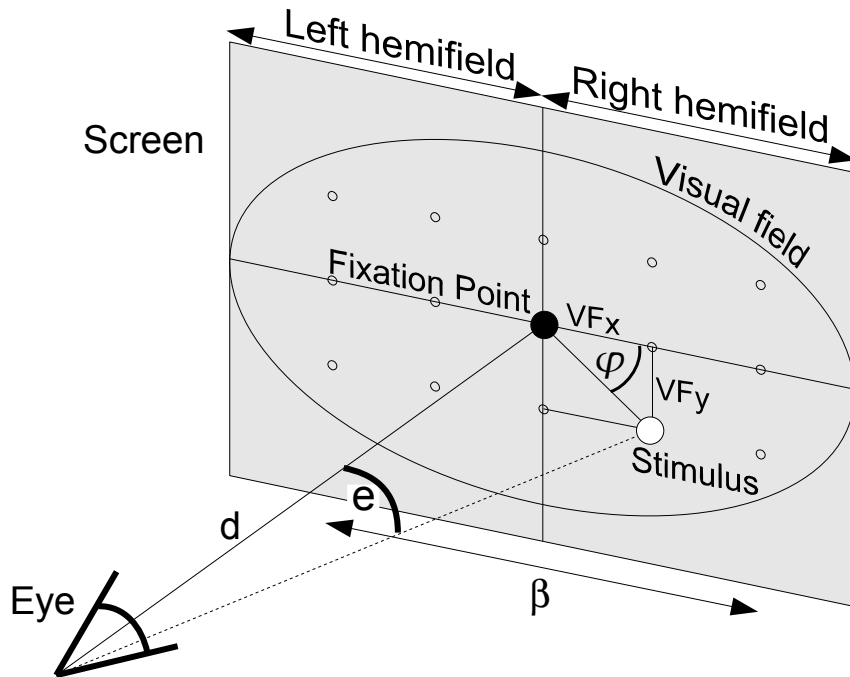
Figure 2.1: This figure illustrates the basic terms in visual field diagnostics. The eye fixates a specific point on the presentation screen (fixation point, black) which is located at the center between the left and right hemifield. A stimulus (white) is presented on the screen at one position of the rectangular test grid (open circles) which stimulates the respective region on the retina when the stimulus position is within the eye's visual field. The distance between fixation point and stimulus is measured in degree (eccentricity e). Together with the polar angle which measures the deviation from the horizontal axis (φ, also in degree), every position is located using this polar coordinate system. Alternatively, visual field positions are located in the Cartesian coordinate system by using the distance, measured in degree of visual angle, between stimulus and the horizontal axis (VF_x) and the vertical axis (VF_y). All measures are dependent on the eye-screen distance d. The angle β is the maximally examined horizontal degree after all grid positions were tested.

assess the visual field. Two main methods are kinetic and static perimetry [Barton and Benatar, 2003]. The kinetic perimetry involves the presentation of *moving* targets of a defined size and intensity. Each stimulus moves from the peripheral toward the central visual field until the subject acknowledges the stimulus by pressing a key on the keyboard. Although the automated static perimetry is most commonly used in the clinical practice for testing visual function, there are situations in which kinetic perimetry is preferable (e.g. in case of patients with visual field loss in the far periphery or in patients with short attention spans). There are some other perimetric procedures such as e.g. the short-wavelength-automated perimetry and motion perimetry, which assess specific aspects of visual function. These perimetric methods are still partly experimental tools [Barton and Benatar, 2003], but it has been shown that these tests are superior to standard perimetry e.g. in identifying glaucoma [Bayer and Erb, 2002, Johnson et al., 1993, Sample et al., 2000].

2.2 The High Resolution Perimetry

One static perimetric test procedure is the High Resolution Perimetry (HRP). It was developed to assess the visual field in high spatial resolution with fairly bright, high-contrast (super-threshold) stimuli. HRP is a perimetric test but uses a flat screen instead of a hemispheric presentation screen as in standard perimetry (see Fig. 2.2). Accordingly, the more specific term campimetry is often used which is derived from the latin word campus (meaning 'flat' or 'plane'). With HRP, visual field loss is investigated in much finer detail than it is possible with conventional perimetry.

In HRP, the stimuli are presented on a computer screen which is positioned at 30 cm to 40 cm in front of the subject while the head of the subject is stabilized with a chin rest (this reduces body and head movements during the diagnostic session). The subject responds by pressing a button whenever the presented test stimuli are perceived. During the complete diagnostic procedure of about 20 minutes, the subject is required to fixate upon a static point positioned usually at the center of the screen (the fixation spot). The validity and reliability of this diagnostic procedure was established elsewhere
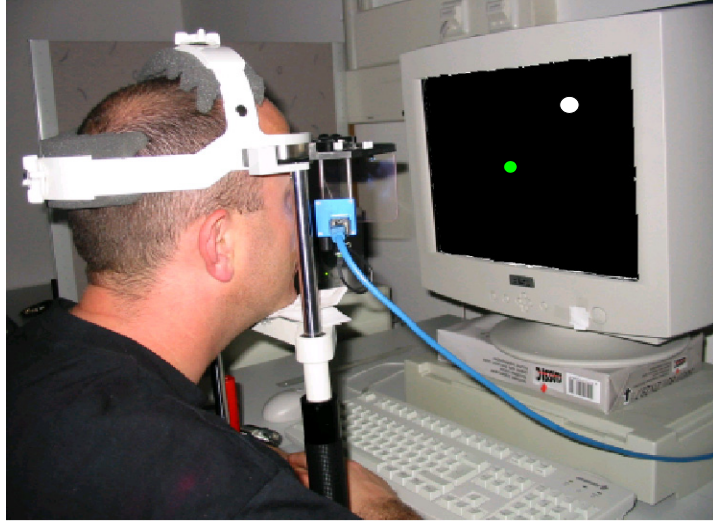
Figure 2.2: The presentation screen is located in front of the subject. A chin rest is used to reduce head and body movements which would otherwise interfere with the diagnostic accuracy. In this figure, an eye tracker is shown measuring eye movements during the diagnostic session. For the complete time of diagnostic testing, the subject is required to fixate upon the central fixation point (located in the center of the screen). Test white stimuli (enlarged for illustrative purposes) appear randomly at different screen locations. The subject is asked to respond to the stimuli by pressing a button.

[Kasten et al., 1997].

Technically, a visual field chart is generated in HRP by a repetitive 4-step algorithm on a dedicated personal computer:

1. Stimulus onset

   (a) The stimulus position is chosen randomly among all previously untested grid positions.

   (b) The number of test positions is determined by the chosen test grid; usually 25 x 19 positions (resulting in 474 stimuli without the fixation point), within a test range between 40 $°$ and 52 $°$ horizontally and 20 $°$ and 32 $°$ vertically, depending on the distance d between screen and subject.

   (c) The stimulus has a constant size (e.g. between 0.43 $°$ and 1.76 $°$).

(d) The presentation time (time from stimulus onset until the stimulus disappears) is constant (e.g. 150 ms).

(e) The stimulus luminance is well above-threshold (super-threshold) with a bright gray color (e.g. $48 \, \text{cd/m}^2$) on a dark gray background (e.g. $0.2 \, \text{cd/m}^2$).

2. Subject response

(a) The reaction time is measured from stimulus onset until the subject responds by pressing a button.

3. Interstimulus interval

(a) The interval between two consecutive stimulus onsets varies at random between the predefined maximal and minimal length (e.g. between 1000 ms and 2000 ms).

4. Fixation catch trial (FCT)

(a) It is required that the subject focuses on the fixation point during the entire diagnostic test session.

(b) The fixation point is colored (e.g. yellow with luminance: $124.2 \, \text{cd/m}^2$).

(c) Whenever the fixation point color alternates at random (e.g. to light green, but optimally without a detectable change in luminance) the subject responds by pressing a key.

(d) The number of presented FCTs is specified as the ratio of all presented test stimuli (e.g. 25 %).

(e) The size of the fixation stimulus is constant during the testing session (e.g. 0.43 °).

(f) The position of the fixation point is also fixed. The position can be chosen before the testing session to be anywhere on the presentation screen, in order to focus the visual field test on a specific visual hemifield or quadrant.

This algorithm is repeated until all positions in the test grid have been tested once. One test usually takes about 20 minutes to complete. The complete diagnostic test procedure is repeated twice. All results are super-positioned forming a diagnostic chart. Between two consecutive sessions, a break of at least 10 minutes is required to reduce the effect of fatigue.

2.3 Diagnostic Charts

The HRP diagnostic output is a topographic chart for each subject which illustrates schematically the visual field defects of the subject. Three single diagnostic tests assessed sequentially after each other are superimposed to obtain a chart that displays visual detections ($chart_{Detection}$) and reaction times ($chart_{ReactionTime}$) to the presented test stimuli. This computer-based perimetric procedure examines the central visual field in a range of $\pm$ 20 °.

2.3.1 Reaction Time Chart

The reaction time chart shows the response time of the subject to the pre-sented stimuli (see Fig. 2.3). The response times vary largely between, and also within subjects.

2.3.2 Detection Chart

The detection chart (see Fig. 2.4) shows the overall response to the presented test stimuli omitting the specific spot related response time.

Each of the 474 spots in the HRP chart was classified as to its detec-tion probability (*Residual Function*) indicating a specific functionality state: intact function (shown in white, the subject responded correctly to 3 out of 3 stimulus presentations), mild-defect (shown in light gray, the subject responded to 2 out of 3 stimulus presentations), moderate-defect (shown in dark gray, the subject responded to only one stimulus presentation) and absolute-defect (shown in black, the subject does not respond to any of the stimulus presentations). The term absolute-defect region is identical to the
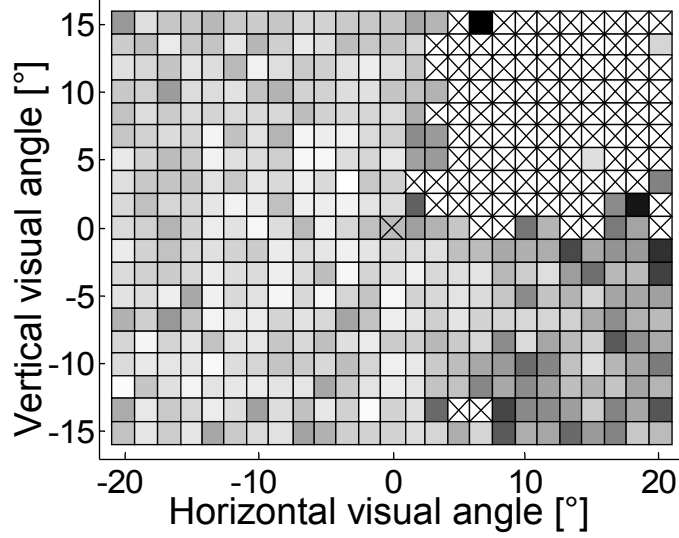
Figure 2.3: Left panel: HRP response time chart. Minimal response time is coded in white (≈ 350 ms), maximal response time is coded in black (≈ 800 ms). A cross represents detection misses. A tendency is apparent that the perception of above-threshold stimuli is slower (dark gray) in intact areas of the bottom right hemifield, in comparison to intact areas (light gray) of the left hemifield.

term 'blind region' or 'scotoma'. Intact regions are sub-sectors of the visual field which are unaffected by the injury having perfect or nearly perfect detection performance. The term 'areas of residual vision' (ARV) or 'relative defect region' is characterized by mild and moderate-defect vision. Here, the response time is slower [Sabel and Kasten, 2000] and the probability of being able to detect super-threshold stimuli is reduced [Kasten et al., 1999]. For the purpose of further analysis, the moderate-defect spots and absolute-defect spots were pooled to one class termed 'impaired spots'. Likewise, the term 'healthy spots' represents the category that includes mild-defect and intact spots. Intact, residual, and defect spots from the baseline diagnostic chart (at time t_0) are accessed for later use by the concept of a stencil matrix:

$$stencil_{Defect}(x,y) = \begin{cases} 1 & \text{iff chart}_{Detection}^{Baseline}(x,y) = 0 \\ 0 & else \end{cases} \qquad (2.1)$$
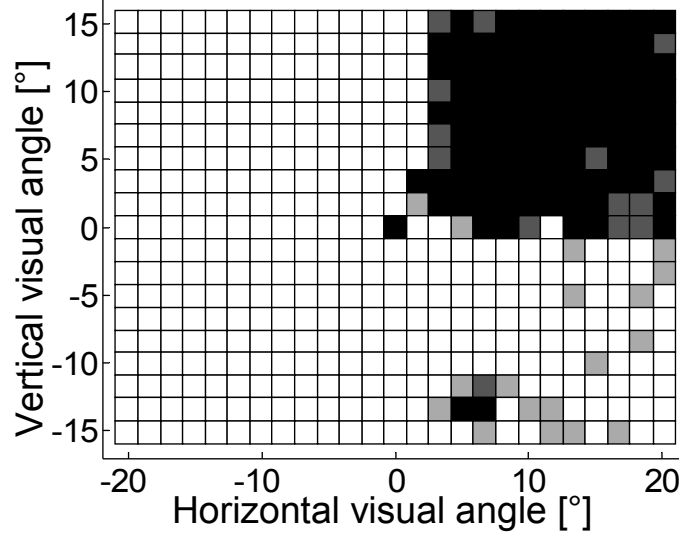
Figure 2.4: The HRP detection chart classifies a spot (which is the smallest atomic unit in the chart) according to the subject's functional state at the respective position (intact perception in white, absolute-defect perception in black, residual vision of unreliable perception (mild- and moderate-defect) in shades of gray).

$$stencil_{Intact}(x, y) = \begin{cases} 1 & \text{iff chart}_{Detection}^{Baseline}(x, y) = 1 \\ 0 & else \end{cases} \tag{2.2}$$

$$stencil_{Residual}(x, y) = \begin{cases} 1 & \text{iff chart}_{Detection}^{Baseline}(x, y) = \frac{1}{3} \cup \frac{2}{3} \\ 0 & else \end{cases} \tag{2.3}$$

2.3.3   Dynamic Chart

The development of individual perception performance as measured with HRP is interesting after brain or eye surgery, conventional treatment, in the process of spontaneous recovery, or as it is used here to assess the differences before (at t_0) and after VRT (at t_1). The dynamic chart visualizes the change over baseline after 6 months of VRT. Just as in the baseline chart, the dynamic chart is subdivided into identical squares. Each square was assigned a symbol after calculating the difference between the HRP charts

from a subject which were assessed at time t_0 and t_1. Depending on their respective 'change over baseline', dynamic spots are defined as follows: (i) spots which were impaired spots at t_0 and became healthy spots at t_1 are termed 'hot spots', and (ii) spots which were impaired spots at t_0 and were still impaired at t_1 are termed 'cold spots'. These detection differences (positions of improvement or without improvement) were calculated for each visual field spot, which then combined to form the 'dynamic chart' (see Fig. 2.5).
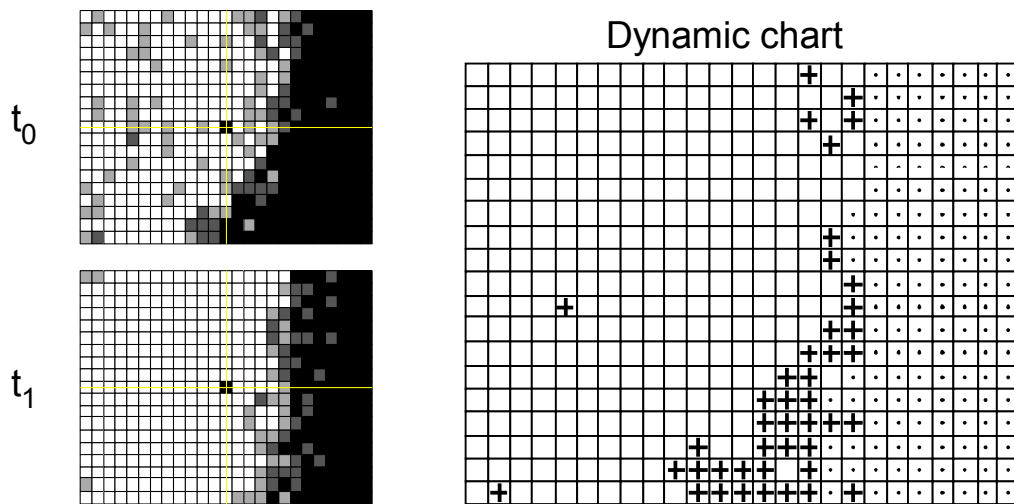
Figure 2.5: Left panel: The baseline diagnostic chart at time t_0 shows a defect-intact border oriented almost vertically and located in the right hemifield. In this example, the lower part of the defect-intact border moved rightwards as it is shown in the post-treatment chart at t_1. Right panel: The dynamics of the visual field were calculated for every position by subtracting the chart at t_1 from the chart at t_0. Accordingly, hot spots are located mostly in the lower right hemifield ('+') whereas cold spots ('.') are observed throughout the complete right hemifield.

2.3.4 False Positive Chart

Another chart is obtained from false positive responses which assesses the zealousness to respond. Random responses would result in a high number of false positives which are evenly distributed throughout the intact visual

field. Subject responses outside the valid time interval (e.g. 150 ms to 1000 ms) after the stimulus onset are termed 'false positives' and a topographic chart of false positives is obtained by associating the false positive response with the last stimulus onset position (see Fig. 2.6, right). Therefore, false positives are comprised of either random reactions unrelated to the stimulus presentations or reactions which are delayed. The number of false positives which occurred in the intact hemifield is used as an exclusion criterion for the subsequent analysis.

Figure 2.6: The chart of false positives (left) shows invalid responses outside the valid time interval (represented by a cross). Only few false positives were detected in the intact area of the right hemifield (the baseline detection c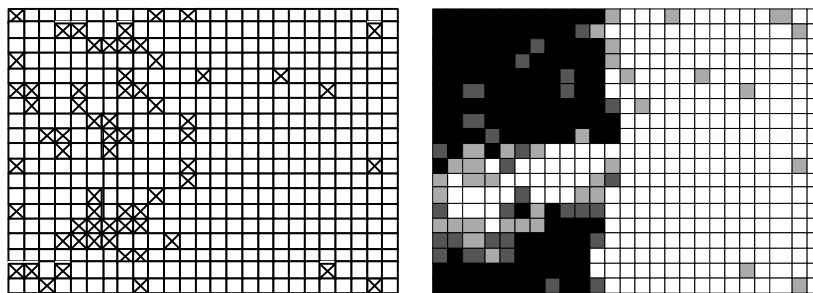hart is shown in the right panel) indicating that the number of false positives which are associated with the intact visual field are within normal variations.

2.4   Data Analysis Infrastructure

Technically, the diagnostic data is efficiently stored in a relational data-bases such that diagnostic results are retrieved for visual inspection or pre-processing. In the database, each stimulus is represented as triple (column, row, response time). The column and row positions are related to the location of the stimulus in the rectangular test grid (see Fig. 2.1). In order to use the prediction model, the raw diagnostic data (sets of triples) are extracted from the database and transfered to the feature computation module. An appropriate environment for feature computation is the Matlab programming environment (distributed by Mathworks). The visualization of diagnostic data is based on system variables which (i) define the minimal and maximal time interval in which a subject response is considered valid and (ii) define the color modus (color, shades of gray or symbolic) which is used to visualize the diagnostic results.

Program for Visualization of Diagnostic Data

In order to visualize the different aspects of the diagnostic results, several presentation modi are available in the graphical user interface (see Fig. 2.7) and process (interpret) the raw diagnostic data triplets as follows:

- The 'Reaction time chart' shows response times for all test stimuli (see Fig. 2.3). No additional interpretation of the triplets is required, except for invalid response times outside the valid time interval, which are omitted.

- The 'Detection chart' shows the strength of damage for all test stimuli (see Fig. 2.4) and is based on the reaction time. Stimulus responses within the valid time interval are considered valid (hits), whereas responses outside the valid time interval or no responses are considered invalid (misses). Hits and misses among three stimulus presentations per location define the *Residual Function* of the respective spot (see chapter 3.3.1).

- The 'False positive chart' shows false positives (response times outside the valid time interval) which are associated with the preceding stimulus location (see Fig. 2.6).

- The mode 'No interpretation' does not filter the raw data in any way (used for data exploration and experimentation).

- The 'Dynamic chart' shows the difference between two diagnostic charts (see Fig. 2.5).

- The 'Cortical projection' shows the diagnostic chart as a cortical projection either onto the left or the right primary visual cortex (see chapter 3.4).

The diagnostic results are shown in a rectangular grid which is identical to the grid used in the diagnostic testing. The position of a spot in the diagnostic chart visualization is determined by (i) the maximal horizontal angle of the diagnostic array (β, measured in degree, see also Fig. 2.1), (ii) by the relative horizontal (F_x) and vertical position (F_y) of the fixation spot with respect to β, (iii) the horizontal (row) and vertical position ($column$) of the stimulus in the test grid and (iv) by the total number of horizontal rows of the test grid (GX). The distance d between eye and screen is contained in the parameter β. The horizontal (VF_x) and vertical (VF_y) components of the visual field position (measured in degree) are calculated as follows:

$$VF_x = \quad row \cdot \frac{\beta}{GX} \quad -F_x \cdot \beta \qquad (2.4)$$

$$VF_y = \quad -column \cdot \frac{\beta}{GX} \quad +F_y \cdot \beta \qquad (2.5)$$

Chart Operators

For the analysis of diagnostic data and in the process of feature selection and construction, many hypothesis were formulated and were examined. A flexible data access is provided by the following operations:
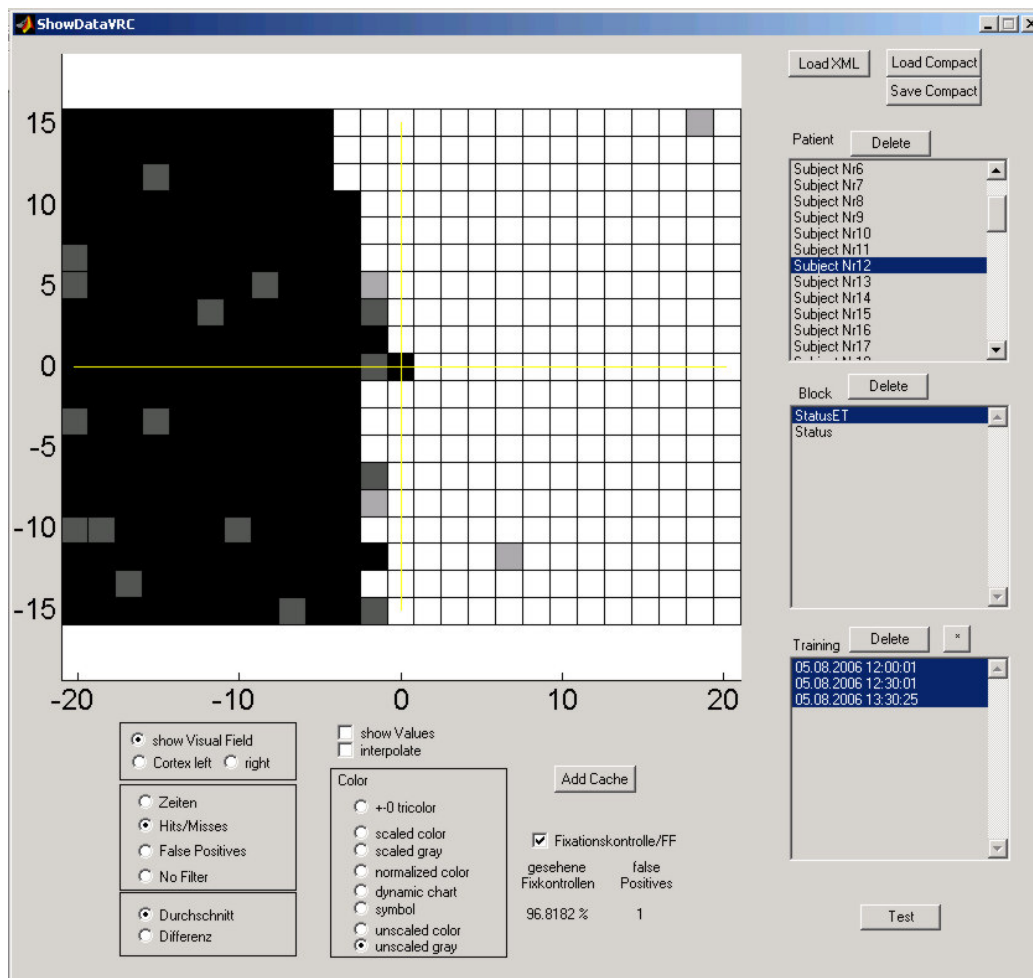
Figure 2.7: A Matlab control center for visualization, feature extraction and feature computation was developed.

- Combination and allocation of diagnostic charts from different subjects and different studies.

- Manipulation of diagnostic charts by using a temporary image buffer (accumulator).

- Calculation of mean and difference of two or more diagnostic charts.

- Batch processing of feature extraction algorithms applied on a set of diagnostic charts.

- Unary operators for calculation of fixation stability and the total number of false positives.

By using the infrastructure presented above, the next chapter introduces features which were extracted and computed from the diagnostic charts.

Chapter 3

Vision Restoration Features

The building of a TOPM requires appropriate features which are clearly associated with the treatment outcome. This chapter introduces features which were constructed from the diagnostic reaction time and detection charts (see Fig. 2.3 and 2.4) and are motivated from neuroscientific literature. As a major part of the Thesis, those features were implemented by considering the anatomy of the visual cortex, results of cortical lesion experiments and expert knowledge in VRT. The predictive value of three newly constructed features was evaluated in a retrospective study. Yet additional features were derived from the specific literature on VRT. The examined features are grouped into global features, local features and features which are not related to diagnostic charts. Whereas global and local features are used in the TOPM, the last group is not considered because the association with the treatment outcome was not clearly shown. A short introduction to feature selection precedes the description of the features.

3.1 Introduction

Feature selection is the task of deciding which features are relevant with respect to the target classification (in this case: treatment outcome). The suitability of features which are used for the TOPM determines the performance of the prediction model even more than the choice of the classification

algorithm. Generally, choosing an internal model is far less important than choosing the right features [Ubaudi, 2005]. This is important especially for the medical domain if automatically collected data sets contain many 'don't care' attributes [Alexopoulos et al., 1999]. Therefore, the feature selection should be evaluated by domain experts using the available numerical information [Lucas and Abu-Hanna, 1999].

For evaluation, the features are usually ranked with respect to their treatment outcome association using a ranking measure. A large variety of such association measures are available for feature ranking, for example, information based measures (information gain), distance based measures (Euclidean distance) or dependence measures (Pearson correlation). A detailed analysis of measures is available elsewhere [Guyon and Elisseeff, 2003, Liu and Motoda, 1998]. Some evaluation measures require a discrete scale (e.g. mutual information gain) such that continuous scale features are at first transformed into discrete features which divides the continuous scale into non-intersecting intervals (also termed 'bins').

Features with low ranking positions are then either removed in filter [Liu and Motoda, 1998] or wrapper models [John et al., 1994] or their impact is reduced by assigning low weights in Bayesian classifiers and neural networks [Langley, 1994]. If the ranking considers single features (univariate) instead of sets of features (multivariate), interdependent relationships between features are lost and the highest ranked features are prone to be mutually redundant if they correlate significantly [Jong et al., 2004]. In wrapper models, feature ranking considers the features in combination with the classifier. The learning process and the evaluation of the classification is repeated with different sets of features. The feature ranking is now determined by the classification error. The advantage of wrapper methods is their multivariate variable evaluation such that interdependent relations are considered. Disadvantages are higher computation time because the learning process is repeated for each iteration. Because the best feature set is determined with respect to the selected classifier, wrapper models can often outperform filter models [John et al., 1994]. In this Thesis, features were evaluated semi-automatically (see chapter 3.6) or an evaluation is available

in the VRT literature. In the next sections, the features from the domain of visual system plasticity are presented.

3.2   Global Features

Global features address information which is derived from the diagnostic chart as an atomic entity and are not specific to spots of the diagnostic charts. Therefore, all spots from the same diagnostic charts share the respective global feature value.

3.2.1   Border Diffuseness

The *Border Diffuseness* is the ratio of residual spots among all spots in the border area of the diagnostic chart. The border area is 5 ° wide and located between the defect and the intact area (see Fig. 3.1, top right). The border is defined as 'diffuse' if many residual spots lay inside the border area and 'sharp' if almost no residual spots are inside the visual scotoma border.

The scotoma border is located with the following algorithm in which the detection chart is interpreted as a gray scale image (see Fig. 2.4):

1. All residual positions were transformed into intact positions (pixels in gray become white)

2. The median filter (size: 3x3 pixel) was used to reduce noise in the diagnostic chart such that isolated spots and fibers are eliminated.

3. The orthogonal border orientation and border position is obtained by convolution with a 3x3 horizontal $\frac{1}{4} \cdot \left(\begin{smallmatrix} 1 & 0 & -1 \\ 4 & 0 & -4 \\ 1 & 0 & -1 \end{smallmatrix} \right)$ and vertical $\frac{1}{4} \cdot \left(\begin{smallmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ -1 & -4 & -1 \end{smallmatrix} \right)$ gradient kernel which approximates the first derivative of the gray scale distribution. The border orientation is used for subsequent analysis (see Fig. 3.12).

4. The vector field thus obtained (each vector has a horizontal and a vertical component) was smoothed by a 3x3 Gaussian kernel:
$\frac{1}{100} \cdot \left(\begin{smallmatrix} 7 & 12 & 7 \\ 12 & 2 & 12 \\ 7 & 12 & 7 \end{smallmatrix} \right)$
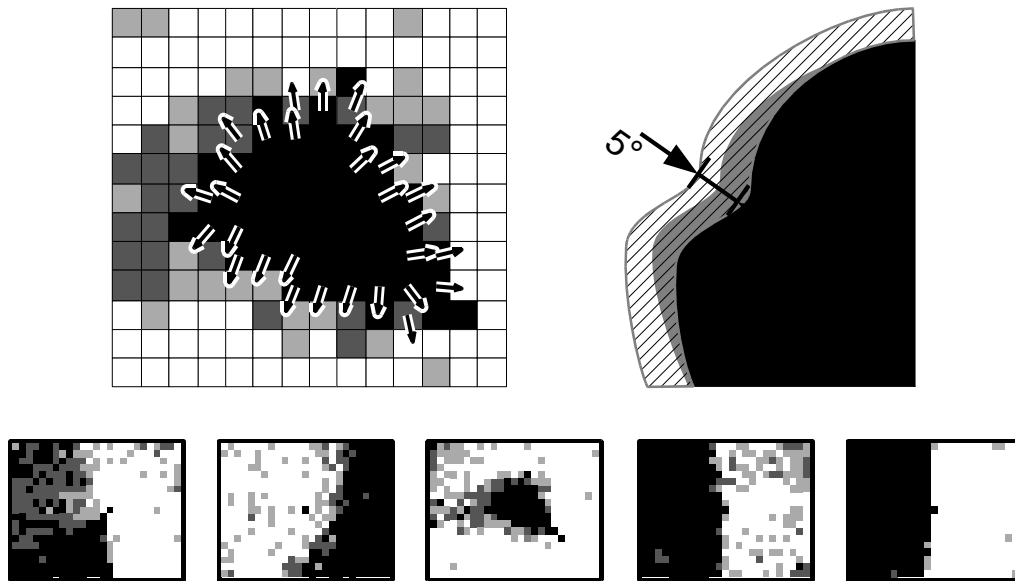
Figure 3.1: Top panels: Detection of border position and border orientation. After the visual field border is detected (shown by white arrows, left) the border area is defined as the area within a distance of 5 ° viewing angle from the border line into the direction of the intact visual field (hatched area, right). Bottom panels: Original examples of different visual diagnostic charts ordered according to *Border Diffuseness* from diffuse (left) to sharp border (right).

  5. The amplitude of vectors in absolute-defect areas which do not border residual or intact areas were set to zero.

   As the result of this algorithm, the intact-defect border as well as the orientation of the border is detected (vectors are oriented orthogonally to the border in direction to the intact visual field (see arrows in Fig. 3.1, top left).

   Border Diffuseness is specific to the exact location of the visual field border, irrespective to the occurrence of residual spots in the remaining chart. An interesting and repetitive observation in the VRT literature is that the improvements after treatment are most pronounced in subjects with many residual positions in the diagnostics chart around the intact-defect border [Sabel and Kasten, 2000]. The extent to which the visual field border is diffuse

or sharp was often identified as a relevant parameter in vision restoration [Kasten et al., 1998b, 1999, Mueller et al., 2003, Sabel, 1999, Sabel and Kasten, 2000].

3.2.2 Reaction Time

The feature *Reaction Time* is derived from the reaction time chart (see Fig. 2.3) and is the average response time of intact spots (such that the reaction time is averaged among the reaction times from three stimulus presentations at the same location) as follows:

$$
\text{Reaction Time} := \frac{\sum_{y=1}^{GY} \sum_{x=1}^{GX} \left(chart_{ReactionTime}^{Baseline}(x,y) \cdot stencil_{Intact}(x,y) \right)}{\sum_{y=1}^{GY} \sum_{x=1}^{GX} stencil_{Intact}(x,y)}
$$

(3.1)

The reaction time varies largely between and also within subjects and the VRT literature reports inconsistent observations. Although one study found no correlation between reaction times and treatment outcome [Poggel, 2002], results of another study (assessing a different group of subjects) showed a clear association: subjects with faster reaction time experienced a more pronounced visual field increase [Mueller et al., 2002].

3.2.3 Conformity to Hemianopia and Quadrantanopia

This feature expresses the degree of similarity of the visual field of a given patient with an 'ideal' hemianopia and quadrantanopia as displayed in Fig. 3.2. These features are based on combined values from the ratio of defect spots ($ratio_{def}^i$) and the homogeneity ($noise^i$) of the i-th quadrant:

$$
degree_{def}^i := ratio_{def}^i \cdot (1 - noise^i)
$$

(3.2)

The homogeneity is obtained by convolution of the quadrant with the 2x2 kernel $\left(\begin{smallmatrix} -1 & 1 \\ 1 & -1 \end{smallmatrix} \right)$. The convolution result is scaled such that the $noise^i$ value is close to '0' if the respective area is homogeneous and close to '1' if the area is noisy.

The conformity to hemianopia and quadrantanopia is then calculated from combining the $degree_{def}^i$ values from all four quadrants such that, for example, quadrantanopia of the second quadrant is represented by many defect spots in the second quadrant with little noise and many intact spots with little noise in the first, third and fourth quadrant (see Fig. 3.2, left).
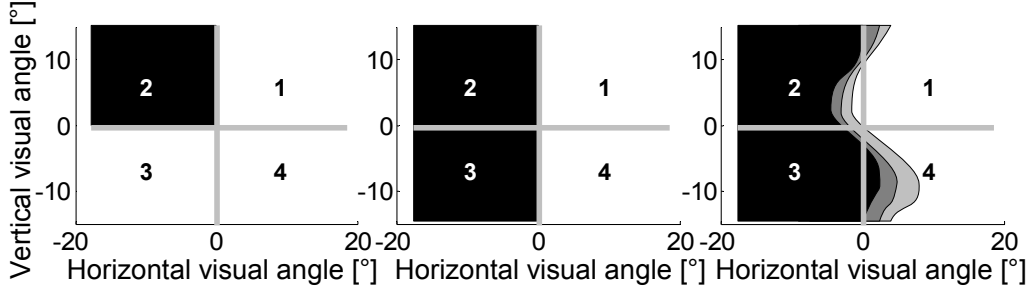

Figure 3.2: Prototypic ideal visual field defects. Homonymous quadrant-anopia (left) in the upper left quadrant and hemianopia (middle) of the left hemifield. In many subjects, the border zone is not as sharp as in the left and middle panel but shows variable performance (i.e. broken up with residual spots forming a diffuse residual border) as in the right panel.

Hemianopia :=
$$max \begin{pmatrix} degree_{def}^1 \cdot degree_{def}^4 \cdot (1 - degree_{def}^2) \cdot (1 - degree_{def}^3) \\ degree_{def}^2 \cdot degree_{def}^3 \cdot (1 - degree_{def}^1) \cdot (1 - degree_{def}^4) \end{pmatrix} \quad (3.3)$$

Quadrantanopia :=
$$max \begin{pmatrix} degree_{def}^1 \cdot (1 - degree_{def}^2) \cdot (1 - degree_{def}^3) \cdot (1 - degree_{def}^4) \\ degree_{def}^2 \cdot (1 - degree_{def}^1) \cdot (1 - degree_{def}^3) \cdot (1 - degree_{def}^4) \\ degree_{def}^3 \cdot (1 - degree_{def}^2) \cdot (1 - degree_{def}^1) \cdot (1 - degree_{def}^4) \\ degree_{def}^4 \cdot (1 - degree_{def}^2) \cdot (1 - degree_{def}^3) \cdot (1 - degree_{def}^1) \end{pmatrix}$$
$$(3.4)$$

The 'ideal' hemianopia where one hemifield is completely intact and the other hemifield is completely blind, is very rare. It is usually caused by total anatomical damage of the visual system in one hemisphere or results

from a complete cut of the optic radiation [Barton and Benatar, 2003]. Both features are considered in the TOPM because the clinical experience showed that restoration is unlikely if a visual field defect is complete and close to the ideal form.

3.2.4 Size of Residual and Defect Area

Both features are equal to the total number of occurrences of the respective spot type in the detection chart:

$$\text{Defect Area} := \sum_{y=1}^{GY} \sum_{x=1}^{GX} stencil_{Defect}(x,y) \tag{3.5}$$

$$\text{Residual Area} := \sum_{y=1}^{GY} \sum_{x=1}^{GX} stencil_{Residual}(x,y) \tag{3.6}$$

The defective sector of the visual field is usually rather compact and covers large areas of the visual field, for example, up to half of the visual field in subjects with hemianopia and about a quarter in quadrantanopia (see Fig. 3.2). Areas of residual vision are usually located at or near the border region separating the seeing field from the blind region although residual vision can also occur in compact areas within the defect or intact visual field.

The literature shows ambivalent results with respect to the predictive value of the features. In one study the size of the visual field defect did not correlate with the outcome [Sabel et al., 2004]. This was in contrast to other studies where the subject group with small defect areas had improved significantly more then the group of subjects with large defect areas [Kasten and Sabel, 1995, Poggel et al., subm.]. It was repeatedly observed that visual field improvements are located in areas where the detection of stimuli is unreliable (termed residual areas, indicated by shades of gray (see Fig. 2.3) [Kasten et al., 1999, Sabel and Kasten, 2000]. Specifically, the size of the residual area before treatment correlated significantly with the treatment outcome [Mueller et al., 2002, Poggel et al., 2004, subm.]. However, the position of residual spots are somewhat unstable because residual functions

are prone to fluctuations caused by fatigue, stimulus saliency and in general, by attention during the diagnostic testing [Poggel et al., 2006].

3.3   Local Features

Local features are directly related to the spot itself and extend the above introduced global features. By using local features the prediction can be applied specifically for each spot in the diagnostic chart.

3.3.1   Residual Function

Residual Function measures the strength of damage to a spot. It is a dichotomous measure and its value is either 0 or $\frac{1}{3}$ indicating that 0 or 1 of 3 presented stimuli were detected by the subject (shown in dark gray and black in Fig. 2.4). This feature is interpreted as the probability of stimulus detection at the respective spot position. Note that *Residual Function* is not a measure of individual electrophysiological activity of single neurons but it is the overall residual function of the corresponding region probed by the stimulus detection. It shows how often the subject has responded to a total of 3 stimulus presentations (detection rate) per specific test position. Spots with a *Residual Function* at baseline above $\frac{1}{3}$ ($\frac{2}{3}$ or 1) are excluded from the analysis and prediction because those spots are classified as 'healthy' (shown in light gray and white in Fig. 2.4). The *Residual Function* of the spot is equal to the detection chart value at the respective grid position (x, y):

$$ResidualFunction(x,y) := \begin{cases} 0 & \text{iff } \text{chart}_{Detection}^{Baseline}(x,y) = 0 \\ \frac{1}{3} & \text{iff } \text{chart}_{Detection}^{Baseline}(x,y) = \frac{1}{3} \end{cases} \qquad (3.7)$$

A high predictive value of this feature was observed in the retrospective analysis (see section 3.6.3): a significantly higher *Residual Function* ($\frac{1}{3}$) was observed in restored spots in comparison to not restored spots where the *Residual Function* was mostly 0. It was hypothesized that such residual

areas are the preferred location for restoration because subjects with hemianopia normally focus their attention on intact visual field regions and tend to neglect small and confusing, imprecise perceptions which originate from residual areas [Kasten et al., 2000]. During the treatment, residual spots could become intact whereas blind spots tend not to improve. This feature has probably the highest predictive weight among all other features.

3.3.2 Neighborhood Activity

The spatial neighborhood around a hot or cold spot includes all spots within a 5 ° radius of visual viewing angle centered around the target spot. This limit of 5 ° (see Fig. 3.3) was chosen for technical reasons: firstly, larger values would possibly average out local effects and secondly, much smaller values are below the spatial resolution of the diagnostic test grid. *Neighborhood Activity* is a measure of how much *Residual Function* is present in the entire neighborhood. This parameter is the average *Residual Function* of all stimulus positions in the entire 5 ° radius, i.e., the functional state of the entire neighborhood. Therefore, a value of '1' is assigned to a neighborhood where all spots are intact and '0' is assigned if all spots in the neighborhood are absolute-defect. The calculation include neither the *Residual Function* of the spot in the neighborhood center, which was addressed above, nor does it include spots of the opposite (contralateral) hemifield.

According to the retrospective analysis in section 3.6.3, improved spots had significantly greater *Neighborhood Activity* in their surround (radius 5 °) than non-improved spots. On average, more than one third of all spots within the baseline-chart neighborhood around hot spots were found to be healthy in comparison to only 1/10 in the neighborhood around cold spots (for details see chapter 3.6). In other words, if the surround is more intact and therefore active, restoration is more likely. Lateral horizontal connections in visual cortex could explain such a center-surround relationship between neurons in the primary visual cortex (see Fig. 3.4, right). Referring to the literature, the reorganization of cortical tissue was simulated in a theoretical study, and practical results also suggest that perilesion excitability is a critical factor for

reorganization after cortical stroke [Goodall et al., 1997, Pettet and Gilbert, 1992].

3.3.3 Neighborhood Homogeneity

The homogeneity is the standard deviation obtained when calculating *Neighborhood Activity* scores. By using this feature together with *Neighborhood Activity*, neighborhoods which are originally composed of more than 20 single values (their respective *Residual Function*) are adequately represented by only 2 values. Furthermore, the defect, residual and intact neighborhoods are well separated from each other by using *Neighborhood Activity* and *Neighborhood Homogeneity* together; Assuming the neighborhood has a medium *Neighborhood Activity* with an average of 0.5, a high standard deviation would result if 50 % of all visual field positions are defective and 50 % are intact (for examples see Fig. 3.5). Another example of a possible neighborhood could have the same *Neighborhood Activity* of 0.5, but a low standard deviation would result, if all positions in the neighborhood are residual, in a rather homogeneous manner.

3.3.4 Visual Field Position

Hot and cold spots are not uniformly distributed in the visual field but are a function of the location of the visual scotoma (see Fig. 3.6) and therefore considered in the TOPM. The VRT literature showed that the vertical eccentricity was positively correlated with the amount of the defect border shift (which is comparable to the definition of the treatment outcome used in the Thesis) [Mueller et al., 2002, Poggel, 2002, Poggel et al., subm.]. Additionally, it was hypothesized that the recovery follows the cortical magnification, hence that improvement occurs more frequently in higher eccentricities [Poggel et al., 2004].
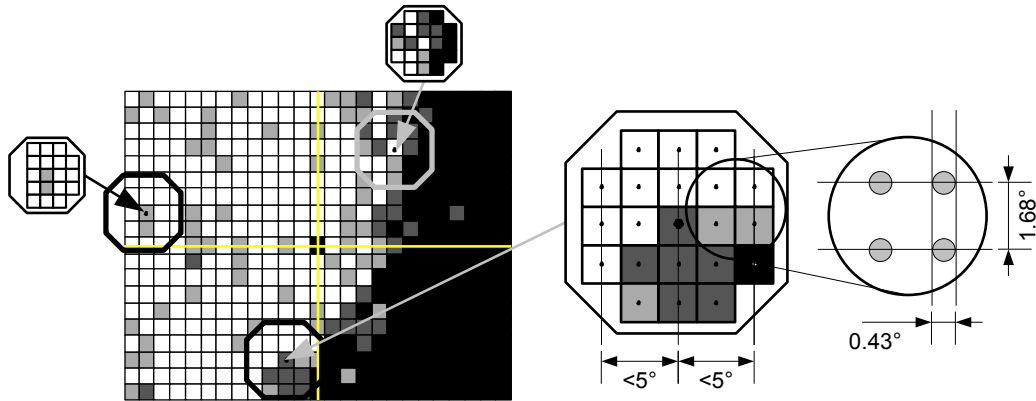
Figure 3.3: This graph illustrates the extraction of three spatial neighborhoods from the baseline diagnostic chart (left). The spatial neighborhood (right) has a radius of 5 ° in visual field coordinates (stimulus size: 0.43 °, inter-stimulus space: 1.68 °).
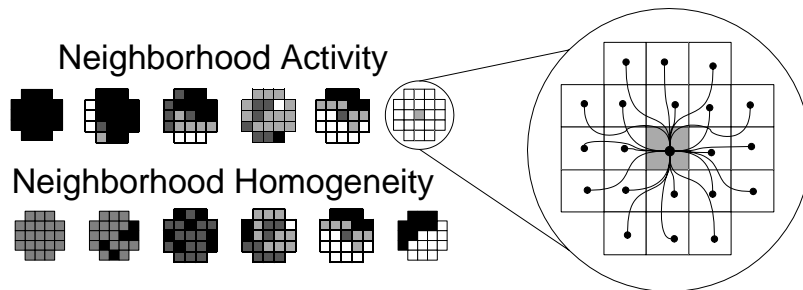
Figure 3.4: This graph shows examples of different neighborhoods and describes them in terms of *Neighborhood-Activity* (top row) and *Neighborhood-Homogeneity* (bottom row). They are ordered from low to high activity and from homogeneous to inhomogeneous. The activity and its standard deviation are based on the concept of the local spatial neighborhood (see also Fig. 3.3) which is inspired by the connectivity between neurons (right) where intrinsic horizontal connections (indicated by branches) exist between neurons. These horizontal connections are anatomically and physiologically well described [Stettler et al., 2002] and their function is to integrate visual information over some cortical distance.
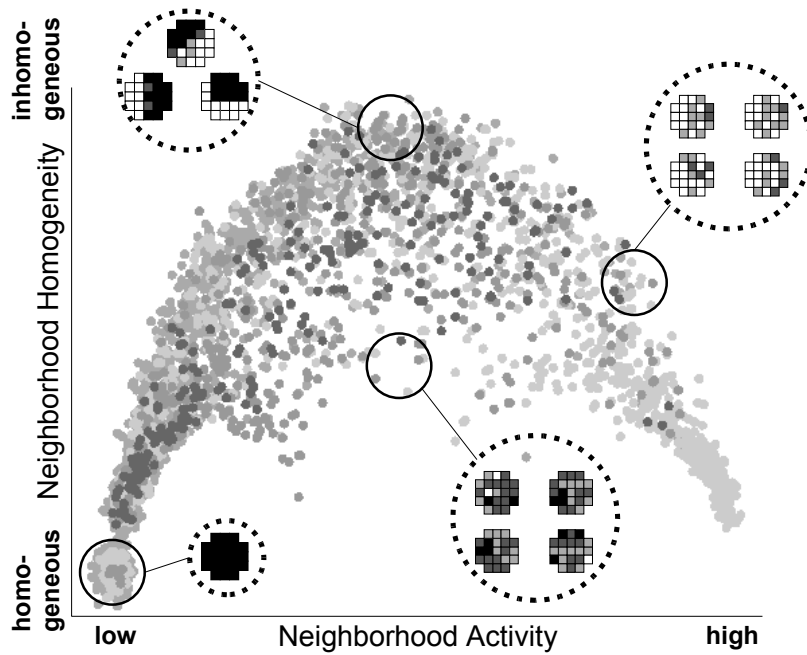
Figure 3.5: By using both, the features *Neighborhood Activity* and *Neighborhood Homogeneity*, the information of the spatial neighborhood is compactly represented. Neighborhood examples and their respective position in the chart are shown in the insets.
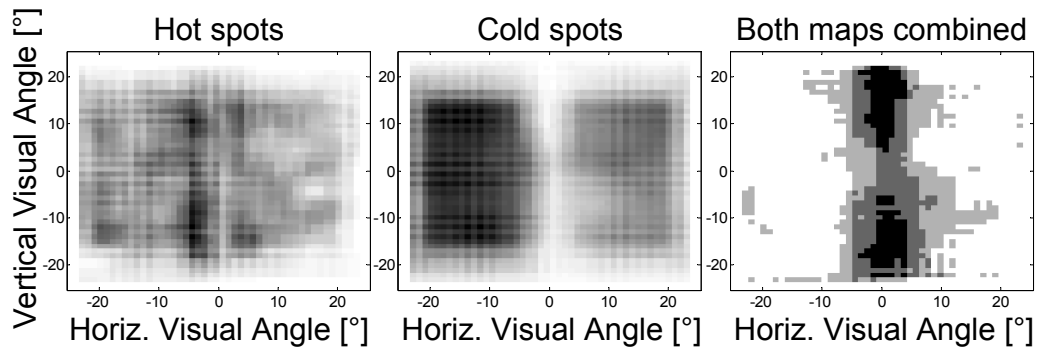
Figure 3.6: The spatial distribution of hot (left) and cold (middle) spots based on 52 subjects with post-genicular damage (black represents highest frequency and white represents no occurrence of hot or cold spots at the respective position). If both charts are combined (right), the effect of the lesion location is eliminated (right, black color: the relative occurrence of hot spots is higher than of cold spots, dark gray: 33 % hot spots and 66 % cold spots, light gray: 20 % hot spots and 80 % cold spots, white: less than 20 % hot spots).

3.3.5 Distance to Scotoma

The distance between a spot and the scotoma border (see Fig. 3.7) is estimated in cortical coordinates after transformation of visual field coordinates (see section 3.4). Cortical coordinates were preferred to visual field coordinates because functionally relevant distances are linear in the visual cortex but non-linear in the visual field due to cortical magnification (see Fig. 1.2).

The Euclidean distance was measured in cortical coordinates between the spot and its next proximal border position and the analysis was restricted to only one hemifield. Possible inter-hemispheric interactions in the visual system [Marks and Hellige, 1999] were not considered in the analysis.
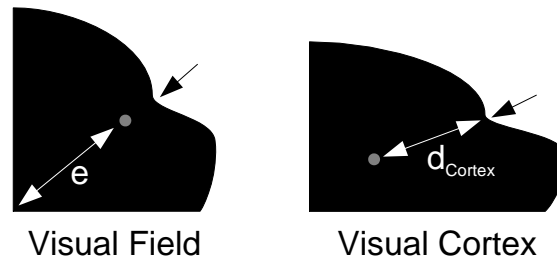

Visual Field   **Visual Cortex**

Figure 3.7: The feature *Distance to Scotoma* measures the distance in cortical coordinates (d_{Cortex}) between a spot (in gray) and the nearest position of the defect-intact border (shown by a black arrow) after the visual field is projected onto the visual cortex by using a projection model (see chapter 3.4). The cortical distance depends at most on the eccentricity e at which the spot is located in the visual field.

According to the feature evaluation study (see section 3.6.3) it was observed that *Distance to Scotoma* is significantly (negatively) correlated with the treatment outcome, (i.e. restoration of vision is greatest in spots close to the visual field border). This finding is compatible with observations in animal studies where reorganization was observed to be restricted to the immediate perilesion area [Schweigart and Eysel, 2002]. A spatially limited amount of information integration was also observed in experiments which examined the size of perceptive fields [Levitt and Lund, 2002].

The spatial limitation of vision restoration is probably attributed to the architecture of interactions (see also Fig. 1.2) in the cortical neuronal net-

work which change their connection weights to control the activity that is transfered to other neurons [Das and Gilbert, 1995]. Cortical correlates for spatial propagation of cortical activation are well-known. They include intrinsic horizontal connections in the visual cortex which connect neurons with overlapping receptive fields and neurons with non-overlapping receptive fields of the same preference [Sincich and Blasdel, 2001]; such as columns of similar ocular dominance [Malach et al., 1993] or preferred orientation [Gilbert and Wiesel, 1989]. Another possible mediator of residual activation are connections [Shmuel et al., 2005] which provide feedback loops from higher (extrastriate) cortical areas. It was further proposed that these kinds of cortical interconnections which are under permanent reorganization [Stettler et al., 2006] play a fundamental role not only in normal visual perception, but also in cortical plasticity during perceptual learning [Gilbert et al., 2000] and after cortical damage [Das and Gilbert, 1995].

An interface to the cortical transformation function was developed such that cortical interactions between different areas of the visual field can be estimated.

3.4 Visuo-cortical Coordinate Transformation

The size or position of damaged areas in the visual cortex is of interest to several applications of visual research, such as calibration of diagnostic systems, correlation analysis or, in this case, data mining of visual field charts. By using a coordinate projection model [Balasubramanian et al., 2002] the size and position of the cortical scotoma is estimated and cortical distances between points in the visual space are computed. The visual field charts are projected onto the primary visual cortex which is located in the occipital cortex (see Fig 3.8).

Such a model shows which cortical position P_C is active if an appropriate stimulus is presented in visual space coordinates at position P_{VS}. The projection function $f : P_{VS} \longrightarrow P_C$ is non-linear and incorporates the cortical magnification factor in which foveal eccentricities are overrepresented in the visual cortex (see also Fig. 1.2). The projection of the high dimensional
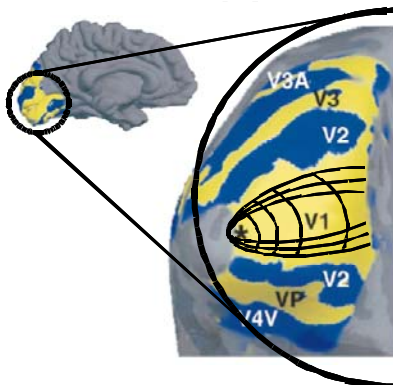
Figure 3.8: Medial view of the visual cortex (in circle). Borders between the different visual processing areas (V1, V2, V3) are shown in the inflated cortex (enlarged). Iso-polar coordinate lines of the projection model are superimposed and originate from the foveal confluent (marked with a star). Adapted figure with kind permission from [Tootell et al., 1998].

visual feature space (color, size, orientation, texture, velocity, position of visual stimuli) onto the two-dimensional visual cortex layer is mathematically equivalent to a compression with loss of accuracy. Due to this dimension reduction, the cortical projection is not homogeneous [Gilbert and Wiesel, 1989, Malach et al., 1993].

The interface is written in Matlab and computes cortical coordinates based on the horizontal and vertical degree of the visual field position and - if available - the size of the individual primary visual cortex.

The coordinate-transformation function $f : P_{VS} \longrightarrow P_C$ transforms (Cartesian or polar) visual space coordinates into cortical coordinates (see Fig. 3.9).

All positions of the visual field are completely but non-uniformly represented in different parts of the primary visual cortex. The transformation considers the eccentricity e taking the cortical magnification factor into account. Cortical magnification is defined as the size of the cortical area which is activated when a visual stimulus of size 1 ° is presented in the visual space at eccentricity e [Dougherty et al., 2003]. The cortical magnification function overrepresents foveal eccentricities reflecting the density of foveal receptors which is higher in the central retina than in the periphery [Duncan
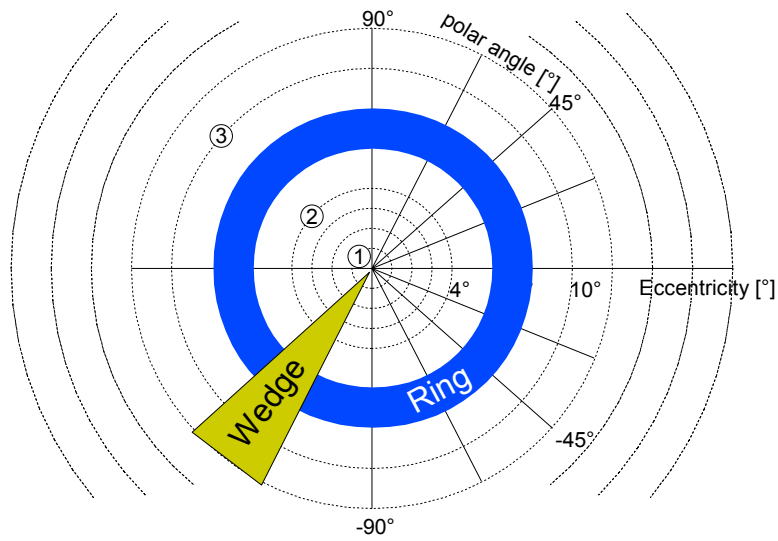
(a) Visual Space
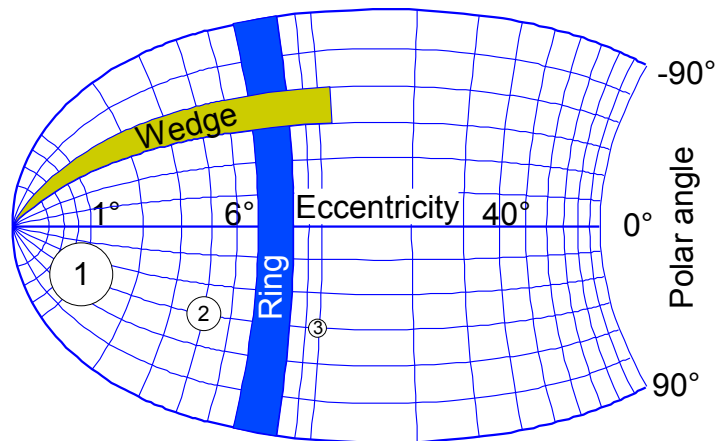
(b) Visual Cortex (right)

Figure 3.9: A: The visual space in polar coordinates. The space is divided into rings of increasing eccentricity and wedges of increasing polar angles representing any point in the visual space. Three different types of visual stimuli are shown: A ring-shaped stimulus (2 ° extent in eccentricity). A wedge-shaped stimulus (15 ° extent in polar angle). Three point-like stimuli of equal size (1 °) at different eccentricities (1-3). B: The inflated primary visual cortex. The visual space (in eccentricity and polar angle) is transformed into cortical space by the complex logarithm. The cortical over-representation of the parafoveal circular stimulus (labeled with 1) is clearly visible in comparison to the peripheral stimulus (labeled with 3). The complex logarithm maps ring-shaped stimuli into vertical lines, and wedge-shaped stimuli into horizontal lines.
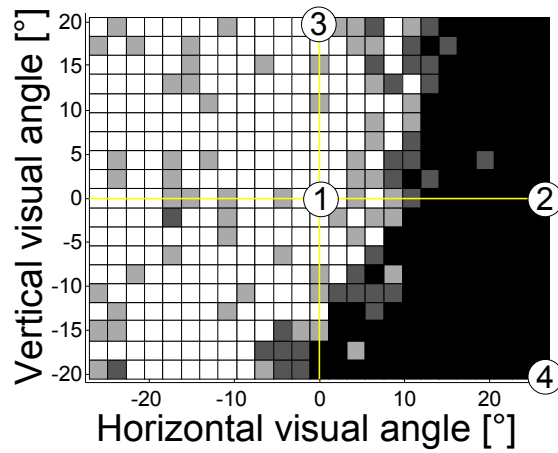
(a) Diagnostic Chart
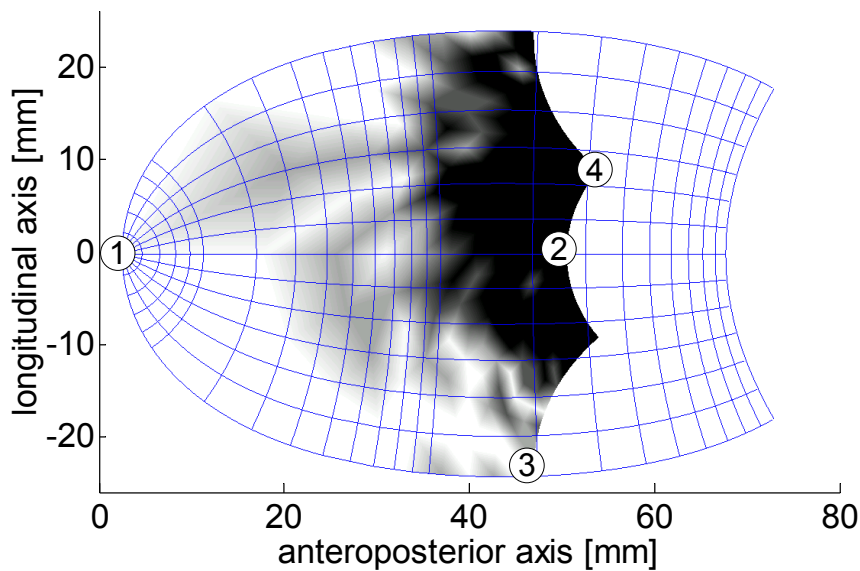
(b) Visual Cortex (left)

Figure 3.10: A: Diagnostic chart with 474 stimuli. A large scotoma is visible (black) in the right hemifield. The left hemifield is almost completely intact (white). Four landmarks are highlighted (1: foveal vision, 2: the most horizontal extent, 3: the most vertical extent, 4: position with highest eccentricity). B: Cortical projection of the visual field chart from (a). To demonstrate the coordinate transformation, landmarks from (a) are shown at correspondent positions. The para-foveal area (eccentricities $\leq$ 10 °) is represented in the left part of the figure around the origin of all iso-polar coordinate lines.

and Boynton, 2003] (see also the projection of equally sized circular stimuli in Fig. 3.9).

By considering the polar angle φ, the transformation incorporates projection differences between the horizontal and vertical meridian [Van Essen et al., 1984]. Results from retinotopic measurements [Schwartz, 1977] suggest that the complex logarithm describes sufficiently the transformation from visual space coordinates (eccentricity and polar angle in degree) into visual cortex coordinates (longitudinal and anteroposterior axis in cortical millimeter). The coordinate transformation is based on a 'wedge-dipol model' of the primate visual cortex topography of the owl monkey [Balasubramanian et al., 2002] and represents any point of the visual hemifield by a complex variable:

$$z = e \cdot exp(\varphi i) \tag{3.8}$$

where e represents eccentricity and φ the polar angle of the visual field position P_{VS}.

To obtain the 'wedge-dipole model' for the primary visual cortex, the complex variable is included into a dipole model (the constants $a = 1.5$ and $b = 170$ determine the positions of two monopoles):

$$w = log((z + a)/(z + b)) \tag{3.9}$$

The original model [Balasubramanian et al., 2002] incorporates the V1, V2, and V3 cortical area but here, only the V1 area is considered. The visual cortex coordinates are obtained from the real and imaginary parts of the complex variable w. The horizontal axis is shifted by $k_1 = 4.73$ mm so that foveal eccentricities are identical with the origin and absolute cortical coordinates were obtained by scaling the model with k_2 to the desired total V1 surface area:

$$P_{CX} = k_2 \cdot (real(w) + k_1) \tag{3.10}$$

$$P_{CY} = k_2 \cdot imag(w) \tag{3.11}$$

The transformation is isotropic (locally invariant to the direction in the visual space) and conformal (visual space angles are preserved in the visual cortex). Because no individual cortex sizes were available for the subjects in the evaluation study, the cortex size was set to the average V1 size of 2399 mm^2 from 29 examined cortices [Andrews et al., 1997].

Visual field charts, which are transformed into visual cortex charts by using the transformation model as described here (see Fig. 3.10), contain a model error caused by the differences in size and shape of the individual cortices. Whereas the error, which is introduced by differences of the cortical shape is relatively small, a substantial error is introduced by the inter-individual size variations (standard deviation: 494 mm^2, n = 29 in [Andrews et al., 1997]) of the cortical lobe. The individual minimal and maximal size of the visual cortex may vary by a factor of 2.5 [Dougherty et al., 2003].

3.5 Inappropriate Features

The following features were derived from the VRT literature but were not considered for the TOPM because the association between the observed features and the treatment outcome is ambivalent, contradictory, not applicable or not existent. However, those features are briefly described.

3.5.1 Subject Age

The VRT literature does not agree on a clear association between *Subject Age* and treatment outcome if different subject samples are examined: one study reported only small non-significant influences of *Subject Age* with the treatment outcome [Mueller et al., 2003] whereas another study reported that the group with age < 50 years showed a better outcome than the subject group with age > 50 years [Kasten and Sabel, 1995]. The opposite was reported elsewhere where subject age was positively correlated with the treatment outcome [Poggel et al., subm.].

3.5.2 Lesion Location and Lesion Age

Lesion Age is not associated with the treatment outcome [Poggel et al., subm., Sabel et al., 2004]. A group of subjects who began the treatment within the first year after their lesion showed no different outcome to the group of subjects who started later than one year after their lesion [Kasten and Sabel, 1995]. Considering the anatomical location and type of visual system damage, the group with damage to the early visual system (optic nerve damage) achieved stronger improvements than the group with post-chiasmatic damage [Kasten et al., 2000]. No difference was found among subjects with post-chiasmatic damage between stroke patients and patients with tumor surgery, skull fracture from accident, hemorrhage [Kasten and Sabel, 1995] or between stroke patients and patients with vascular malformations/aneurysm surgery or between the affected hemisphere [Poggel et al., subm.]. Because all subjects in the database have similar parameters (post-chiasmatic damage, location of treatment, time since lesion), the type of lesion, location and time is not considered in the TOPM.

3.5.3 Treatment Schedule and Location of Treatment

The TOPM training set was derived with all subject's treatment duration being similar (about 180 hours) and the treatment location was selected equally according to standard operating procedures. Therefore, *Treatment Schedule* and *Location of Treatment* were not considered in the TOPM. According to the VRT literature, treatment results were not stable if the treatment lasted for less than 30 hours [Kasten and Sabel, 1995]. Another study showed that neither total duration nor total number of treatment session did influence the treatment outcome significantly [Poggel et al., subm.].

However, the treatment outcome is specific to the location of stimulation. No improvement was observed in the untreated visual field sectors in contrast to the treated sectors [Kasten et al., 1998a]. Likewise, more pronounced improvement was observed in the quadrant where an attentional cuing was used during treatment in comparison to the quadrant without attentional cuing [Poggel et al., 2004].

3.5.4   Diagnostic Quality Parameters

The subject groups are pre-selected to include only subjects with low fixation instability. Therefore, the parameter *Fixation Stability* is not considered in the TOPM. The criteria of sufficient fixation stability is examined in chapter 5 because eye movements can severely affect the validity of the visual field diagnostics. In the pre-selected group of subjects, *Fixation Stability* was not correlated significantly with the treatment outcome (see also Table 3.1). These findings are similar to those reported elsewhere [Kasten et al., 2006]. Furthermore, neither alertness nor vigilance showed any value for predicting the treatment outcome [Poggel et al., subm.].

3.6   Retrospective Feature Evaluation

In this section, the results of an evaluation study are presented. Methods are described to evaluate the local features *Residual Function*, *Neighborhood Activity* and *Distance to Scotoma* with respect to their predictive value. The evaluation is necessary because the features were created recently and have never been tested.

A hemianopic subject sample (n = 23) was selected from the archives for this retrospective evaluation. By comparing the diagnostic charts before vs. after treatment, two groups of visual field spots with improvement (hot spots) and without improvement (cold spots) were identified at first. Then, feature averages were computed separately for the groups of hot and cold spots. If the new features are useful in prediction, the respective group averages μ_{hot} and μ_{cold} which are derived from the baseline diagnostic charts should differ significantly between the group of hot and cold spots.

All three features are motivated by a priori knowledge derived from the neuro-scientific literature:

- *Distance to Scotoma* is derived from observed restoration patterns after retinal lesions [Das and Gilbert, 1995, Giannikopoulos and Eysel, 2006], deprivation [Gilbert and Wiesel, 1992] and after nerve segregation [Buonomano and Merzenich, 1998].

- *Residual Function* is based on the observation that restoration is most pronounced if cortical damage is diffuse rather than absolute [Steward, 2006].

- *Neighborhood Activity* is derived from the observed perilesion activity which was identified as a critical factor in cortical reorganization [Goodall et al., 1997, Sober et al., 1997].

3.6.1   The Subject Sample

The subject sample which was used for feature evaluation and for building the TOPM determines the group of subjects whose treatment can be predicted by the TOPM. Because the prediction error may increase significantly if the TOPM is applied to subjects which are different from the subject sample in the evaluation and building phase, a detailed description of the appropriate subject sample and the methods of feature extraction is necessary.

The visual field of hemianopic subjects (n = 23, 16 male, 7 female) was assessed by HRP before and after carrying out VRT for 6 months. The data was retrospectively selected from a pool of subjects which have previously been treated in Magdeburg in the course of two independent studies: Study 1: [Mueller et al., 2003], Study 2: [Mast et al., 2006]. Although the diagnostic parameters differ between both studies from which the subjects were taken, no significant differences (two-tailored T-test) were found between the response times in the baseline ($\mu_1 = 450\,ms \pm 16.5\,ms$, $\mu_2 = 448 \pm 12.8\,ms$) or in the post-treatment diagnostic charts ($\mu_1 = 427\,ms \pm 11.0\,ms$, $\mu_2 = 415 \pm 10.0\,ms$) and therefore all subjects were pooled into one group. The subjects had an average age at start of training of 52 ± 13 (mean $\pm$ S.D.) years and their visual field defects were caused by post-chiasmatic damage of the visual system (age of lesion: 1.2 ± 1.3 years in Study 1 and 3.8 ± 1.1 years in Study 2). To be included in the present study the following inclusion criteria had to be met: Stroke or ischemia in the occipital cortex (A. cerebri posterior or A. media) as evidenced by medical documentation, and the availability of at least 3 independent visual field diagnostic tests at baseline and post-treatment, respectively. These three visual field tests were

then superimposed to generate a visual field chart. Exclusion criteria were (i) confounding eye disease, such as nystagmus or strabismus, (ii) more than 4 % false positive positions at baseline or post-training examination in HRP, and (iii) less then 90 % of detected fixation catch trials in HRP. On the basis of the visual field charts at baseline (see also the appendix), the visual field defects (22 homonymous, 1 heteronymous) were classified as follows: Hemianopia to the left (n = 10) or to the right side (n = 4), quadrantanopia in the upper (n = 2) or lower left quadrant (n = 1), quadrantanopia in the upper right quadrant (n = 2), quadrantanopia with partial defect in multiple quadrants (n = 3) and local scotoma (n = 1).

3.6.2 Measuring the Association between Features and Treatment Outcome

Several association measures were calculated to quantify the statistical relationship between the local features and the treatment outcome. Furthermore, feature inter-dependencies were examined.

Testing Significant Differences between Hot and Cold Spots

The mean value of all features was calculated for hot (μ_{hot}) and cold spots (μ_{cold}) separately in order to examine the group differences. In order to exclude the influence of the individual subject's performance on the averages for cold and hot spots, the average was determined at first for each subject regarding the two groups of hot and cold spots. Next, the average among all subjects was calculated for the respective group of hot and cold spots. This difference was tested for statistical significance with the 2-tailed Mann-Whitney-U-Test.

Testing the Correlation between Feature and Treatment Outcome

The first correlation (Pearson's Rho) was subject-based and establishes the relationship between topographic features and the number of observed hot spots per subject (a chart related measure). This correlation required the

replacement of the original features by dummy variables because the three features are spot-related (local), whereas the number of observed hot spots per subject is chart-related (global). Each feature was divided into two intervals (the lower and the upper interval) by using the median of the respective feature from all spots (the complete set of spots was collapsed among all subjects). A subject related dummy variable was obtained by counting the number of spots in each subject's chart contained in the upper interval (in the case of *Residual Function* and *Neighborhood Activity*) or contained in the lower interval (in the case of *Distance to Scotoma*). The second correlation measure was based on the level of spots, directly. In order to examine whether a location with high or low values (with respect to the pre-treatment features) showed post-treatment restoration or not, an association measure (ranking based non-parametric Spearman's Rho) correlates each spot's feature value with its binary post-treatment classification (cold or hot spot). In order to perform this correlation independent of subject specific properties, the following modified bootstrap resampling method was used:

1. Prepare 23 sets Π_1 to Π_{23} such that Π_i contains all feature-outcome tuples of all spots of the i-th subject; all Π_i together form the set $\Sigma = \{\Pi_1, .., \Pi_{23}\}$.

2. Draw only one sample π from each Π_i with replacement resulting in a sample set of spots $\Sigma^* = \{\pi_1, .., \pi_{23}\}$.

3. Compute the Spearman correlation coefficient ρ on the basis of each spot's feature value and its binary treatment outcome (cold or hot spot) of Σ^*.

4. Perform the second and third step for 1000 iterations resulting in the correlation coefficients ρ_1 to ρ_{1000}.

5. Compute the median among all ρ_i.

As the result of these steps, a subject-independent correlation measure was obtained for the spot's restoration response (hot or cold). On the basis of the distribution of the ρ_i, 95 % confidence intervals were computed to test for statistical significance.

3.6.3 Differences between Hot and Cold Spots

In order to compare the restoration of the subject sample with prior studies, the differences of detected stimuli between baseline charts and post-treatment charts were calculated. On average, the subject sample improved from 293.6 ± 12.4 detected stimuli before to 322.5 ± 12.7 after VRT (of a total of 474 test stimuli). The average absolute improvement was thus 6 % ± 2.7 % (mean ± S.E., p = 0.01, paired T-test) which is in the range of prior studies [Kasten et al., 1998a, Mueller et al., 2003, subm., Poggel et al., 2004, Sabel et al., 2004]. Twenty three subjects were studied with 474 testing positions each. A total sample of 10,902 baseline visual field spots was collected from all of the subjects, collectively. This included 688 hot spots and 3,426 cold spots. Figure 3.11 shows the distribution of hot spots among all 23 subjects.
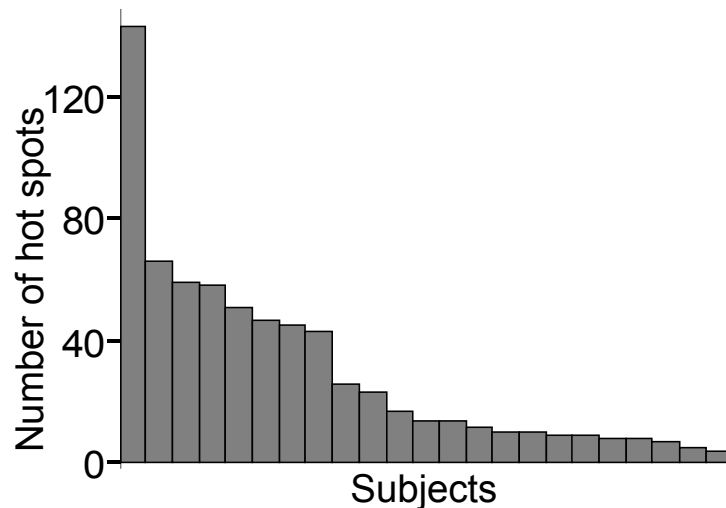
Figure 3.11: Number of hot spots after VRT for each of the 23 subjects. The charts of all subjects are shown in the appendix.

Residual Function

Hot spots (μ_{hot} = 0.16 ± 0.01) have a significantly higher (p < 0.01) *Residual Function* at baseline than cold spots (μ_{cold} = 0.04 ± 0.01). At baseline, hot spots consist of absolute-defect (black color in Fig. 2.4) and residual (moderate-defect, dark gray color in Fig. 2.4) spots in similar

proportion (46 % were moderate-defect, 54 % were absolute-defect). Contrasting this, 93 % of cold spots were absolute-defect at baseline and only 7 % were moderate-defect. Thus, a spot with a detection probability of 33 % (moderate-defect) at baseline is likely to become a hot spot after treatment.

Neighborhood Activity

The feature *Neighborhood Activity* measures the average probability of stimulus detection (at baseline) in the local spatial neighborhood of a 5 ° radius around cold and hot spots. It was found that hot spots have a significantly ($p < 0.01$) higher *Neighborhood Activity* ($\mu_{hot} = 0.37 \pm 0.03$) than cold spots ($\mu_{cold} = 0.10 \pm 0.03$) in their respective spatial neighborhood. This demonstrates that hot spots are surrounded by significantly more intact areas or areas of residual vision than cold spots.

For controlling the robustness of the obtained results, it was tested as to whether any difference in the *Neighborhood Activity* between hot and cold spot exists only locally or in the complete hemifield. The feature *Hemifield Neighborhood Activity* averages all stimulus positions in the ipsilateral hemifield (where the testing spot is located) but excluded the directly adjacent spatial neighborhood of 5 ° (see Fig. 3.12 bottom left). The *Hemifield Neighborhood Activity* was not statistically different between cold ($\mu_{cold} = 0.32 \pm 0.05$) and hot spots ($\mu_{hot} = 0.34 \pm 0.05$). Thus, the complete hemifield did not contain more intact areas around hot spots than around cold spots. This indicates that the influence of residual function on restoration is probably limited to the immediate neighborhood.

Another robustness test examines the differences in the neighborhood between cold and hot spots at higher spatial resolution using cortical distances. The *Neighborhood Activity* was measured in a thin 'corridor' as a function of cortical distance to the respective hot or cold spot (see Fig. 3.12, right).
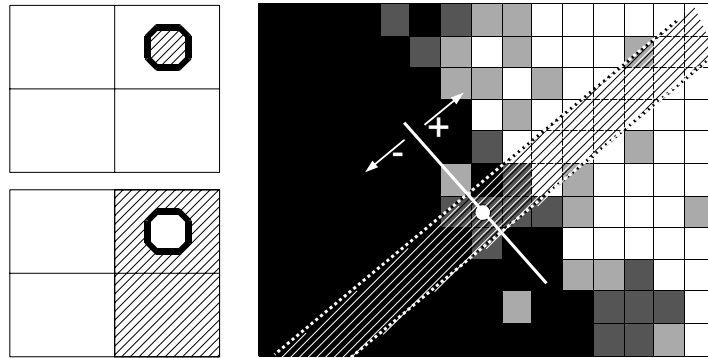
Figure 3.12: Left: The *Neighborhood Activity* (top left) is computed on the basis of the spot's spatial neighborhood (area with hatching) whereas *Hemifield Neighborhood Activity* (bottom left) is computed on the basis of the spot's hemifield minus the spot's spatial neighborhood (area with hatching). Right: The spatial neighborhood was examined one-dimensionally but with high spatial resolution in a thin corridor (area with hatching between the 2 dashed lines) which is defined by any selected spot (white circle) and runs orthogonal to the scotoma border. Positive distances are assigned to the part of the corridor in direction to the intact area and negative distances are assigned to the part of the corridor in opposite direction to the intact area.

This analysis is restricted 3-fold:

- to moderate-defect spots

- to the orthogonal orientation of the scotoma border

- to the ipsilateral hemifield.

Only moderate-defect spots are considered because the moderate-defect spots are typically located around the scotoma border and are almost equally distributed among cold and hot spots (45 % of moderate-defect spots are cold spots and 55 % are hot spots). This contrasts with absolute-defect spots which usually form huge homogeneous areas of cold spots. The second restriction is necessary to reduce the artefact which results from the high distances to the spots from the often coast-like form of the scotoma border. Therefore, the examined neighborhood is only one dimensional and considers spots which are located on a virtual line connecting the spot and the

border position of minimal distance to the spot; hence the orientation is orthogonal to the border (see Fig. 3.12, right). Thirdly, the analysis is restricted to the ipsilateral hemifield because the contralateral hemifield (which is typically intact) would otherwise interfere with this analysis. This analysis of *Neighborhood Activity* as a function of cortical distance requires the transformation of diagnostic charts into cortical charts which was described above (see chapter 3.4).

The results show that the activity around hot and cold spots differs maximally ($> 1/5$ of the complete scale) in the close neighborhood (see Fig. 3.13). Hence, the residual activity around hot spots is spatially limited and does not generalize to areas at greater distances.
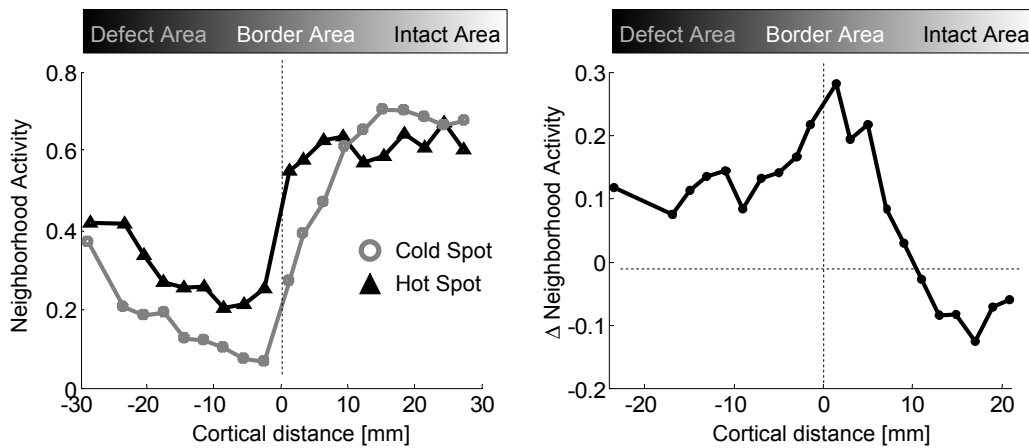

Figure 3.13: Left panel: *Neighborhood Activity* of cold and hot spots (collapsed into one set from all subjects) as a function of distance measured from hot and cold spots. Only moderate-defect spots which are typically found near the intact defect border are considered. The examined spatial neighborhood is one dimensional and restricted to a thin corridor located on a virtual line which runs orthogonal to the scotoma border (see Fig. 3.12). The corridor itself is limited to the damaged hemifield. Positive distances are assigned to the part of the corridor oriented towards the direction of the intact area and negative distances represent the opposite direction. Right panel: The difference of both functions (hot minus cold) shows that the maximal difference (which is equal to more than 1/5 of the complete scale) is found $< 4\ mm$ from the neighborhood center.

Distance to Scotoma

Hot spots ($\mu_{hot} = 3.2 \pm 0.67\ mm$) were found significantly closer ($p < 0.01$) to the scotoma border than cold spots ($\mu_{cold} = 5.9 \pm 0.65\ mm$). As Fig. 3.14 shows, the majority of hot spots (75 %) were located within 4 mm from the scotoma border. This analysis was carried out using visual cortex coordinates to consider cortical magnification.
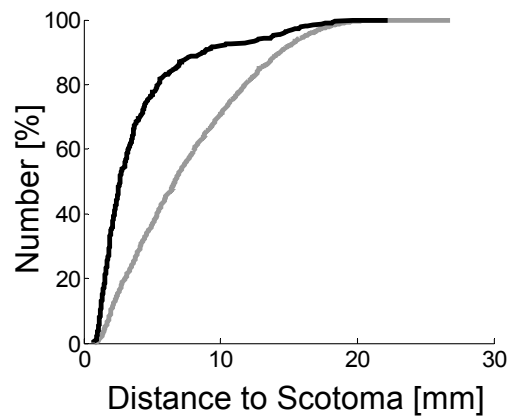

Figure 3.14: The cumulative distribution for *Distance to Scotoma* in cortical coordinates (spots with *Distance to Scotoma* = 0 are not considered in the graph). Hot (black) and cold spots (gray) of all subjects are pooled into one set.

Correlation of Features and Treatment Outcome

Significant correlations (see Table 3.1) were found between the treatment outcome and the topographic features. The number of detected fixation catch trials in baseline and post-training charts (an indirect measure of eye movements) did not correlate significantly with the number of hot spots.

Feature Interdependencies

To study how *Residual Function* and *Neighborhood Activity* interact, the number of cold and hot spots was plotted as cumulative function (Fig. 3.15 left). Generally, the neighborhood was found to be more active around

Table 3.1: Correlation coefficients between features and treatment outcome in a resampling design at the local level of spots (second column, Spearman's Rho) and at the global level of individual subject charts (third column, Pearson's Rho). For the latter, the occurrences of spots per subject which satisfy (i) *Residual Function* identical to the median of 0, (ii) *Neighborhood Activity* above the median of 0.03 and (iii) *Distance to Scotoma* below the median of 4.93 mm were counted.

Feature	Spearman Correlation with type of spot (cold,hot)	Pearson Correlation with number of hot spots per subject
Residual Function	0.73**	0.70**
Neighborhood Activity	0.59**	0.56**
Distance to Scotoma	-0.19	0.44*
Baseline fixation	N.A.[1]	0.17
Post-treatment fix.	N.A.[1]	0.07

Abbreviations: ** = correlation significant at the 0.01 level, * = correlation significant at the 0.05 level, [1]=the fixation measures are chart-related, a correlation with individual spots is not applicable.

moderate-defect spots than around absolute-defect spots (independent of the classification of cold or hot spots). Another interdependent relationship of *Neighborhood Activity* was found with the feature *Distance to Scotoma*. As Fig. 3.15 (right) shows, the neighborhood does not differ between hot and cold spots as long as they are located on the scotoma border. In contrast, if hot and cold spots are located at some distance from the visual field border, the neighborhood around hot spots was found to be significantly more active than around cold spots.

The mutual distribution of hot and cold spots with respect to *Distance to Scotoma* and *Neighborhood Activity* is shown in Fig. 3.16. In areas where the distribution of hot and cold spots is similar (close to the scotoma border) the TOPM would have potentially higher prediction errors.

The topographic features are statistically dependent among each other because a significant correlation (Person's Rho) on the individual level was found among all dummy variables of the three feature combinations: 1 with 2 = 0.85, 1 with 3 = 0.58, 2 with 3 = 0.66 (1 = *Residual Function*, 2 = *Neighborhood Activity*, 3 = *Distance to Scotoma*).

3.7 Conclusion

Global features were identified from the VRT literature and were considered only if their association with the treatment outcome was evidently described. Local features were constructed on the basis of the vision plasticity literature and their predictive value was examined in an evaluation study. The results of the study should be interpreted with caution because the scotoma border was only defined by a behavioral perimetric task and was not directly visualized anatomically in the cortex. Furthermore, the *Residual Function* of sensory areas is assessed only indirectly by the ability to detect above-threshold stimuli. This includes both, the perception of the stimulus and the execution of the motor response. It is furthermore possible that eye movement artefacts might contribute to the findings but efforts were made to reduce their influence (see chapter 5). In conclusion, the evaluation study found that all three features (*Residual Function*, *Neighborhood Activity* and
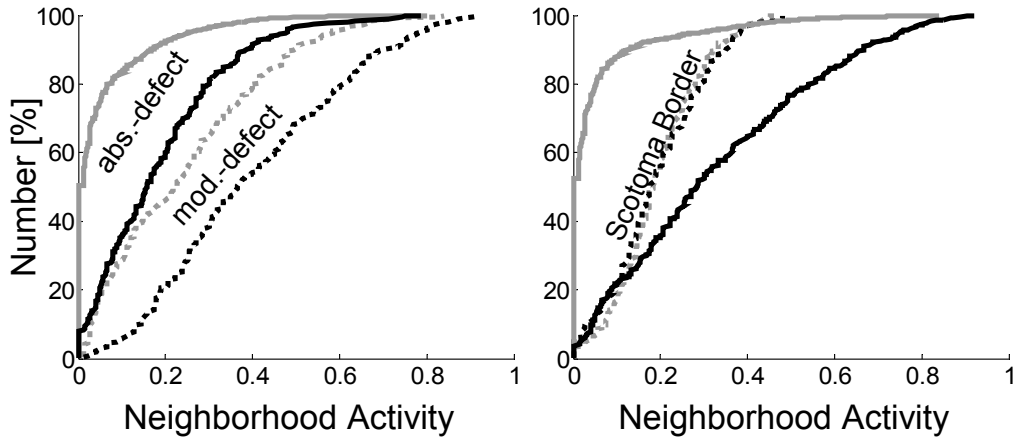
Figure 3.15: Hot (black) and cold spots (gray) of all subjects are collapsed into one graph. Left: The cumulative distribution of *Neighborhood Activity* is shown separately for moderate-defect (dashed) and absolute-defect (solid) spots. Right: The cumulative distribution of *Neighborhood Activity* is shown separately for spots located on the scotoma border (dashed, *Distance to Scotoma* = 0) vs. beyond the scotoma border (solid, *Distance to Scotoma* > 0).
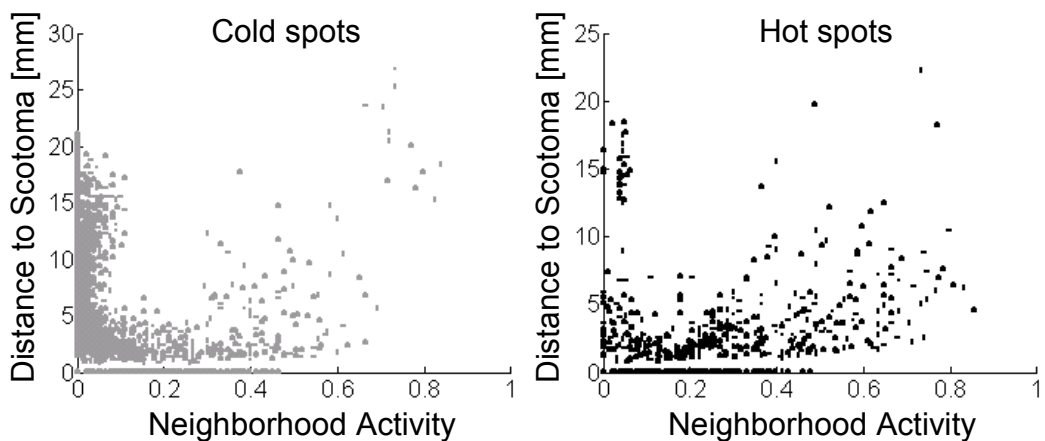

Figure 3.16: The scatter plots of cold (left) and hot spots (right) with respect to *Neighborhood Activity* and *Distance to Scotoma*.

Distance to Scotoma) have a higher predictive value which shows that these features can be used for building a TOPM. However, a prediction model should be used which is robust to shared variance among all three features.

The results which were presented in this chapter were derived from a well selected and largely homogeneous subject group. Furthermore, algorithms were presented which consider subject related dependencies between spots. Descriptive statistic results from a larger and more inhomogeneous set of subjects are shown in appendix B. These statistical measures were computed from spots of all 52 subjects pooled into one set and therefore neglecting subject related dependencies between spots.

Chapter 4

Treatment Outcome Prediction Model

4.1   Introduction

A prediction model for the field of visual system plasticity was developed using Self-Organizing-Maps (SOM) as the core of the TOPM. Besides the generally accepted recommendation to favor simple algorithms rather than complex algorithms [Sajda, 2006], no consensus is available which favors a specific algorithm for a specific problem. The SOM is a relatively 'lightweight' algorithm which offers excellent data visualization methods and can also be used for classification and prediction. It is applied to data obtained from patients with visual field defects caused from brain damage after stroke, brain trauma or other etiologies. Because neither all patients nor all areas of the visual field necessarily profit from VRT, it is desirable to find methods that efficiently predict vision restoration. The following is a useful definition of medical prognosis [Abu-Hanna and Lucas, 2001]:

> 'Medical prognosis is the prediction of the future course and outcome of disease processes, which may either concern their natural course or their outcome after treatment.'

The TOPM described here, predicts the outcome of intrasubject visual field areas. The local prediction is spot-based (a spot is a point in the visual field)

73

and the treatment outcome of a specific spot is either improvement (termed 'hot spot') or no improvement (termed 'cold spot'). Spots showing functional deterioration are not considered here because they are very rare and within the expected normal variability. The TOPM methodology is comprised of the following three steps (see Fig. 4.1):

1. Building a TOPM which can predict the treatment outcome based on the baseline diagnosis by using diagnostic charts from patients who completed VRT.

2. Extracting predetermined features (see Table 4.1) from baseline diagnostic charts. In these baseline performance charts, the spots are classified as being either 'impaired' (spots with 0 or 1 out of 3 possible detections) or 'intact' (spots with 2 or 3 out of 3 detections).

3. Predicting the treatment outcome for all impaired spots by using the TOPM derived in step 1 based on the features from step 2.

Table 4.1: Specifications of Features which are used for the TOPM.

Feature	Minimal value	Maximal value	Scale	Type
Residual Area	18	117	Ordinal	global
Defect Area	54	219	Ordinal	global
Border Diffuseness	0.1	0.7	Continuous	global
Reaction Time [ms]	380	540	Continuous	global
Hemianopia	0	0.8	Continuous	global
Quadrantanopia	0	0.8	Continuous	global
Horizontal Position [°]	-24	18	Continuous	local
Vertical Position [°]	-12	16	Continuous	local
Neighborhood Activity	0	0.6	Continuous	local
Neighborh. Homogeneity	0	0.4	Continuous	local
Dist. to Scotoma [mm]	0	15.2	Continuous	local
Residual Function	0	0.33	Binary	local

A supervised database was designed to support SOM learning and cross validation which contained pairs of baseline and post-treatment diagnostic charts from 52 selected patients. Charts from another three patients that
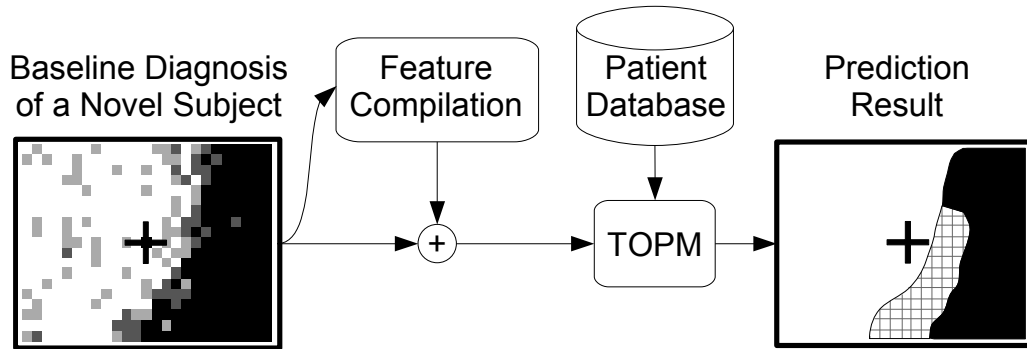
Figure 4.1: The flow chart for prediction of the treatment outcome. The baseline diagnosis (left) assesses the visual field of the patient before treatment (the cross represents the center of vision). From this diagnostic chart, several features are extracted directly or were computed incorporating a priori knowledge. The TOPM calculates the predicted treatment outcome (right) of impaired spots based on cases in the training set. The prediction result shows prognosed improvements (hatching) and lack of improvement (black area) after VRT.

were not part of the learning or test set were used to show the prediction for those three cases. All patients had homonymous visual damage except for one case with some heteronymous elements on top of a homonymous defect. Diagnostic data were retrospectively selected from a patient pool of two different prior studies [Mast et al., 2006, Mueller et al., 2003] in which the patients participated. Inclusion criteria were (i) post-chiasmatic damage, (ii) availability of at least three binocular diagnostic tests in pre- and post-treatment diagnosis and (iii) that the patient participation in the experimental group, not in the control group. In order to reduce any potential error source caused by eye movements, patients were excluded from the analysis who showed poor fixation performance (only < 90 % of color based fixation catch trials were detected). Both studies were carried out in concordance with the declaration of Helsinki.

4.2 The Kohonen Self-Organizing-Map

In addition to selecting appropriate features, the development of a TOPM requires the selection of an appropriate classification algorithm. No generally accepted recommendation exists for choosing the optimal classifier for data analysis and classification. Rather, the choice is determined by experience, knowledge and personal preference of the user. SOM's, introduced by Teuvo Kohonen [Kohonen, 1990], are based on relatively 'simple' mathematical operations. The most complex operation in SOM learning is the calculation of the Euclidian distance in multidimensional space. SOM's have therefore convinced experts of their usability in the medical domain [Wyatt and Altman, 1995]. They are a preferred tool for exploratory data analysis. Besides their data visualization capabilities, SOMs are used for correlation hunting, unsupervised and supervised cluster analysis, multivariate feature analysis as well as for novelty detection and classification [Vesanto, 1999]. In practice, SOMs reduce the high dimensional feature space to low dimensions. The SOM surface is usually two-dimensional and consists of map units which enable the data to be visualized while preserving the topography of the original feature space. The learning procedure incorporates a competition scheme (every map unit competes with each other to best match the data samples) and a cooperation scheme (the best matched map unit and its local neighborhood on the map participates in learning) [Kohonen, 1990]. Each map unit represents a specific position in the input data space and becomes a highly sensitive feature detector during the learning process [Vesanto, 1999]. In supervised learning, the SOM units are labeled according to the data labels of the samples which are located close to the SOM units in the feature space. After the map units are labeled, novel data samples are classified by using k-nearest prototype classification (k-Nearest-Neighbor with k equal to 1). The classification depends in turn on the label of the SOM unit with the smallest Euclidian distance to the novel sample. Preprocessing prior to learning should include (i) removal of weak features from the learning database which are clearly not associated with the target feature and (ii) normalization of all features to equal variance. The latter point is required

as the evaluation of the best matching map unit is based on the Euclidian distance which overweights features with higher variance to the disadvantage of features with lower variance [Vesanto, 1999]. For further improvement of the classification accuracy, the learning vector quantization is recommended to be used after SOM-learning. This tunes the SOM units to maximize the margins between map units of different classes [Vesanto, 1999]. This tuning function was not used here due to the high computational costs in combination with k-fold cross-validation. In classification, the SOM architecture supports the evaluation of the degree of matching, between novel samples which were used in the learning phase. The degree of matching can influence the prediction accuracy. Whereas novel cases that lay within the training range are interpolated without concern, extrapolating samples outside the training range is problematic [Veropoulos, 2001]. The SOM allows the detection of outliers and measures the degree of concordance between the sample and the training data set by considering the distance between the sample and its nearest map unit or, by measuring the distance to all map units such that an overall response map is created [Vesanto, 1999]. In practice, SOMs conveniently combine both, data exploration and prediction.

4.2.1 A SOM Learning Example

During SOM learning, the SOM approximates the data distribution of the training set Λ. As an example, a set of training samples (2-dimensional) is shown in Fig. 4.2. The feature space Φ is therefore spanned by two features. For illustrative purposes, graphical attributes are used to locate a specific point in the two-dimensional feature space. Value ranges of the first feature are represented by different patterns of hatching and values of the second feature are represented by a brightness distribution. Hatching patterns or color distributions are less precise than explicit numbers but allow a better understanding of the data distribution by using visual aids.

A SOM with 4 map units and their connection structure is shown in Fig. 4.3 (top right). Each map unit has a storage capacity of two numbers (associated with the two features) representing a specific location in the feature
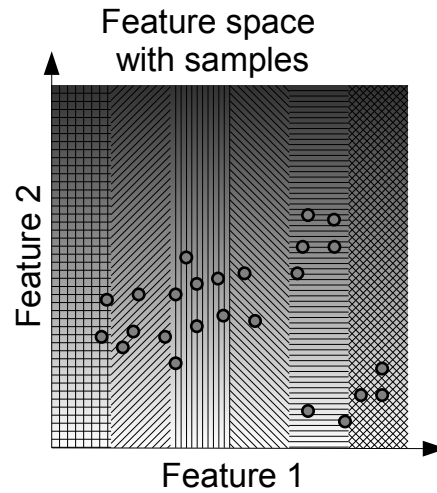
Figure 4.2: The distribution of training samples in a two dimensional feature space.

space. Assuming that the SOM units are initially located in the 'corners' of the feature space (indicated by white crosses in Fig. 4.3, left) the feature values for the respective SOM unit are indicated by the respective brightness and hatching pattern (Fig. 4.3, bottom right). A SOM structure which shows the respective feature values by e.g. a brightness distribution is termed 'component plane'.

Both, training samples and initialized location of the SOM units are shown together in Fig. 4.4, left. During learning, the SOM units are moving towards the training samples according to the following algorithm:

For a training sample λ in the training set Λ do

1. Find the map unit m (also termed 'winner') among the set of all SOM units M with least Euclidian distance d such that

$$m = argmin_M \{d(\lambda, m)\}$$

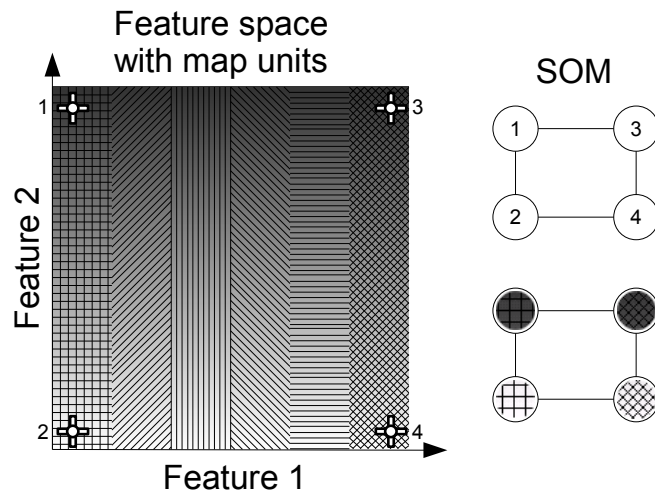2. For building the set N, add to m all units in M which are in the

Figure 4.3: The SOM structure with 4 units (top right). Each map unit represents a location in the feature space (marked with white crosses, left). The SOM component plane (bottom right) shows the feature value for the respective map unit.

neighborhood of m. The neighborhood is defined by connecting arcs between the SOM units.

3. For all units n in N compute the new location pos'_n in the feature space with respect to the old location pos_n, the location of the training sample pos_λ and a step size γ according to:

$$pos'_n = pos_n + \gamma \cdot (pos_\lambda - pos_n)$$

An example is presented in Fig. 4.4 (middle). The first training sample (marked with a triangle in the feature space) is selected at the beginning of the first iteration. At first, the nearest SOM unit is determined (the unit 2 at the bottom left of the SOM), then its neighbors are found by following the connecting arcs (units 1 and 4). The winner unit and all neighbors in the set N (marked with the rectangle on the SOM plane) participate in the location update and move 'towards' the training sample (indicated by arrows in the feature space). After one complete iteration through all data samples, all map units which were either the winner unit itself or in the neighborhood

of a winner unit, have updated their location. Usually, the learning phase consists of hundreds of iterations. After learning, the data distribution is approximated by the map units (see Fig. 4.4, right) as the map units are now located in regions with a high density of data samples.
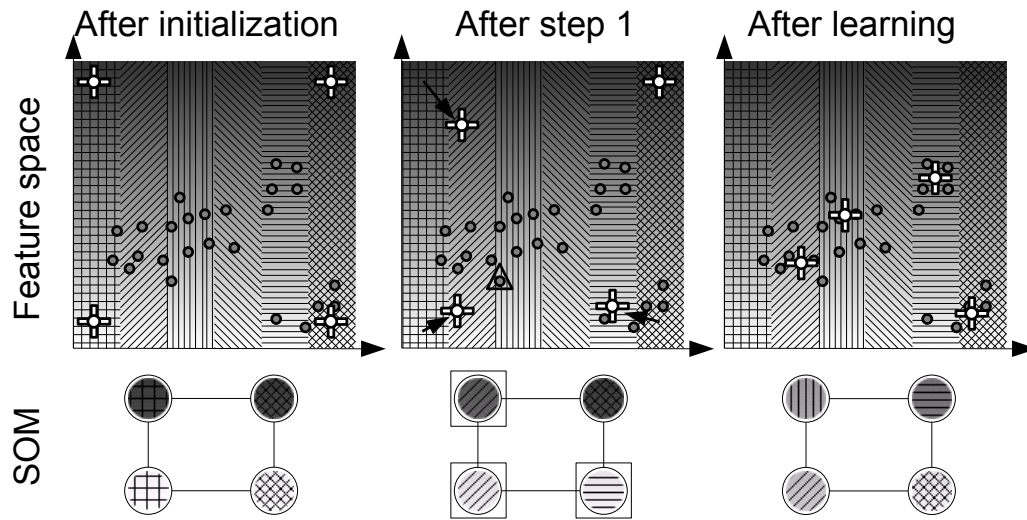

Figure 4.4: The SOM learning phase. The winner unit is determined for the training sample (the sample is marked with the triangle) in a competition such that the unit with the least distance to the training sample wins. In a cooperative way the winner unit and all units in the winner's neighborhood (marked with rectangles) move toward the actual training sample (middle, the movement is indicated by arrows). After learning, the SOM units have approximated the training samples (right).

The learning procedure results in a SOM which can be used for data visualization. The component plane shows the location of SOM units in the feature space for the first (illustrated by hatching) and second feature (illustrated by brightness). Usually, in order to enhance the visibility, the component planes are shown separately for each feature. Each SOM unit is regarded as a cluster and represents all training samples within the cluster. By visual inspection of the component planes, it is possible to derive a hypothesis about the data distribution or about correlations between the features. Therefore it could be concluded from the SOM component plane in Fig. 4.4, bottom right, that no or only few samples exists with values of

feature 1 corresponding to the vertical-horizontal crossed hatching and with values of feature 2 corresponding to dark gray or black brightness.
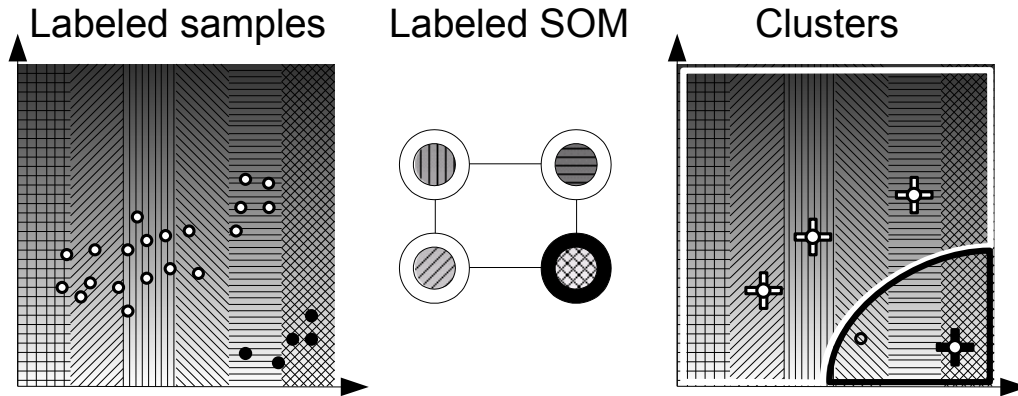

Figure 4.5: The training samples are labeled as of positive (white circles) and of negative (black circles) class (left). After unsupervised learning has finished (shown in Fig. 4.4, right), the SOM units are labeled according to the labels of training samples in their surround (middle). The class information of the SOM units is then used to separate the feature space into positive and negative clusters (right). A novel sample (open circle) is located within the negative cluster. The predicted class is therefore 'negative'.

The algorithm described above belongs to unsupervised learning. No class information is used that separates the data samples (e.g. representing patients) from e.g. the two classes of patients with poor treatment and good treatment outcome. However, to use the SOM for prediction, the class information is necessary (hence supervised learning is required). Therefore, the class information is added to the samples presented above (see Fig. 4.5, left). The class variable is dichotomous and the training samples are therefore either positive (white circles) or of negative (black circles) class. After unsupervised learning has finished, the class information of the data samples is used to determine the class label of the map units. The map unit label (black or white) is determined by finding the most frequent class label of the data samples in the surround of the map units (the labeling is shown in Fig. 4.5, middle).

The labeled SOM is ready to be used for the prediction of novel samples. The class information of the SOM units determine the separation of the feature space Φ into clusters with different class labels (see Fig. 4.5, right). The classification of a novel sample which is located in the bottom right corner of the feature space is therefore classified as belonging to the 'negative' class.

4.3 Data Visualization

SOM-based data exploration is usually carried out on two dimensional component planes. These planes exist for each feature and are compromised of the map units (the horizontal and vertical axis of the component planes have no specific name or meaningful interpretation). These planes are feature-specific projections onto the SOM surface after learning has finished. Each map unit represents a specific point in the n-dimensional feature space Φ (where n is the number of features). Importantly, due to the cooperative concept in SOM learning, adjacent map units (neighboring units in the SOM map) have similar feature values resulting in a smooth brightness distribution. The component planes and map units are shown for all 12 features in Fig. 4.6.

Each map unit always has the same position in each of the 12 component planes. Map units are labeled as hot ('+') and cold spots ('0') and separate the feature space into regions of these two classes. The labels are not shown for each map unit because the map is separated compactly between hot and cold spot map units. Each data sample in the training data base is assigned to exactly one map unit (minimizing the distance between map unit and data sample). The topography of hot and cold spots in the component planes after learning (shown in Fig. 4.6) is relatively robust if the process of SOM learning is repeated from the same initialization state. Preferably, the distance between the closest cold and hot spot map units in Euclidian space should be relatively broad in comparison to the distance between map units belonging to the same class (see Fig. 4.7).

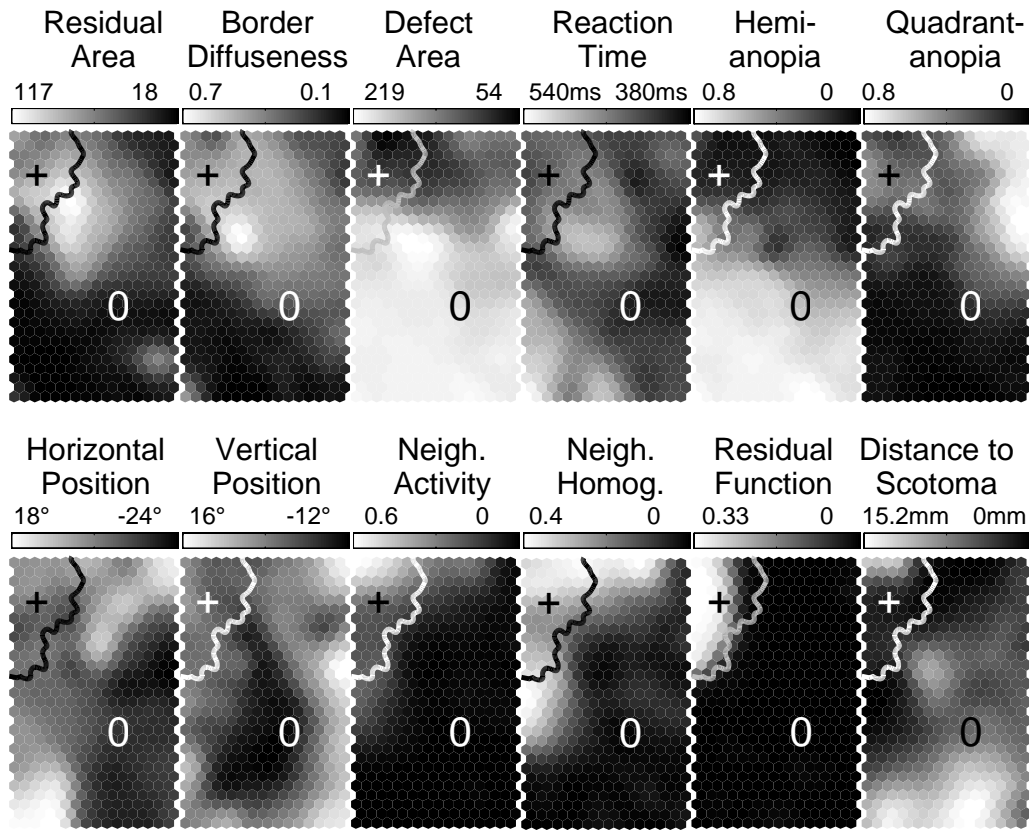This would indicate that there is a 'natural' difference between cold and

Figure 4.6: Component planes of global (top row) and local (bottom row) features extracted from the baseline diagnostic. The distribution of values of the respective feature is shown for hot ('+') and cold ('0') spots. The SOM is comprised of 16 x 29 map units in a hexagonal grid such that each map units is connected with its 6 neighbors.

hot spots in the sample distribution which allows a more robust discrimination between both classes in prediction.

By visual inspection of the component planes, hypotheses about the relationship between the treatment outcome and the various features as well as interdependencies are formulated. For example, visual inspection shows that high values in the size of the feature *Residual Area* tend to be associated with hot spots whereas most cold spots are not (low values of this feature in the cold spot region). This example shows a more pronounced separation between hot and cold spots in contrast to *Reaction Time* where the component
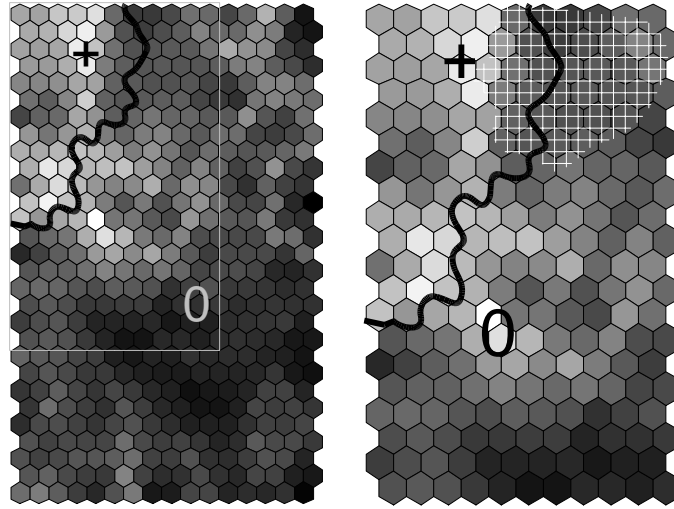
Figure 4.7: The distance matrix shows the distances (white indicates high distances, black indicates small distances) between neighboring map units in the grid (labeled either as hot ('+') or cold ('0') spot) in the input feature space Φ. A region of the map (rectangle in the left figure) contains map units labeled as hot spots. These units are enlarged in the right panel showing a region at the top border (white hatching) where the map units of both classes (hot and cold spots) are not well separated. This makes classification more difficult. According to the component planes (see Fig. 4.6) this region refers to spots of low *Residual Function* in charts with small numbers of *Defect Area*.

chart appears more uniform (gray). The following hypotheses are proposed as a result of the visual analysis indicating areas for further analysis by statistical and experimental means: (i) almost no restoration (cold spots) is observed in charts with small *Residual Area* (black color), with low *Border Diffuseness* (black color) or without any affinity to *Quadrantanopia* (black color), (ii) a univariate model with *Hemianopia* or *Reaction Time* alone is not appropriate as separator between hot and cold spots and (iii) *Residual Area* and *Border Diffuseness* are strongly positively correlated, *Hemianopia* and *Quadrantanopia* are strongly negatively correlated considering the color distribution of the respective features in the component planes.

Again, hypotheses were formulated for the component planes of local features (see Fig. 4.6, bottom row): (i) *Residual Function* is the best dis-

criminator between hot and cold spots; only few map units in the hot spot region have a low value of *Residual Function*, (ii) all map units which are labeled 'hot spot' represent data samples which have a *Neighborhood Activity* and *Neighborhood Homogeneity* well above zero (white and gray color), (iii) a univariate model with *Horizontal* or *Vertical Position* alone is not appropriate to discriminate between cold and hot spots because map units of both classes represent data samples of middle values (gray), (iv) most data samples with high *Distance to Scotoma* (white color) are cold spots but not all hot spots are close to the scotoma border (*Distance to Scotoma* $\approx$ 0, black color) and (v) the features *Neighborhood Homogeneity* and *Neighborhood Activity* are strongly correlated (the color formation looks familiar in both component planes). In general, the component charts indicated that global features of the visual charts as a whole are less appropriate for classification than local features. The colors in the later group of features (Fig. 4.6, bottom row) show a better association with the map unit labeling of hot and cold spots in comparison to the color distribution of the global features (Fig. 4.6, top row).

4.4 Treatment Outcome Prediction

Forecasts predicted by the TOPM are shown for three patients (see Fig. 4.8, left column). Those patients were selected randomly and their charts were not used in SOM training. The predicted charts are obtained by using the developed TOPM according to the schema shown in Fig. 4.1. Values of the global and local features are initially extracted for each spot such that the location of each spot in the 12 dimensional input feature space Φ is located. Secondly, by using the k-nearest prototype algorithm (identical to the k-nearest neighbor algorithm with $k = 1$), the best matching SOM unit is determined for each spot which determines thereafter the predicted class (hot or cold) of the respective spot. The actual as well as the predicted treatment outcome is shown in Fig. 4.8 (second and third column). The actual treatment outcome is computed as follows:

$$\text{Actual outcome}(x,y) = \begin{cases} \text{'hot spot'} & \text{iff chart}_{Detection}^{Baseline}(x,y) <= \frac{1}{3} \cap \\ & \quad \text{chart}_{Detection}^{Post}(x,y) >= \frac{2}{3} \\ \text{'cold spot'} & \text{iff chart}_{Detection}^{Baseline}(x,y) <= \frac{1}{3} \cap \\ & \quad \text{chart}_{Detection}^{Post}(x,y) <= \frac{1}{3} \end{cases} \quad (4.1)$$

To localize the best matching SOM unit for spots of a specific diagnostic chart, the spot's representation in the feature space Φ is projected onto the SOM in Fig. 4.8 (right column). Almost all samples of the first chart (Fig. 4.8, top row) are located at the top right position in the SOM. According to the component planes, this area is more strongly related to quadrantanopia, with small *Residual Area* and low *Residual Function*. Spots of the second diagnostic chart (Fig. 4.8, middle row) are widely spread on the SOM. The top left corner in the SOM chart is related to maximal *Neighborhood Activity* and the aggregation in the SOM chart center is related to negative horizontal coordinates covering positive and negative vertical coordinates with low *Neighborhood Activity*. This describes the left defect hemifield. In the third diagnostic chart, the strongest aggregation of spots is closely located to the hot spot border within the region of cold spots (Fig. 4.8, bottom row). This region is related to negative vertical coordinates with a similarity to both, quadrantanopia and hemianopia with sharp borders. These spots belong to the lower left part of the semi-quadrantanopic diagnostic chart which has also many defects in the upper quadrant of the left hemifield.

4.5   Model Evaluation

A k-fold-cross validation design was selected to evaluate the performance of the TOPM [Veropoulos, 2001]. Cross validation is a resampling method where the test data set is first divided into k disjunctive sets. Instead of calculating only one evaluation measure, k-fold-cross validation offers a more robust evaluation [Fawcett, 2004]. Therefore, the performance measurement procedure is repeated k times. In each iteration, all except one set are pooled
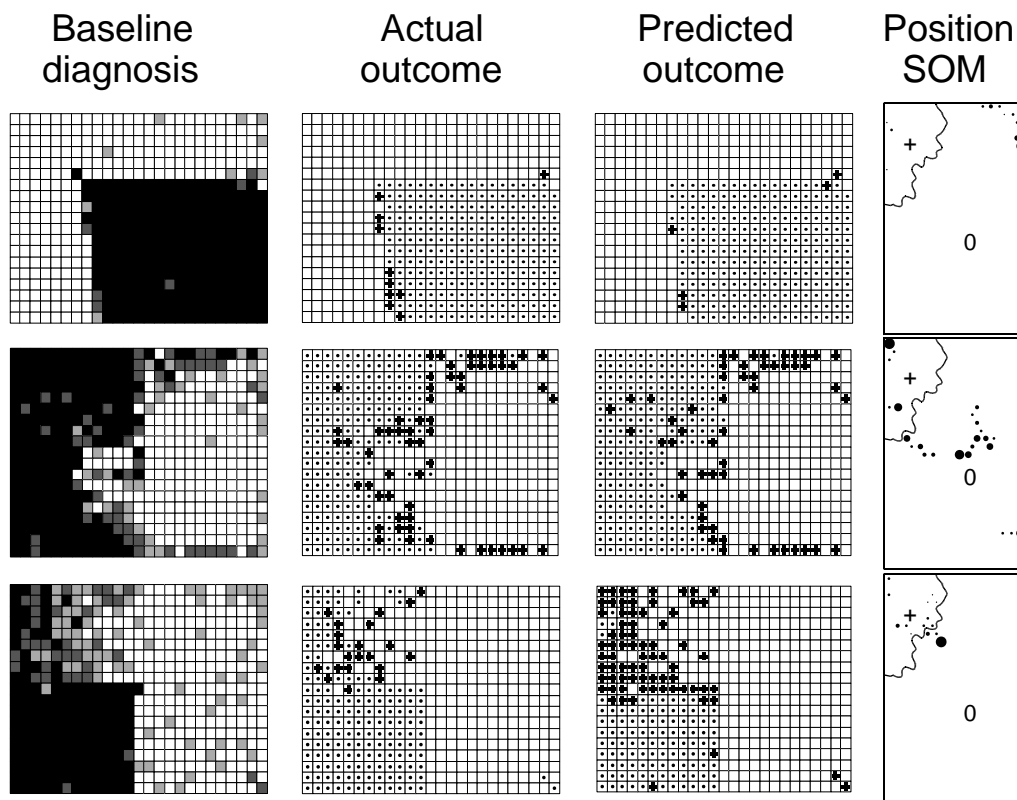
Figure 4.8: Diagnostic charts of three patients which were not part of the training data set (first column, left). The actual treatment outcome is shown in the second column (hot '+' or cold '.' spot). The third column shows the predicted treatment outcome for each impaired spot at baseline. The location of spots from individual baseline charts projected onto the SOM is shown in the fourth column (right). The respective SOM location differs widely among the three patients (higher frequencies are indicated by a larger circle size) and reflects the respective projection of global and local features onto the SOM surface.

to build the learning set Λ which is used to train the SOM model. The remaining set (Γ) is used as test set. The whole training database is therefore used for training and evaluation as well. The result is the average among the k evaluation measurements. In comparison to other methods, cross validation is an unbiased, robust estimation of the generalization error, but with higher standard deviation of the obtained measures [Green and Ohlsson, 2007]. With 10-fold cross validation, 10 classifiers were trained and then evaluated with the test samples. Standard evaluation measures were used [Green and Ohlsson, 2007] and the average True-Positive-Rate ($TPR = 44\%\ \pm 4.7\%$, how many hot spots were classified correctly) and the False-Positive-Rate ($FPR = 6\%\ \pm\ 1.9\%$, how many cold spots were classified as hot spots) was calculated among all 10 classifiers according to:

$$TPR = \frac{TP}{P} \tag{4.2}$$

$$FPR = \frac{FP}{N} \tag{4.3}$$

which is based on the number of positive samples (P, hot spots) and negative samples (N, cold spots) in Γ, as well as the number of correctly classified positive samples (TP) and incorrectly classified negative samples (FP).

The average accuracy ($ACC = 84.2\%\ \pm\ 1.4\%$) measures the amount of correctly classified samples and was found to be slightly better than the average baseline accuracy ($ACC_{baseline} = 81\%\ \pm 1.3\%$) which is the ratio of samples labeled with the most frequent class. Both measures are computed as follows (the former uses the number of correctly classified negative samples, TN):

$$ACC = \frac{TP + TN}{P + N} \tag{4.4}$$

$$ACC_{baseline} = \frac{max(P, N)}{P + N} \tag{4.5}$$

The small difference between accuracy and baseline accuracy is a result of

the strong skewness of the class distribution; because the training database contained 7026 cold spots but only 1689 hot spots. Interestingly, in some 'unfortunate' classification problems the accuracy of established classifiers is even below the baseline accuracy [Holte, 1993]. This is acceptable in the case of unequal misclassification costs where the prediction performance of the preferred class is of higher interest than the prediction performance of the non-preferred class.

4.5.1   Receiver Operating Curve Analysis

A measure which is robust to skewed class distributions is the Receiver Operating Characteristic (ROC) [Fawcett, 2004] and is therefore used as a standard in performance evaluation of classifiers in the medical domain [Swets and Pickett, 1982]. The ROC measure is a combination of the TPR and FPR measure of the same classifier. Both measures are strongly connected and are therefore presented together - a high TPR usually leads to a high FPR and vice versa. All 10 trained classifiers are located close to each other in the ROC space represented by the value of their respective TPR and FPR (points in Fig. 4.9).

Classifiers which appear on the left hand side in the ROC space (as it is the case here) are called conservative because they classify 'only with strong evidence' (low FPR but also low TPR) [Fawcett, 2004].

A more robust performance measure is achieved if a continuous ROC curve and the area under the ROC curve (AUC = 0.81) is computed for the classifier instead of just a single point in the ROC space [Pepe, 2005]. The position on the curve determines the favored value combinations of FPR and TPR. Especially in problems where the misclassification costs are not identical, the ROC curve allows chosing the position on the ROC curve for which the total costs of both, misclassification of the positive and negative class are minimal. To obtain a ROC-curve, a score is required which is interpreted as probability that the sample belongs to the predicted class. Most classifiers use an internal threshold for determining the class which can then be used as probability score. The score must at least have an ordinal
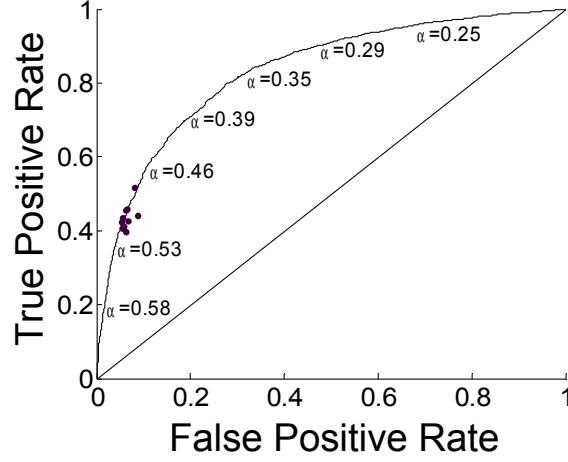
Figure 4.9: The receiver operating characteristic (ROC) for the TOPM in prognosis of hot and cold spots in 10-fold cross-validation (single points). The continuous ROC-curve was calculated by using 1-nearest neighbor classification considering different thresholds α (see text). The performance of any classifier that appears around the diagonal is equal to random guessing [Fawcett, 2004].

scale but the range is not limited (between $-\infty$ and $+\infty$). The curve is obtained by using an increasing threshold α which iterates through the full range between $-\infty$ and $+\infty$. For each iteration, all samples (i in Γ) are labeled with respect to their score and α as follows:

$$
label_i = \begin{cases} \text{'hot spot'} & \text{iff } \alpha \geq score_i \\ \text{'cold spot'} & \text{iff } \alpha < score_i \end{cases} \tag{4.6}
$$

The respective rate of correctly labeled positive samples (TPR) and incorrectly labeled negative samples (FPR) determines a point in the ROC space [Fawcett, 2004] forming a continuous curve (see Fig. 4.9). Because classification of samples by using the k-NN classifier is not based on a single score value but on distances to all samples in the training set, an explicit construction of the score is necessary such as the ratio of map units with positive class among the k nearest map units as in [Soltanian-Zadeh et al., 2004]. Another measure is proposed here, and is based on the distance be-

tween the sample and the nearest map unit with positive class (d_{hot}) and negative class (d_{cold}). The score of the i-th sample is the relative distance:

$$score_i = \frac{d_{cold}}{d_{hot} + d_{cold}} \tag{4.7}$$

The relative distance is a value between '0' (indicating that the i-th sample is close to a map unit labeled cold spot and distant to the next hot spot map unit) and '1' (close to hot spot and distant to cold spot). The labeling of being either hot spot or cold spot is ambiguous when the score is exactly 0.5. In this case, the sample is equally distant to both, the next 'hot spot' *and* the next 'cold spot' map unit.

4.5.2 Clinical Evaluation

The appropriateness of the TOPM to several subclasses of patients is of clinical relevance. By separating the samples with respect to patient affiliation, leave-one-out cross validation showed that the performance is not equal for each subject resulting in a high range of individual performance measures (minimal accuracy = 29 %, maximal accuracy = 98 %). A strong correlation between accuracy and the feature *Defect Area* ($\rho = 0.67$) shows (see Fig. 4.10) that the accuracy of diagnostic charts with many defect areas is better than in charts with only few defect areas (this is in agreement with the observation that hot and cold spot map units are not well separated in the region with low *Defect Area*, see Fig 4.7). This is probably a result of the overrepresentation of diagnostic charts with large numbers of defect areas in the training data set (50 % of all patients have a very large *Defect Area* covering more than one third of the total diagnostic area).
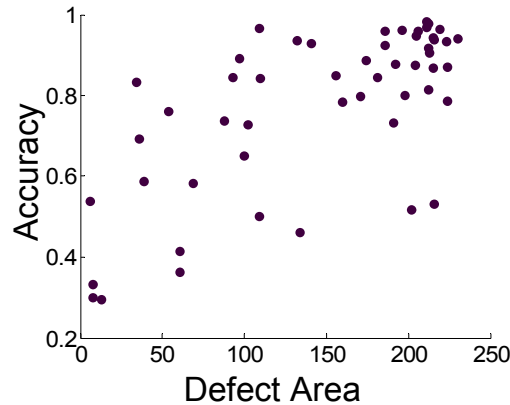
Figure 4.10: In comparison to 10-fold cross validation where the training database was divided into 10 parts, the spots in the database were divided into 52 parts with respect to patient affiliation. The prediction accuracy was determined for all individual charts in leave-one-out cross validation and where plotted against the feature *Defect Area*. Diagnostic charts with a large *Defect Area* have a better prognostic accuracy than charts with only few defect areas.

4.5.3 Comparison to a Linear Regression Model

The same features and outcome variable were used in combination with a linear regression model instead of the SOM model. Again, evaluation measures were obtained using 10-fold cross validation. The performance measures (AUC = 0.86, ACC = 85 %, $ACC_{baseline}$ = 81 %, TPR = 52 %, FPR = 8 %) of the linear regression model are slightly better (see Fig. 4.11) than the performance measures of the SOM model presented above. This might be a hint that the association between the features and the target variable (treatment outcome) is linear rather than non-linear. Another possible reason why the performance of the SOM model is not significantly better than the linear regression model is probably the high class skewness. If the positive class is extremely underrepresented in the training set Λ, it is possible that only few map units are labeled with the positive class. This is because the labeling is determined by the most frequent class of all samples near the map unit in the feature space Φ. Therefore, another SOM was trained with 1265 map units (instead of only 464) which could better represent the positive class

and could outperform (AUC = 0.89, ACC = 88 %, $ACC_{baseline}$ = 81 %, TPR = 62 %, FPR = 6 %) the linear regression model in 10-fold cross validation. Consequently, the learning phase of this extended SOM model is computationally more expensive. However, overfitting was not yet observed because the classification performance in cross validation increased together with the higher number of map units.
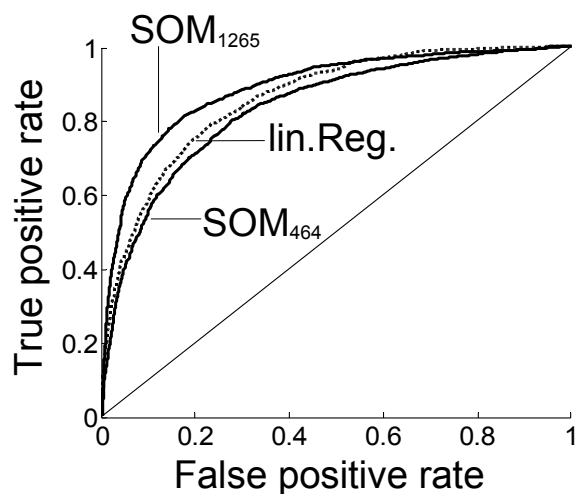

Figure 4.11: The ROC curve in 10-fold cross validation for the linear regression model (dotted) and two SOMs with 464 and 1265 map units (solid).

4.6 Discussion

Care must be taken when prediction models are used in the medical field. Unlike other classification problems (e.g. in the financial domain) the misclassification of patients could lead to changes in patient's treatment regimen. Patients could face serious consequences such as emotional stress (healthy subjects are misclassified, wrong promises etc.), inappropriate treatment, or no treatment at all [Pepe, 2005]. It has been suggested that such an analysis should go beyond the evaluation of statistical measures such as the area under the ROC curve. Therefore, additional evaluations are necessary to assure clinical usefulness of the prediction model [Abu-Hanna and Lucas, 2001].

In the case of VRT, misclassifications of treatment areas could result in longer treatment time or reduced treatment efficiency. It was reported elsewhere [Schwarzer et al., 2000] that many published prediction models are not useful for clinical application, although they apparently perform better than selected statistical competitors. Reasons attributed to the failure of these models include weak robustness and low semantic relation with the clinical domain [Abu-Hanna and Lucas, 2001]. There are some guiding principles which should be observed in assessing the clinical credibility, effectiveness and evidence of generality and accuracy of prediction models in medicine [Wyatt and Altman, 1995]. TOPMs should not rely solely on correlation analysis of features and the treatment outcome [Shillabeer and Roddick, 2006]. To extent the correlation analysis, statistical tests were used to clarify that the difference of baseline feature means from cold spots and hot spots is statistically significant (see chapter 3.6). This ascertains that the prediction model contains features which are able to separate the set of cold spots sufficiently from the set of hot spots. Domain experts should verify the appropriateness of the proposed classification principles on the basis of patient samples [Ruseckaite, 1999]. This was fulfilled as this study was regularly audited by VRT domain experts. By close examination of the SOM units that determine the class label (see Fig. 4.8, right) it is possible to analyze the rules on which the classification is determined as it is suggested elsewhere [Shillabeer and Roddick, 2006]. Furthermore, it is recommended to use *a priori* knowledge such that the robustness of the prediction model is enhanced [Lucas and Abu-Hanna, 1999]; physiological knowledge of visual cortex organization and reorganization, visual cortex anatomy and the experience of domain experts was considered to construct appropriate features. Moreover, the association between the features and the treatment outcome should be assessed in another and different population to reduce the risk of spurious association between feature and class information [Smith and Ebrahim, 2002]. The feature extraction phase was based on two independent studies done within the last 10 years. Further work is required to confirm the clinical benefits of the developed TOPM for VRT because results should be replicated, confirmed and well documented [Shillabeer and Roddick, 2006]. These requires future

rigorous testing, ideally in a double blind, randomized and placebo controlled investigation [Lucas and Abu-Hanna, 1999].

The SOM was chosen as the core of the prediction model. Its non-linearity and self-organization methodology allows a comprehensive adaptation to the data distribution. Although other prediction models may perform comparably well, SOMs simplify the process of data mining and the feature selection phase, as they conveniently combine both prediction and data exploration.

Chapter 5

Assessing the Data Quality

5.1 Introduction

During HRP, subjects are required to maintain a sufficient fixation stability. Otherwise, the diagnostic chart should be excluded from further analysis. Eye movements can produce ambiguous results in perimetric assessments. Today, fixation stability is adequately assessed by using eye trackers. In the past, no appropriate eye tracking data were available from the charts which were used in building the TOPM. Instead, a color based fixation catch trial (FCT) was used for measuring fixation stability. However, no reliable data is available describing the sensitivity of the FCT to the occurrence of eye movements. A close examination of fixation stability during HRP is necessary because the therapy outcome as well as the prediction power of the TOPM rests on the assumption that the eyes are held steady during visual field testing. Therefore, this chapter addresses two main points: (i) the occurrence of eye movements during HRP and (ii) examining the sensitivity of the FCT to eye movements.

5.1.1 Eye Movements during HRP

Eye movements were observed as a compensation strategy in hemianopic subjects [Zihl, 1995]. Those subjects learn to scan the visual area by using the remaining healthy field of vision in combination with eye movements.

However, eye movements can cause ambiguity in perimetric results. For example, the defect hemifield appears to be less defective [Jamara et al., 2003], small isolated scotoma tend to vanish or the scotoma edges are blurred out [Demirel and Vingrys, 1994].

Only few studies examined eye movements for longer than a couple of minutes. It is therefore not clear how strongly eye movements increase in tasks which require subjects to fixate properly for more than 10 min as in HRP. Short recording times are preferred to longer recording times in experimental studies because of subject discomfort and technical limits. The eye tracker recording system requires that the head and the eye tracker are firmly fixed. This makes subjects feel uncomfortable [Eizenman et al., 1992]. A problem in head mounted eye tracker systems is that the error increases steadily due to decalibration caused by small movements of the eye tracker relative to the head. In most fixations tasks, the visual environment is not very 'attractive' and boredom can lead to uncontrolled eye movements. Furthermore, limits of the statistical software [Kasten et al., 2006] and an expensive frame by frame analysis of recorded images in camera based systems can limit the recording time [Mutlukan et al., 1993, Trauzettel-Klosinski and Reinhard, 1998].

When examining fixation stability of hemianopic subjects compared to normal subjects, eye movements were measured during a long term fixation task. This required a stable and steady fixation for at least 17 min. A non-invasive eye tracker was used which is robust to decalibration and offered maximal comfort to the subjects during the test.

5.1.2 FCT Sensitivity and their Relation to Eye Movements

The FCT is technically less complex than eye tracking and is never prone to loose calibration over time. The FCT exploits inhomogeneities in the spatial distribution of color specific cones in the retina. The density of the cones is higher in central vision and decreases exponentially with higher eccentricity. This makes the perception of color a more difficult task in the periphery

in comparison to central vision. The FCT is based on the discrimination between two colors of almost equal luminance.

Such a behavior based eye movement assessment is widely used, for example, in the Tuebinger Automated Perimeter [Lachenmayr, 1992] as well as in other perimetric tasks [Kasten et al., 1997]. However, it is not clear to what extent peripheral color discrimination is appropriate for eye movement detection. It depends on size, color and luminance of the test stimulus [Abramov et al., 1991]. Therefore, the sensitivity of the color based FCT to eye movements was examined by presenting single and multiple targets in the periphery and by correlating the individual FCT detection rate with the respective eye movement measures obtained with an eyetracking system.

5.2 Methods

5.2.1 Subject Sample

In the first experiment, 13 hemianopic subjects (11 male and 2 female) with an age of 45.6 $\pm$ 10.3 years were examined. All hemianopic subjects participated in earlier studies and were familiar with the fixation task. The homonymous visual field defects were caused by cerebral lesions. The exclusion criteria of the hemianopic group were: known seizures of photosensitive epilepsy, evidence of spontaneous remission or unstable baseline, total blindness, central scotoma, inability to fixate or nystagmus, neglect, diplopia, glaucoma or other retinal disorders. For comparison, six normal subjects (4 male and 2 female) with an average age of 40.2 $\pm$ 11.0 years were examined who were age matched with the hemianopic subjects (exclusion criteria: visual acuity < 0.3). Normal subjects were inexperienced with the fixation task and completed a training test to become familiar with the task. Only the second test was used in the analysis. Hemianopic and normal subjects declared written consent prior to participation.

For the second part, which assessed the FCT, a larger group of 15 (11 female, 4 male) normal subjects with normal binocular vision and an average age of 30 $\pm$ 8 years was recruited. The inclusion criteria were normal results

in (i) visus testing, (ii) Ishihara-Tables and (iii) the 100 Hue-Test.

5.2.2   Fixation Task

The testing parameters of HRP are equivalent to those presented in chapter 2. All subjects were aware that eye movements were monitored during the HRP session.

5.2.3   Eye Movement Recording

A non-invasive eye tracker (Tobii 1750, Tobii Technology AB, Sweden) was used to record gaze positions from the left and the right eye. The signal was then averaged among both eyes to obtain maximal accuracy. A chin rest was used but the head clamp was not. The gaze data of the eye tracker is independent of body and head movements. The individual calibration of the eye tracker was completed in about 20 s. Manufacturer's specifications show that the long-term error is 1 ° with an accuracy of 0.5 ° and a spatial resolution of 0.25 ° at a recording frequency of 50 Hz [Tobii, 2004]. Additionally, the eye recognition rate (ERR, ratio of samples were both eyes were detected by the eye tracker) was recorded for each subject describing quantitatively how well the eye tracker recognized the pupils of the subject.

5.2.4   Measuring Horizontal Saccades

Horizontal saccades were measured during the entire fixation task. Four subjects (out of 13) were not considered in this analysis because their recording time was interrupted by voluntary breaks. In the remaining subjects, a total and continuous recording time of 17 min was obtained. The total recording time was divided into three periods, the first 5 min, the middle period and the last 5 min. The occurrence of horizontal saccades was than compared between the first and the last period. Furthermore, saccades are reported in higher temporal resolution of one minute intervals for the total testing time. The saccades and their respective amplitudes were extracted from the stream of gaze positions which was recorded by the eye tracker. A saccade is charac-

terized by the velocity of gaze positions. The specifications for the minimal (15 °/s) and maximal (1000 °/s) velocity limits were derived from measurements in normal and pathologic subjects [Boghen et al., 1974]. The end of a saccade is determined by: (i) the velocity is equal to 0 $°/s$ or (ii) the eye movement direction is inverted or (iii) the velocity exceeds 1000 °/s or (iv) the saccade duration exceeds 1000 ms (see Fig. 5.1). Saccades are reported in absolute numbers using four amplitude classes (2 $°-3$ $°, 3$ $°-4$ $°, 4$ $°-5$ $°$ and > 5 $°$). Absolute numbers of saccades are reported instead of relative numbers because small saccades outnumber large saccades by far and are possibly a source of sensor noise. If large saccades are expressed in percent of all observed saccades (including numerous small saccades produced by sensor noise) the numerical robustness of the analysis is at risk.
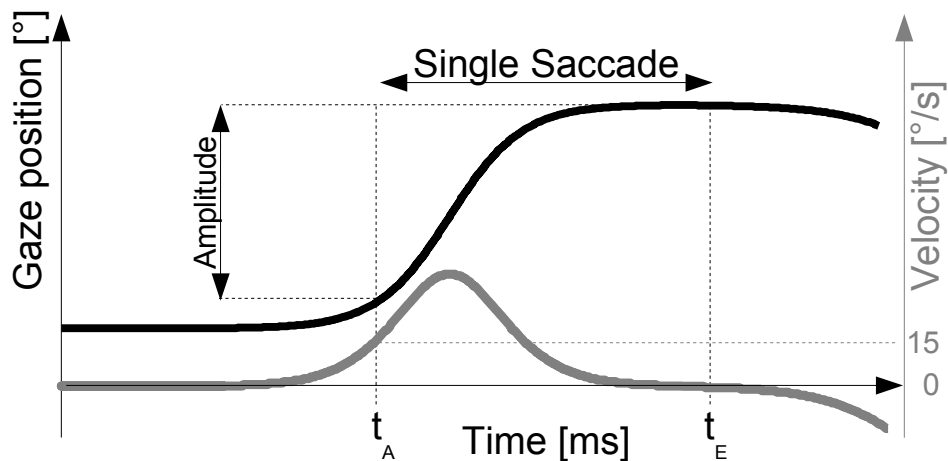

Figure 5.1: Scheme for describing the detection of saccades. The velocity of eye movements (gray color) is obtained from the gaze trajectory (black color) by computation of the first derivative. The beginning of a saccade is characterized by a velocity > 15 °/s (t_A). The end of the saccade is determined by a velocity equal to 0 °/s or when the movement direction is inverted (t_E). The saccade amplitude is the distance between the gaze positions measured at time t_A and t_E.

5.2.5 Measuring the FCT Sensitivity to Eye Movements

The ratio of detected color changes was measured in three experiments. In the first experiment, normal FCT detection ratios were collected during HRP (as described in section 2.2). The color of the fixation spot (size: 0.15 °) changed randomly from yellow (luminance: 138.3 cd/m^2) to light green (luminance: 133.5 cd/m^2, presentation time: 200 ms). The ratio of detected color changes was correlated with different eye movement features (mean and standard deviation of gaze positions). In the second experiment, a rectangular grid of 63 colored targets (interstimulus distance: 0.76 °, stimulus size: 0.4 °, total width: 9 °, total height: 6 °) was presented in eccentricities ≤ 5.4 ° (see Fig. 5.2, left). In this multiple target condition (MTC), one of the 63 targets was randomly selected changing the color from yellow (luminance: 138.3 cd/m^2) to light green (luminance: 133.5 cd/m^2, presentation time: 150 ms) with a difference in brightness < 5 cd/m^2. The trial was completed after the color change was presented in all of the 63 targets. It is of special interest at which eccentricity the ratio falls below a critical value of 90 % correct detections.

The third experiment was conducted using the single target condition (STC). The color was alternated from green (luminance: 133.5 cd/m^2) to yellow (luminance: 138.3 cd/m^2) at exactly one location (stimulus size: 0.4 °, presentation time: 150 ms, see Fig. 5.2, right). To assess the detection of color changes as a function of horizontal eccentricity, a fixation cross (luminance: 171.7 cd/m^2) was displaced on the horizontal meridian after each trial. After a total of 10 trials, the color change of the single target was assessed in horizontal eccentricities from ± 1 ° to ± 9 ° eccentricity in steps of 2 °. In the STC trials, above-threshold stimuli (size: 0.4 °, luminance: 75.6 cd/m^2) were presented as distractors for 150 ms at random position as they would usually appear in a perimetric testing. Subjects were instructed to respond to all distractors and all color changes. Each individual trial in the STC and MTC conditions lasted about 4.5 min. Fixational stability was assessed continuously throughout the trials by the Tobii eyetracking system. Trials with insufficient fixation (standard deviation of the mean gaze posi-
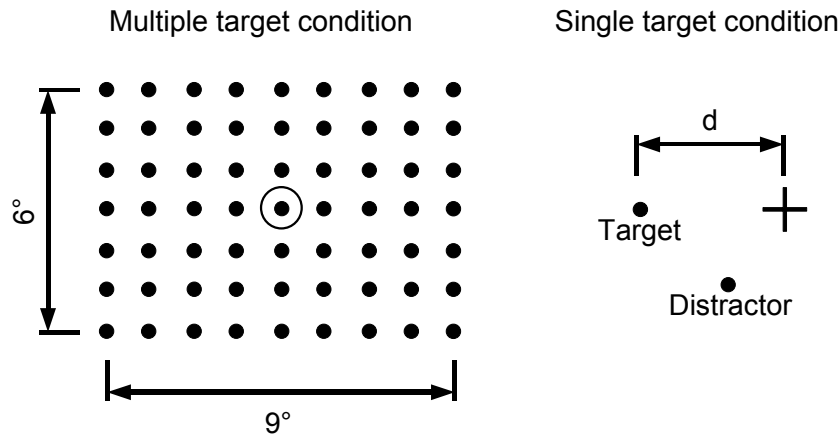
Figure 5.2: Assessment of colored based FCT. Left: In the multiple target condition, a rectangular grid of $7 \cdot 9 = 63$ stimuli in yellow color are continuously presented on the presentation screen covering an area of 9 ° times 6 °. The color changed from yellow to light green only at one randomly chosen grid position. The subject was instructed to fixate the center spot (marked with an open circle, the circle was not shown in the experiment). Right: In the single target condition the subject was instructed to fixate the cross. Distracting targets were shown at random positions within an area of $\pm$ 21 ° eccentricity. After an average of four distractor presentations, the FCT was presented resulting in a color change of the target from light green to yellow. After a total of 10 tests the fixation cross was located in 10 different horizontal distances d to the target at $\pm$ 1 °, $\pm$ 3 °, $\pm$ 5 °, $\pm$ 7 ° and $\pm$ 9 ° eccentricity.

tion > 2 °) or with an ERR < 80 % were not considered in the STC or MTC analysis. Both experiments were performed in a dark room after an adaptation time of several minutes. The background color of the screen was dark gray (luminance: 15 cd/m^2).

5.3   Results

The occurrence of horizontal saccades was measured during a long-term fixation task of 17 min. Saccades with smaller amplitudes were much more frequently observed than saccades with larger amplitudes (see Fig. 5.3). In order to report only robust results which are almost free from sensor noise

(about 1 °) or small eye movements (about 1 °), all saccades with amplitudes < 2 ° are omitted in the subsequent analysis.
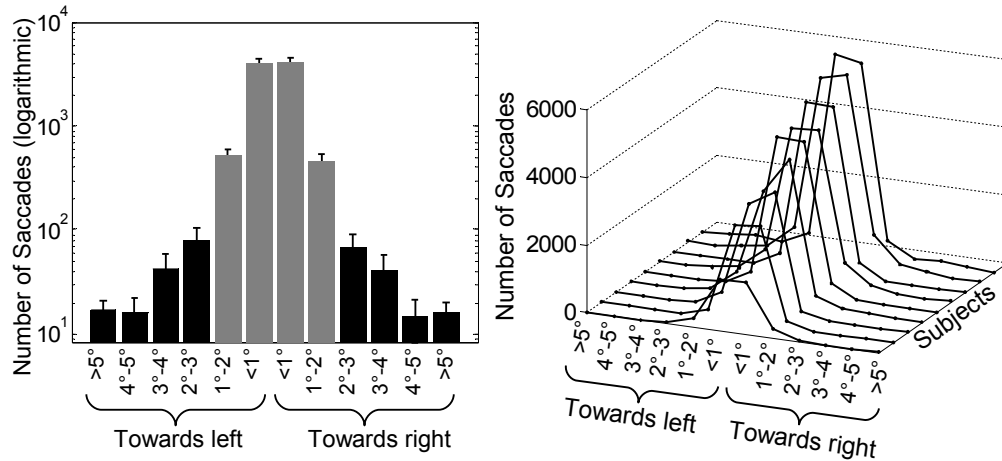

Figure 5.3: Total number of horizontal saccades (mean ± S.E.) with left and right direction during a 17 min recording in hemianopic subjects. Left: The number of saccades is shown in a logarithmic scale from all hemianopic subjects pooled into one set. Saccades < 2 ° (gray) are omitted in subsequent analysis because they are possibly confounded by sensor noise (about 1°) or are produced from small fixation instabilities (about 1 °). Right: The same data is shown in the left panel for each hemianopic subject.

During the fixation task, the average gaze position of normal and hemianopic subjects (n = 19) was 1.2 ° ± 0.6 ° horizontally and 0.7 ° ± 0.2 ° vertically. However, the standard deviation (0.9 ° ± 0.1 ° horizontally and 1.1 ° ± 0.2 ° vertically) of the average gaze position indicates that eye movements were made during the 17 min of uninterrupted testing time. Individual results of all measurements are provided in Table 5.1.

5.3.1 Long-term Measurement of Horizontal Saccades

When comparing the first with the last period of 5 min, an increase of eye movements was observed in both groups (see Fig. 5.4). Hemianopic subjects made on average significantly fewer eye movements ($\mu = 61 \pm 20.3$) within the first period than in the last period of measurement ($\mu = 100 \pm 28.3$). This is an increase by a factor of 1.6. In normal subjects, eye movements are

Table 5.1: Eye movement measurements and other variables from hemianopic (row 1-13) and normal subjects (row 14-19). The columns 'Saccad. First Perd.' and 'Saccad. Last Perd.' show the number of observed saccades $> 2°$ in the first and in the last 5 min of the HRP testing
.

Size De-fect	FCT	Vertical Fixation (mean±std) [°]	Horizontal Fixation (mean±std) [°]	Saccad. First Perd.	Saccad. Last Perd.	ERR [%]
166	96	0.61 ± 1.5	0.21 ± 1.14	83	214	71
235	88	0.92 ± 1.0	0.27 ± 0.6	20	76	66
90	100	0.04 ± 0.6	2.42 ± 0.7	0	44	82
85	100	2.79 ± 0.6	0.47 ± 0.6	0	9	63
99	100	1.45 ± 1.0	1.07 ± 0.9	112	154	81
196	80	0.12 ± 0.9	0.17 ± 0.9	*	*	67
157	97	0.28 ± 0.8	0.33 ± 0.5	*	*	93
104	81	1.18 ± 0.8	1.18 ± 0.45	13	27	80
219	90	1.04 ± 1.4	4.43 ± 1.8	*	*	90
187	99	0.89 ± 0.6	0.05 ± 0.9	171	240	77
377	75	0.14 ± 1.2	10.77 ± 1.8	116	104	89
38	98	0.9 ± 0.8	0.61 ± 0.7	49	31	82
57	95	0.55 ± 0.6	0.61 ± 1.0	*	*	82
#	100	0.54 ± 1.0	-0.06 ± 0.9	13	52	86
#	99	0.37 ± 0.5	0.52 ± 0.5	6	12	97
#	100	0.49 ± 0.8	-0.27 ± 0.7	10	17	96
#	100	-0.81 ± 4.0	1.1 ± 1.4	22	601	31
#	100	1.92 ± 1.3	-1.42 ± 1.4	32	152	78
#	100	0.43 ± 0.6	-0.22 ± 0.4	1	15	88

*: subject was not considered in this part of the analysis
#: not applicable

less but not significantly less frequent in the first period ($\mu = 14 \pm 4.6$) when compared to the last period ($\mu = 142 \pm 94.4$). This is an increase by a factor of 10.1. The increase of saccades as a function of testing time is shown in Fig. 5.5.
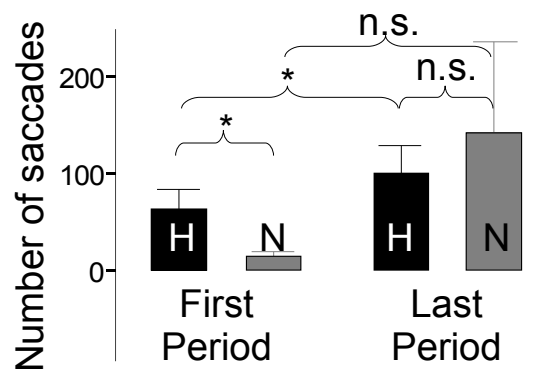

Figure 5.4: Number of saccades (mean $\pm$ S.E.) comparing the first and last period of 5 min recording in hemianopic (H) and in normal (N) subjects.
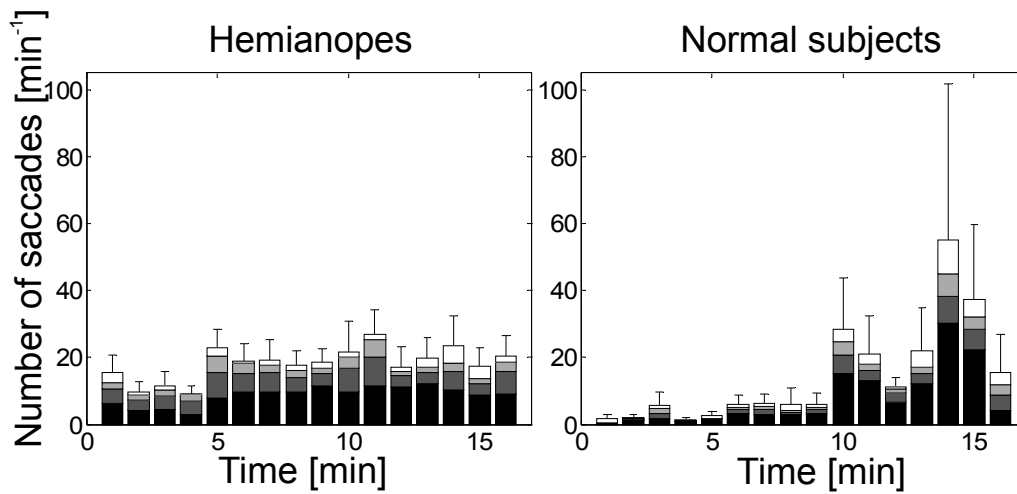
Figure 5.5: Number of horizontal saccades (mean ± S.E.) as a function of testing time in the group of hemianopic (left panel) and in normal subjects (right panel) separated in the amplitudes: 2 °-3 ° (black), 3 °-4 ° (dark gray), 4 °-5 ° (light gray), > 5 ° (white). The amplitude is shown irrespective of saccade direction or fixation position.

5.3.2 Examination of the Color Based FCT

A medium correlation (Pearson) between the number of detected fixation catch trials and (i) the standard deviation horizontal/vertical ($\rho = -0.36$, n.s.) or (ii) with the occurrence of eye movements ($\rho = -0.24$, n.s.) was observed. Reasonably, the correlation is negative indicating that more fixation catch trials are detected if less eye movements are made. If the sensitivity of the color based FCT is assessed as a function of eccentricity, less FCTs are detected in the periphery (see Fig. 5.6). The criteria of significant detection loss (> 10 %) is found at about 2 ° for the MTC and about 4 ° for the STC.
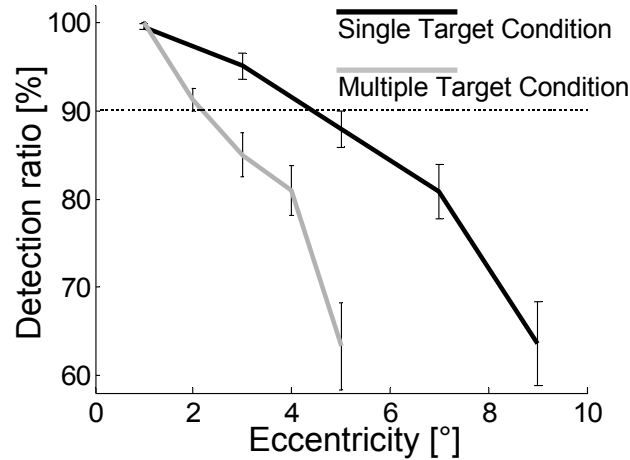
Figure 5.6: Ratio (mean ± S.E.) of detected color changes in the single- (STC) and multiple-target-condition (MTC) as a function of eccentric stimulus presentation. The critical detection loss (dotted line) is observed at an eccentricity of 2 ° (MTC) and 4 ° (STC), respectively.

5.4 Discussion

5.4.1 Saccades of Amplitudes > 2 °

Most of the time, all subjects made eye movements inside the central area of 2 ° around the fixation point. This range was exceeded by the hemianopic and normal group in 32 % and 20 % of the time, respectively in a continuous recording time of 17 min. This is in concordance with the observation that normal subjects made fixations with amplitudes > 2 ° in 18 % of the total testing time of 15 min [Eizenman et al., 1992]. Eye movements at larger amplitudes were observed to be more frequent in eye movements studies with longer testing time [Demirel and Vingrys, 1994]. It is not surprising that eye movements are less often observed if the testing time is 4 min or less as in [Kasten et al., 2006].

5.4.2 Increase of Eye Movements over Time

It was observed that the number of saccades increased over time in normal subjects by a factor of 10 and in hemianopic subjects by a factor of

1.6 (significant increase). The task of detecting above-threshold stimuli was probably too easy for healthy subjects as they showed fatigue earlier than the group of hemianopic subjects who were more experienced with the fixation task. Others have reported similar results: The standard deviation of the fixation position (a measure of the magnitude of eye movements) increased every minute by 0.56 °, resulting in a standard deviation of 2.55 ° after 3 min recording time [Eizenman et al., 1992]. Normal subjects had a steady fixation instability of > 5 ° during the recording time of 30 min [Demirel and Vingrys, 1994]. Fatigue (which is identified by slow drifts of fixation followed by fast corrective saccades) is a possible reason for the occurrence of eye movements [Eizenman et al., 1992]. Further reasons are boredom, increasing distraction of randomly appearing peripheral stimuli and blanking out (disappearance) of the target due to sensory adaptation [Demirel and Vingrys, 1994]. Furthermore, subjects reported a 'washing out' effect of the background after some minutes, resulting in a 'hypnotizing sensation' [Wong et al., 1995].

During fixation tasks, small eye movements were considered as part of the natural scanning behavior in normal subjects [Trauzettel-Klosinski and Reinhard, 1998]. Fixation instability of 1 ° is assumed to be 'normal' in fixation tasks [Jamara et al., 2003]. Accordingly, it was reported that 20 % of inexperienced normal subjects had significant fixation loss during Humphrey perimetry (which is a diagnostic system similar to HRP) [Eizenman et al., 1992]. This suggests that even well trained and motivated normal subjects are not able to maintain accurate fixation if the testing time exceeds a duration of 5 min [Demirel and Vingrys, 1994].

5.4.3 Eye Movements in Hemianopic and Normal Subjects

Within the first period of 5 min, hemianopic subjects made on average significantly more eye movements (> 2 °) than normal subjects (by a factor of 4.4). Hemianopic subjects were quite experienced with the fixation task vs. the normal subjects had practiced only for 15-20 min prior to the experiment.

The observation that hemianopic subjects made more eye movements than normal subjects could be explained by compensatory eye movements which help hemianopic subjects to increase their field of view. It was observed that the magnitude of eye movements is higher in subjects with larger scotoma especially in the absence of macular sparing [Trauzettel-Klosinski and Reinhard, 1998]. The fluctuation of perimetric results was more pronounced in subjects with greater scotoma size which was attributed to eye movements [Demirel and Vingrys, 1994]. This is in agreement with an observed medium correlation ($\rho = 0.5$, n.s.) between the size of the scotoma and the number of observed saccades during the fixation task in hemianopic subjects.

It was observed elsewhere that compensatory eye movements are more pronounced if the subject had a longer time to learn a compensation strategy [Pambakian et al., 2000] and that such a compensatory behavior is even induced in normal subjects when an artificial hemifield-scotoma is simulated. Normal subjects needed more time for perception tasks, made more errors and made more fixations equivalent to the scanning behavior of hemianopic subjects [Tant et al., 2002].

5.4.4 Color Based Fixation Catch Trial

The color based FCT is a less expensive and technically less complex alternative to eye tracking in tasks where small fixation instability is tolerated. The results of the FCT experiment are probably influenced by miss-perceptions because subjects reported symptoms comparable to the scintillating grid illusion [VanRullen and Dong, 2003]. The criteria for critical detection loss ($> 10\%$) was observed at a distance between fixation and catch trial of more than $2\,°$ in the MTC and $4\,°$ in the less difficult STC. The examined FCT is based on the inhomogeneity in the spatial distribution of color specific cones. The number of cones drops below 50% at an eccentricity $> 1.75\,°$ [Curcio et al., 1991, Osterberg, 1935]. Accordingly, it was shown that the color sensitivity deteriorates in the periphery [Newton and Eskew, 2003]. To achieve the same color appearance (hue), the stimulus size in the periphery at $10\,°$ eccentricity must increase by a factor of about 27 for the green color and

by a factor of about 5 for the yellow color [Abramov et al., 1991, Iivanainen and Rovamo, 1992]. If the size of the peripheral stimulus is $< 2\,°$ the color discrimination thresholds are different between periphery and fovea [Nagy and Doyal, 1993, Parry et al., 2006] which is exploited by the FCT. The FCT is obviously not strongly associated with the occurrence of eye movements. Both correlations of individual FCT values with standard deviation of mean gaze position ($\rho = -0.36$) and number of saccades ($\rho = -0.24$) are only medium and not significant. This suggests that the FCT measure is probably influenced by perceptual factors such as fatigue and alertness (even if the subjects fixate properly, they could miss the FCT presentation) or by motoric aspects (although subjects have perceived the catch trial they were not able to press the button within the valid time interval of 1000 ms).

5.5 Conclusion

The criteria of sufficient fixation stability was set to 10 % and charts with a FCT detection loss $> 10\%$ were not used for building the TOPM or constructing the features. As it was shown, this criteria of sufficient fixation stability was on average achieved by the normal group in the MTC experiment even if the stimulus was presented at an eccentricity of $2\,°$. Although the FCT could be more difficult for hemianopic subjects (which were not tested in the MCT experiment), it is not possible to conclude that all diagnostic charts are free of eye movement artefacts of amplitudes $\leq 2\,°$ wheras the probability is small that steady eye movements $> 4\,°$ were not detected with the FCT paradigm.

If eye tracking equipment is available, it should be used rather than a behaviorally based methods like the FCT, even though it is more complex and in some systems prone to decalibration. State of the art eye trackers are comfortable for the subjects and have a small margin of error. A behavior based control like the FCT should be used only in tasks where eye movements with amplitudes ($\leq 4\,°$) cannot interfere with the test results. Then, FCT provides an approximate measure of large eye movements ($> 4\,°$).

As it was reported in the introduction, most studies limit the testing time

to a couple of minutes. According to results from this study, this practice is strongly recommended because a significant increase of saccades was observed in hemianopic subjects between the first and last period of testing (first period: 1-5 min, last period: 12-16 min). For the field of HRP which requires a testing time of more than 20 min, a combination of HRP and eye tracking technology is recommended, so that appropriate actions can be undertaken when the fixation stability begins to deteriorate. Or it is recommended to split the 20 min testing into sessions with short testing time with pauses in between.

Chapter 6

Summary and Outlook

A treatment outcome prediction model was presented for patients with damage to the visual system. It allows a prognosis of restoration to be made regarding the improvement of vision or lack thereof in patients with visual field damage. As the basis of the prediction model, features were developed which describe specific properties of the visual field diagnostic charts. The features incorporate *a priori* knowledge and address several topics in the broad field of vision plasticity. As this Thesis shows, the features were found to be statistically associated with the treatment outcome and are therefore appropriate to be used in the treatment outcome prediction model (TOPM). The TOPM core architecture is a Self-Organizing-Map which offers methods for prediction as well as for data visualization. The model was applied on pre-treatment diagnostic charts from 52 patients with visual field damage who have finished the treatment. Only diagnostic charts were used which were minimally confounded by eye movements as it was assessed by an eye tracker study. Finally, cross validation was used to derive error measures showing that the TOPM is able to predict potential areas of improvement after vision stimulation treatment.

Several applications exist for the presented TOPM. If the TOPM is applied on pre-treatment diagnostic charts of appropriate patients, a better therapy may be planned by using the predicted treatment outcome. In case the prognosis shows only limited improvements, an alternative treatment

may be suggested. If the prognosis shows improvements of specific areas in the visual field of the patient, the treatment may be located to areas of greatest recovery potential, neglecting areas without predicted improvement. The features were developed by using knowledge from visual system plasticity. If those features are used to examine diagnostic charts in prospective and retrospective studies, hypothesis about the mechanisms of vision restoration can be examined by comparing the feature values with multiple clinical parameters. The integration of new features into the open architecture is possible with little effort such that new hypothesis can be evaluated by using the available diagnostic data.

In summary, the TOPM may be used to enhance the therapy efficiancy and by supporting the analysis of the mechanisms of vision restoration it provides a technological platform for improving restoration of visual disorders caused by brain damage.

Appendix A

Examples of Diagnostic Charts

The following pages show the diagnostic charts (from hemianopic subjects, n = 23) which were used in the feature evaluation study from chapter 3.6. Baseline charts (left panel) and post-treatment detection charts (middle panel) were assessed before and after 6 months of VRT. These charts show intact areas (white), absolute defect areas (black) and residual areas (shades of gray). The right panel displays the dynamic charts which were calculated by comparing the post-treatment charts with the baseline charts. These charts display the occurrence of improvement (hot spots, '+') or lack of improvement (cold spots, '.'). Spots which were healthy at baseline are not shown. The vertical and horizontal axis is shown in vertical and horizontal visual degree. The charts are ordered with respect to the number of observed hot spots. The charts in the first row correspond to the subject with strongest improvement after VRT.
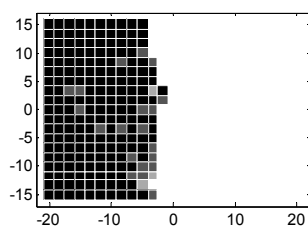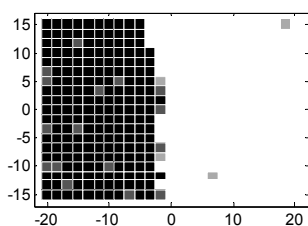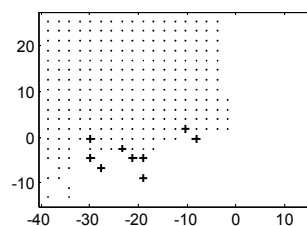
Baseline chart   Post-treatment chart   Dynamic chart

Baseline chart   Post-treatment chart   Dynamic chart
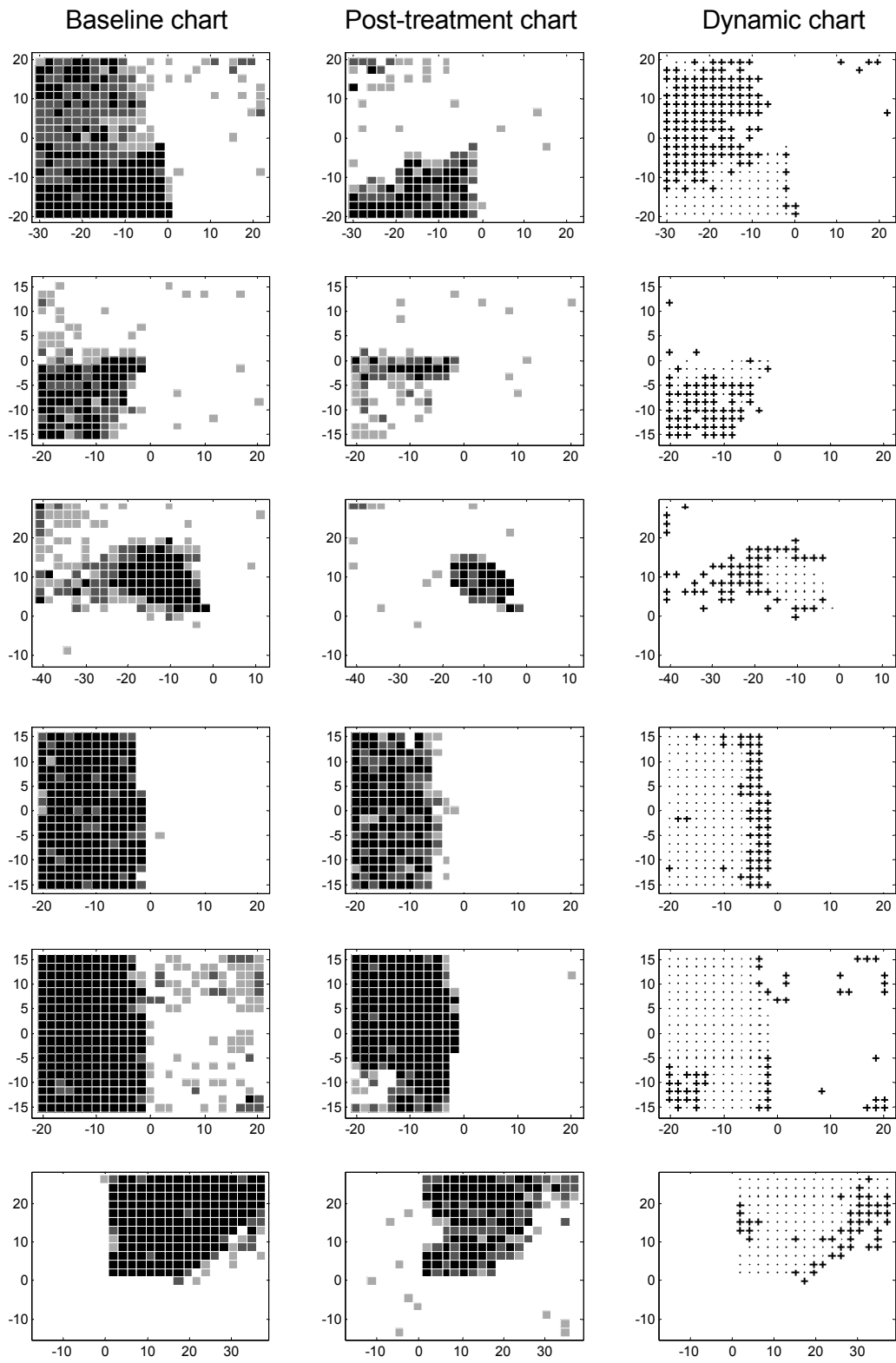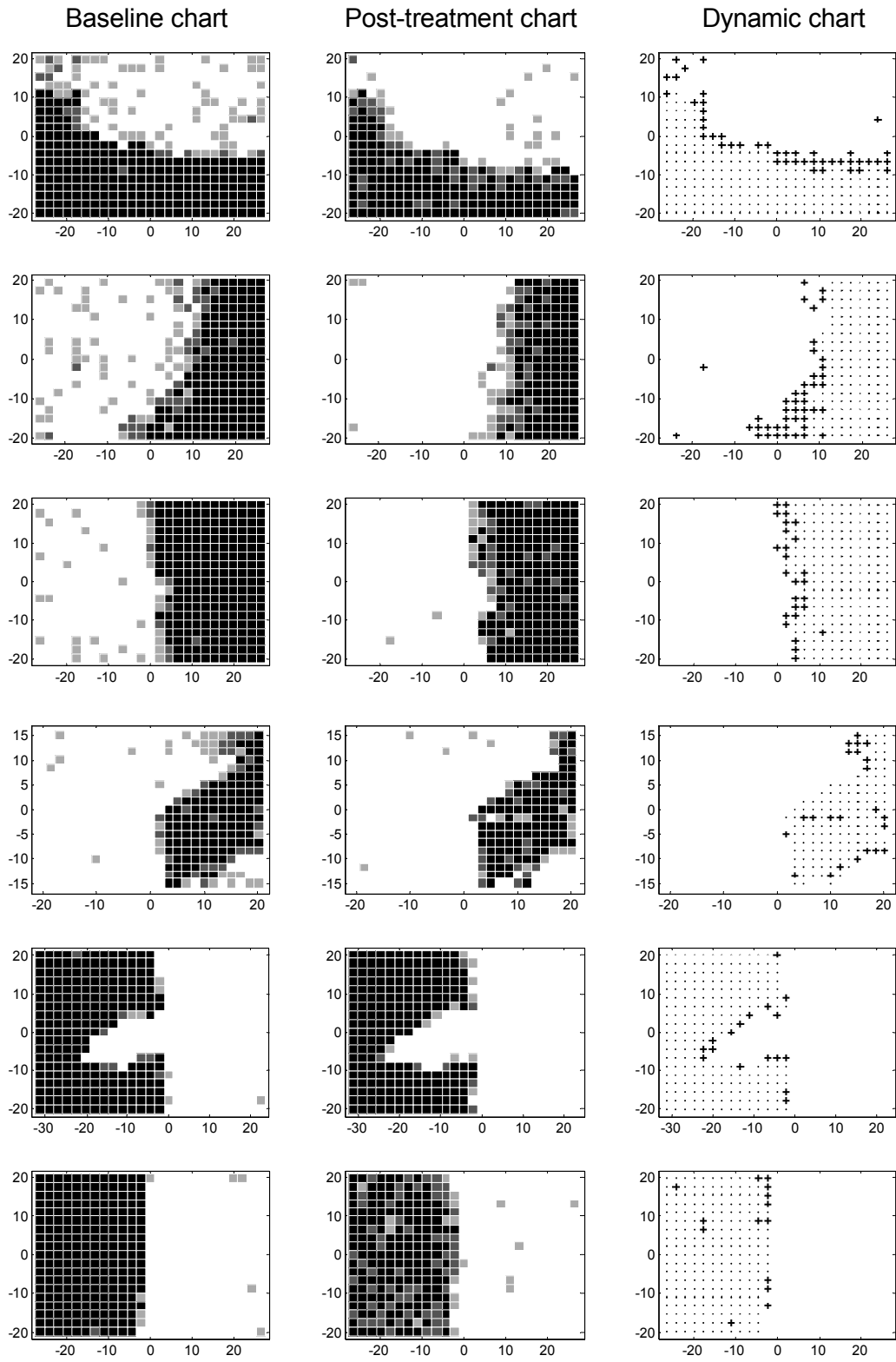
Baseline chart   Post-treatment chart   Dynamic chart
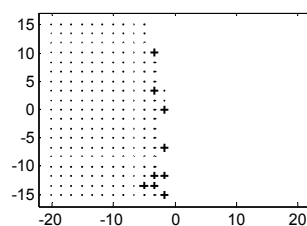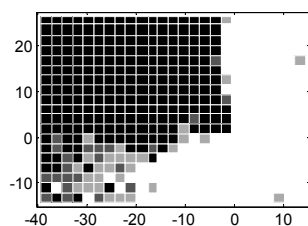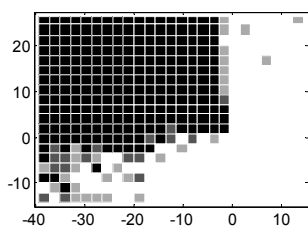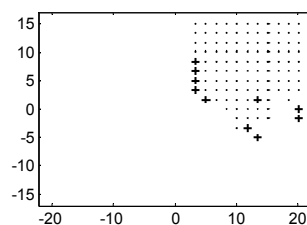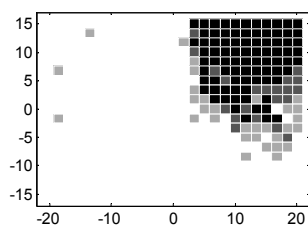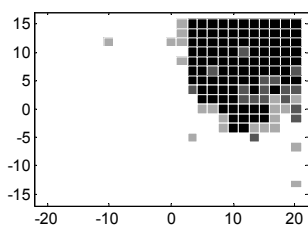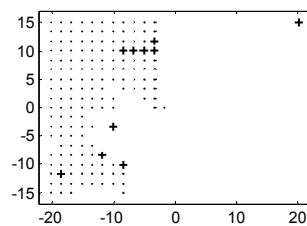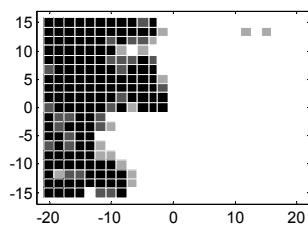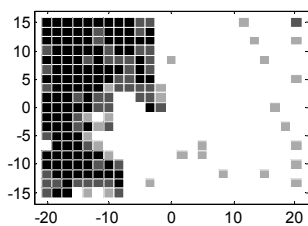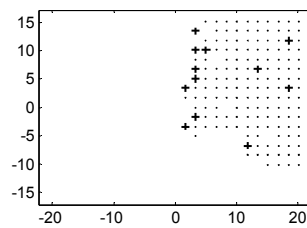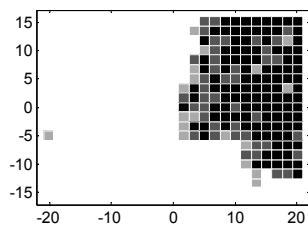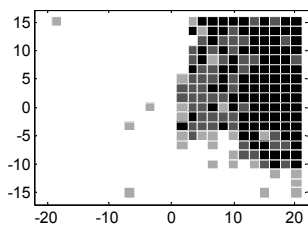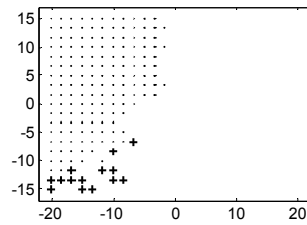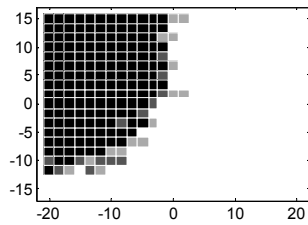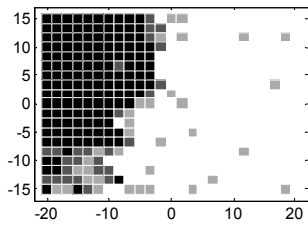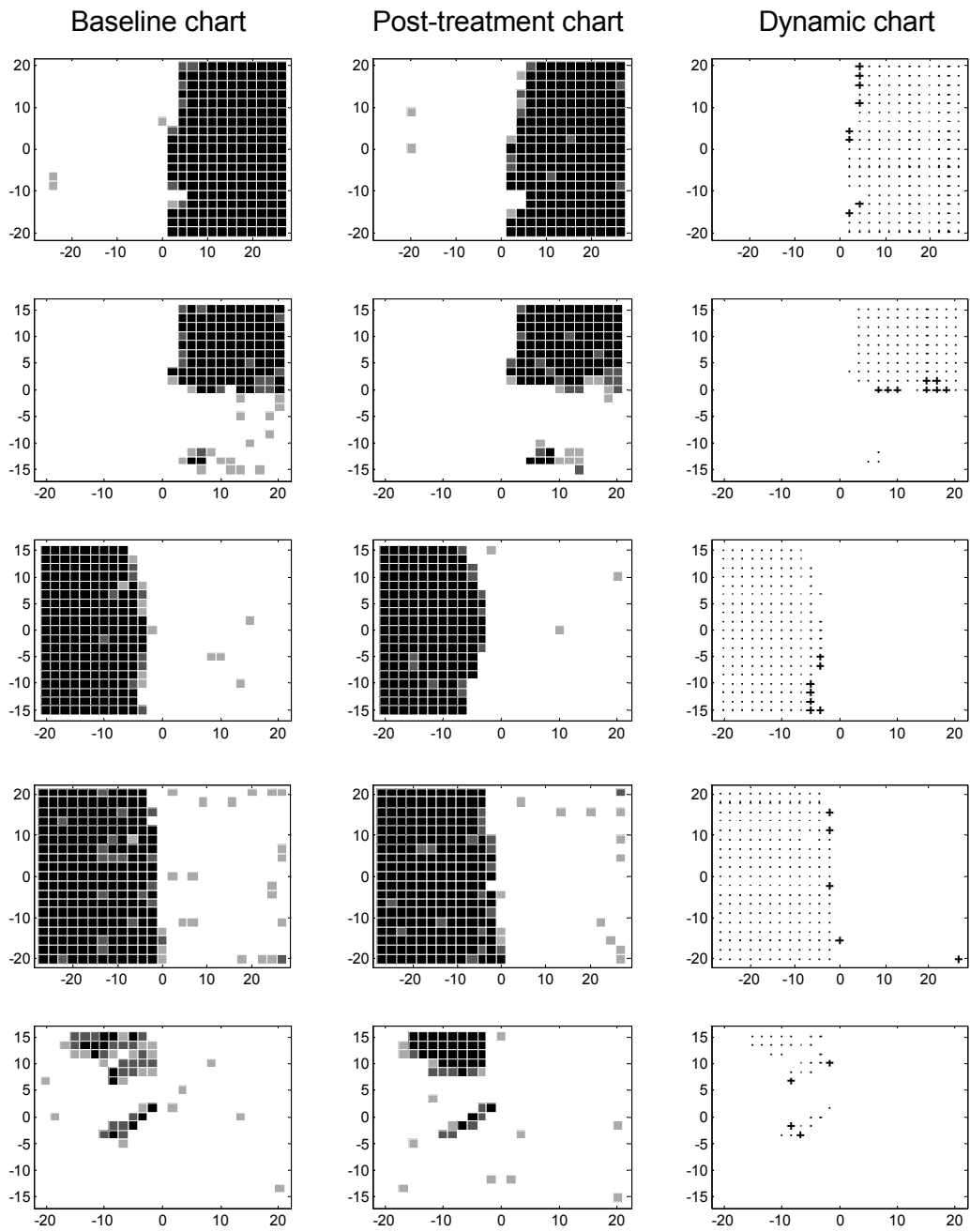
Appendix B

Extended Descriptive Statistics

Table B.1: Descriptive statistics of spots from 52 subjects which were pooled into one set. The statistic results does not consider patient affiliation (see chapter 3.6 for a subject independent analysis). The Pearson correlation coefficient (ρ) between features and treatment outcome, the respective group means (μ) and standard deviations (δ) are shown for hot (n=1689) and cold spots (n=7026).

| Feature	ρ	cold spot			hot spot		
		μ	$\pm$	δ	μ	$\pm$	δ
Neigh. Activity	+0.49	0.06	$\pm$	0.11	0.25	$\pm$	0.22
Neigh. Homogeneity	+0.42	0.09	$\pm$	0.13	0.25	$\pm$	0.14
Residual Function	+0.38	0.02	$\pm$	0.08	0.12	$\pm$	0.16
Residual Area	+0.26	41.60	$\pm$	29.6	63.05	$\pm$	40.00
Dist. to Scotoma [mm]	-0.24	7.31	$\pm$	5.3	4.00	$\pm$	5.60
Defect Area	-0.24	181.9	$\pm$	47.8	150.4	$\pm$	65.21
Hemianopia	-0.18	0.51	$\pm$	0.31	0.36	$\pm$	0.33
Border Diffuseness	+0.17	0.32	$\pm$	0.17	0.39	$\pm$	0.18
Reaction Time	+0.05	442.7	$\pm$	47.91	449.0	$\pm$	41.80
Quadrantanopia	+0.04	0.29	$\pm$	0.23	0.32	$\pm$	0.21
Horizontal Position [°]	+0.03	-5.00	$\pm$	15.4	-3.74	$\pm$	14.04
Vertical Position [°]	-0.01	0.73	$\pm$	11.11	0.37	$\pm$	11.05

Bibliography

I. Abramov, J. Gordon, and H. Chan. Color appearance in the peripheral retina: effects of stimulus size. *J Opt Soc Am A*, 8(2):404–14, 1991.

A. Abu-Hanna and P. J. F. Lucas. Editorial: Prognostic models in medicine. *Methods of Information in Medicine*, 2001.

R. Agrawal, R. Imielinski, and A.N. Swami. Mining association rules between sets of items in large databases. In Peter Buneman and Sushil Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C., 26–28 1993.

E. Alexopoulos, G. Dounias, and K. Vemmos. Medical diagnosis of stroke using inductive machine learning. In *The Workshop on Machine Learning in Medical Applications, Advanced Course in Artificial Intelligence, July 5-16*, Chania, Greece, 1999.

T. J. Andrews, S. D. Halpern, and D. Purves. Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract. *J Neurosci*, 17(8):2859–68, 1997.

M. Balasubramanian, J. Polimeni, and E. L. Schwartz. The v1 -v2-v3 complex: quasiconformal dipole maps in primate striate and extra-striate cortex. *Neural Netw*, 15(10):1157–63, 2002.

J. S. Barton and M. Benatar. *Field of Vision: a manual and atlas of perimetry.* Humana Press, Totowa, NJ, 2003.

A. U. Bayer and C. Erb. Short wavelength automated perimetry, frequency doubling technology perimetry, and pattern electroretinography for prediction of progressive glaucomatous standard visual field defects. *Ophthalmology*, 109(5):1009–17, 2002.

D. Boghen, B. T. Troost, R. B. Daroff, L. F. Dell'Osso, and J. E. Birkett. Velocity characteristics of normal human saccades. *Invest Ophthalmol*, 13 (8):619–23, 1974.

C. Borgelt and R. Kruse. Learning probabilistic and possibilistic networks: Theory and applications. In *Proc. 7th International Fuzzy Systems Association World Congress (IFSA'97)*, volume 1, pages 19–24, 1997.

C. Borgelt and R. Kruse. *Graphical Models - Methods for Data Analysis and Mining*. Wiley, Chichester, 2002.

D. V. Buonomano and M. M. Merzenich. Cortical plasticity: from synapses to maps. *Annu Rev Neurosci*, 21:149–86, 1998.

L.M. Chalupa and J.S. Werner. *The Visual Neurosciences*. Bradford Book, 2004.

Y. M. Chino, 3rd Smith, E. L., J. H. Kaas, Y. Sasaki, and H. Cheng. Receptive-field properties of deafferentated visual cortical neurons after topographic map reorganization in adult cats. *J Neurosci*, 15(3 Pt 2): 2417–33, 1995.

C. A. Curcio, K. A. Allen, K. R. Sloan, C. L. Lerea, J. B. Hurley, I. B. Klock, and A. H. Milam. Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J Comp Neurol*, 312(4):610–24, 1991.

F. Klawonn D. Nauck and R. Kruse. *Foundations of Neuro-Fuzzy Systems*. Wiley, Chichester, United Kingdom, 1997.

A. Das and C. D. Gilbert. Long- range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. *Nature*, 375:780–84, 1995.

S. Demirel and A.J. Vingrys. Eye movements during perimetry and the effect that fixation instability has on perimetric outcomes. *J Glaucoma*, 3:28–35, 1994.

R. F. Dougherty, V. M. Koch, A. A. Brewer, B. Fischer, J. Modersitzki, and B. A. Wandell. Visual field representations and locations of visual areas v1/2/3 in human visual cortex. *J Vis*, 3(10):586–98, 2003.

R. O. Duncan and G. M. Boynton. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4): 659–71, 2003.

M. Eizenman, G. E. Trope, M. Fortinsky, and P. H. Murphy. Stability of fixation in healthy subjects during automated perimetry. *Can J Ophthalmol*, 27(7):336–40, 1992.

T. Fawcett. Roc graphs: Notes and practical considerations for researchers. Technical report, 2004.

U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *Ai Magazine*, 17:37–54, 1996.

D. V. Giannikopoulos and U. T. Eysel. Dynamics and specificity of cortical map reorganization after retinal lesions. *Proc Natl Acad Sci U S A*, 103 (28):10805–10, 2006.

C. D. Gilbert. Neuronal dynamics and perceptual learning. *Current Biology*, 4(7):627–29, 1994.

C. D. Gilbert and T. N. Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. of Neuroscience*, 9 (7):2432–2442, 1989.

C. D. Gilbert and T.N. Wiesel. Receptive field dynamics in adult primary visual cortex. *Nature*, 356(6365):150–2, 1992.

C. D. Gilbert, M. Ito, M. Kapadia, and G. Westheimer. Interactions between attention, context and learning in primary visual cortex. *Vision Res*, 40 (10-12):1217–26, 2000.

S. Goodall, J. A. Reggia, Y. Chen, E. Ruppin, and C. Whitney. A computational model of acute focal cortical lesions. *Stroke*, 28(1):101–9, 1997.

C. S. Gray, J. M. French, D. Bates, N. E. Cartlidge, G. S. Venables, and O. F. James. Recovery of visual fields in acute stroke: homonymous hemianopia associated with adverse prognosis. *Age Ageing*, 18(6):419–21, 1989.

M. Green and M. Ohlsson. Comparison of standard resampling methods for performance estimation of artificial neural network ensembles. In *The Third International Conference on Computational Intelligence in Medicine and Healthcare, July 25-27*, Plymouth, UK, 2007.

T. Guenther, J. Gudlin, B. A. Sabel, and I. Mueller. Data infrastructure of prediction models for the diagnosis and treatment outcome in patients with visual system dysfunction. In *Proceedings of The Third International Conference on Computational Intelligence in Medicine and Healthcare, July 25-27*, Plymouth, UK, 2007a.

T. Guenther, M. Preuss, and I. Mueller. Coordinate-transformation module connecting the visual cortex and the visual space. In *Proceedings of The International Conference on Artificial Intelligence and Pattern Recognition, July 9-12*, Orlando, USA, 2007b.

T. Guenther, A. Lipfert, I. Mueller, E. Kasten, and B. A. Sabel. Long-term eye movement assessment in a perimetric fixation task. subm.a.

T. Guenther, I. Mueller, R. Kruse, and B. A. Sabel. Treatment outcome prediction model of visual field plasticity using self-organizing-maps. subm.b.

T. Guenther, I. Mueller, M. Preuss, F. Wolf, and B. A. Sabel. Local topographic influences on vision restoration hot spots after brain damage. subm.c.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157 – 1182, 2003.

J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2006.

D. J. Hand, M. Heikki, and S. Padhraic. *Principles of Data Mining*. The MIT Press, 2001.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2002.

S. J. Heinen and A. A. Skavenski. Recovery of visual responses in foveal v1 neurons following bilateral foveal lesions in adult monkey. *Exp Brain Res*, 83(3):670–4, 1991.

L. Henriksson, A. Raninen, R. Nasanen, L. Hyvarinen, and S. Vanni. Training-induced cortical representation of a hemianopic hemifield. *J Neurol Neurosurg Psychiatry*, 78(1):74–81, 2007.

R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, Volume 11, Number 1 / April, 1993: 63–90, 1993.

A. Iivanainen and J. Rovamo. The effects of colour adaptation and stimulus size on white perception as a function of eccentricity in man. *Vision Res*, 32(6):1131–5, 1992.

A.K. Jain and R. C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. ISBN 0-13-022278-X.

R. J. Jamara, F. Van De Velde, and E. Peli. Scanning eye movements in homonymous hemianopia documented by scanning laser ophthalmoscope retinal perimetry. *Optom Vis Sci*, 80(7):495–504, 2003.

G.H. John, R. Kohavi, and K. Pfleger. Irrelevant features and the subset selection problem. *International Conference on Machine Learning*, pages 121–129, 1994.

C. A. Johnson, A. J. Adams, E. J. Casson, and J. D. Brandt. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Arch Ophthalmol*, 111(5):645–50, 1993.

K. Jong, J. Mary, A Cornuéjols, E. Marchiori, and M. Sebag. Ensemble feature ranking. In *Lecture Notes In Computer Science;*, volume 3202, pages 267 – 278. Pisa, 2004.

L. Julkunen, O. Tenovuo, S. Jaaskelainen, and H. Hamalainen. Rehabilitation of chronic post-stroke visual field defect with computer-assisted training: a clinical and neurophysiological study. *Restor Neurol Neurosci*, 21(1-2):19–28, 2003.

L. Julkunen, O. Tenovuo, V. Vorobyev, J. Hiltunen, M. Teras, S. K. Jaaskelainen, and H. Hamalainen. Functional brain imaging, clinical and neurophysiological outcome of visual rehabilitation in a chronic stroke patient. *Restor Neurol Neurosci*, 24(2):123–32, 2006.

E. R. Kandel and J. H. Schartz. *Neurowissenschaften. Eine Einführung.* Spektrum Akademischer Verlag, 1996.

A. Karlsson. Introduction to linear regression analysis. *Journal Of The Royal Statistical Society Series A*, 127(3):856–857, 2007.

E. Kasten and B. A. Sabel. Visual field enlargement after computer training in brain- damaged patients with homonymous deficits: an open pilot trial. *Restorative Neurology and Neuroscience*, 8:113–27, 1995.

E. Kasten, H. Strasburger, and B. A. Sabel. Programs for diagnosis and therapy of visual field deficits in vision rehabilitation. *Spat Vis*, 10(4): 499–503, 1997.

E. Kasten, S. Wuest, W. Behrens-Baumann, and B. A. Sabel. Computer-based training for the treatment of partial blindness. *Nat Med*, 4(9):1083–7, 1998a.

E. Kasten, S. Wuest, and B. A. Sabel. Residual vision in transition zones in patients with cerebral blindness. *J Clin Exp Neuropsychol*, 20(5):581–98, 1998b.

E. Kasten, D. A. Poggel, E. Müller-Oehring, J. Gothe, T. Schulte, and B. A. Sabel. Restoration of vision ii : Residual functions and training-induced visual field enlargement in brain-damaged patients. *Restorative Neurology and Neuroscience*, 15:273–287, 1999.

E. Kasten, D. A. Poggel, and B. A. Sabel. Computer-based training of stimulus detection improves color and simple pattern recognition in the defective field of hemianopic subjects. *J Cogn Neurosci*, 12(6):1001–12, 2000.

E. Kasten, U. Bunzenthal, and B. A. Sabel. Visual field recovery after vision restoration therapy (vrt) is independent of eye movements: an eye tracker study. *Behav Brain Res*, 175(1):18–26, 2006.

T. Kohonen. The self-organizing map. In *Proceedings of the IEEE, September*, volume Vol. 78, No. 9, pages 1464–1480, Boston, 1990.

R. Kruse and C. Borgelt. Data mining with graphical models. *Proc. 12th Int. Symp. Computer Science for Environmental Protection*, 1:17–30, 1998.

B. J. Lachenmayr. *Perimetrie*. Thieme, Stuttgart, 1992.

P. Langley. Selection of relevant features in machine learning. *in AAAI Fall Symposium on Relevance*, pages 140–144, 1994.

J. B. Levitt and J. S. Lund. The spatial extent over which neurons in macaque striate cortex pool visual signals. *Vis Neurosci*, 19(4):439–52, 2002.

H. Liu and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, Norwell, 1998.

P. J. Lucas and A. Abu-Hanna. Prognostic methods in medicine. *Artif Intell Med*, 15(2):105–19, 1999.

G. D. Magoulas and A. Prentza. Machine learning in medical applications. In *Workshop on Machine Learning in Medical Applications.* Advanced Course on Artificial Intelligence, 2001.

R. Malach, Y. Amir, M. Harel, and A. Grinvald. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc Natl Acad Sci U S A*, 90(22):10469–73, 1993.

N. L. Marks and J. B. Hellige. Effects of bilateral stimulation and stimulus redundancy on interhemispheric interaction. *Neuropsychology*, 13(4):475–87, 1999.

R. S. Marshall, J. J. Ferrera, A. Barnes, X. Zhang, A. O'Brien K, M. Chmayssani, J. Hirsch, and R. M. Lazar. Brain activity associated with stimulation therapy of the visual borderzone in hemianopic stroke patients. *Neurorehabil Neural Repair*, 2007.

S. Marsland and I. Buchan. Clinical quality needs complex adaptive systems and machine learning. *Medinfo*, 11(Pt 1):644–7, 2004.

H. Mast, M. Preuss, and B. A. Sabel. Vision restoration therapy vs. exploration training in patients with visual field loss. In *From neural plasticity to neurorehabilitation, June 26-30*, Zurich, Switzerland, 2006.

T. M. Mitchell. *Machine Learning.* McGraw-Hill, New York, 1997.

I. Mueller, D. A. Poggel, E. Kasten, and B. A. Sabel. Predicting the outcome of visual restitution training: A retrospective clinical study on visual system plasticity. In *30th Annual Meeting Society for Neuroscience, November 2-7*, Orlando, USA, 2002. Society for Neuroscience.

I. Mueller, D. A. Poggel, S. Kenkel, E. Kasten, and B. A. Sabel. Vision restoration therapy after brain damage: Subjective improvements of activities of daily life and their relationship to visual field enlargements. *Visual Impairment Research*, 5(3):157–178, 2003.

I. Mueller, H. Mast, and B. A. Sabel. Recovery of visual function after brain injury: a large-sample study using vision restoration therapy. subm.

E. Mutlukan, B. E. Damato, and S. M. Tavadia. Clustering of fixation targets in multifixation campimetry. *Eye*, 7 (Pt 1):131–3, 1993.

A. L. Nagy and J. A. Doyal. Red-green color discrimination as a function of stimulus field size in peripheral vision. *J Opt Soc Am A*, 10(6):1147–56, 1993.

J. R. Newton and J. Eskew, R. T. Chromatic detection and discrimination in the periphery: a postreceptoral loss of color sensitivity. *Vis Neurosci*, 20(5):511–21, 2003.

A. Nürnberger, W. Pedrycz, and R. Kruse. Neural network approaches. pages 304–317, 2002.

G. Osterberg. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica*, 6(1):11–97, 1935.

A. L. Pambakian, D. S. Wooding, N. Patel, A. B. Morland, C. Kennard, and S. K. Mannan. Scanning the visual world: a study of patients with homonymous hemianopia. *J Neurol Neurosurg Psychiatry*, 69(6):751–9, 2000.

N. R. Parry, D. J. McKeefry, and I. J. Murray. Variant and invariant color perception in the near peripheral retina. *J Opt Soc Am A Opt Image Sci Vis*, 23(7):1586–97, 2006.

A. Pascual-Leone, A. Amedi, F. Fregni, and L. B. Merabet. The plastic human brain cortex. *Annu Rev Neurosci*, 28:377–401, 2005.

M. S. Pepe. Evaluating technologies for classification and prediction in medicine. *Stat Med*, 24(24):3687–96, 2005.

J. Peter. *Introduction to expert systems.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1986. ISBN 0-201-14223-6.

M. W. Pettet and C. D. Gilbert. Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 89(17):8366–70, 1992.

B. Pleger, A. F. Foerster, W. Widdig, M. Henschel, V. Nicolas, A. Jansen, A. Frank, S. Knecht, P. Schwenkreis, and M. Tegenthoff. Functional magnetic resonance imaging mirrors recovery of visual perception after repetitive tachistoscopic stimulation in patients with partial cortical blindness. *Neurosci Lett*, 335(3):192–6, 2003.

D. A. Poggel, E. Kasten, and B. A. Sabel. Attentional cueing improves vision restoration therapy in patients with visual field defects. *Neurology*, 63(11): 2069–76, 2004.

D. A. Poggel, E. Kasten, E. M. Muller-Oehring, U. Bunzenthal, and B. A. Sabel. Improving residual vision by attentional cueing in patients with brain lesions. *Brain Res*, 1097(1):142–8, 2006.

D.A. Poggel. *Effects of visuo-spatial attention on the restitution of visual field defects in patients with cerebral lesions.* Dissertation, 2002.

D.A Poggel, I. Mueller, E. Kasten, and B. A. Sabel. Multifactorial predictors and outcome variables of vision restoration therapy in patients with visual field loss. subm.

D. Purves, G.J. Augustine, D. Fitzpatrick, W.C. Hall, A.-S. Lamantia, J.O. McNamara, and S.M. Williams. *Neuroscience (3rd ed.).* Sunderland, MA: Sinauer Associates, 2004.

J. R. Quinlan. *C4.5: programs for machine learning.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993. ISBN 1-55860-238-0.

R. Ruseckaite. Computer interactive system for ascertainment of visual perception disorders. In *The Workshop on Machine Learning in Medical Applications, Advanced Course in Artificial Intelligence, July 5-16*, pages 27–29, Chania, Greece, 1999.

B. A. Sabel. Editorial: Residual vision and plasticity after visual system damage. *Restor Neurol Neurosci*, 15(2-3):73–9, 1999.

B. A. Sabel and E. Kasten. Restoration of vision by training of residual functions. *Current Opinion in Ophthalmology*, 11:430–436, 2000.

B. A. Sabel, S. Kenkel, and E. Kasten. Vision restoration therapy (vrt) efficacy as assessed by comparative perimetric analysis and subjective questionnaires. *Restorative Neurology and Neuroscience*, 22:399–420, 2004.

P. Sajda. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng*, 8:537–65, 2006.

P. A. Sample, C. F. Bosworth, E. Z. Blumenthal, C. Girkin, and R. N. Weinreb. Visual function-specific perimetry for indirect comparison of different ganglion cell populations in glaucoma. *Invest Ophthalmol Vis Sci*, 41(7):1783–90, 2000.

E. L. Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biol Cybern*, 25(4):181–94, 1977.

G. Schwarzer, W. Vach, and M. Schumacher. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med*, 19(4):541–61, 2000.

G. Schweigart and U. T. Eysel. Activity-dependent receptive field changes in the surround of adult cat visual cortex lesions. *Eur J Neurosci*, 15(10): 1585–96, 2002.

A. Shillabeer and J. F. Roddick. On the arguments against the application of data mining to medical data analysis. *Intelligent data Analysis in BioMedicine and Pharmacology*, 2006.

A. Shmuel, M. Korman, A. Sterkin, M. Harel, S. Ullman, R. Malach, and A. Grinvald. Retinotopic axis specificity and selective clustering of feedback projections from v2 to v1 in the owl monkey. *J Neurosci*, 25(8): 2117–31, 2005.

L. C. Sincich and G. G. Blasdel. Oriented axon projections in primary visual cortex of the monkey. *J Neurosci*, 21(12):4416–26, 2001.

A. Skowron, J. Komorowski, Z. Pawlak, and L. Polkowski. Rough sets perspective on data and knowledge. pages 134–149, 2002.

G. D. Smith and S. Ebrahim. Data dredging, bias, or confounding. *Bmj*, 325 (7378):1437–8, 2002.

S. J. Sober, J. M. Stark, D. S. Yamasaki, and W. W. Lytton. Receptive field changes after strokelike cortical ablation: A role for activation dynamics. *J. Neurophysiol.*, 78:1–6, 1997.

H. Soltanian-Zadeh, F. Rafiee-Rad, and S Pourabdollah-Nejad. Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognition*, 37(10):1973–1986, 2004.

P. G. Spry, C. A. Johnson, A. M. McKendrick, and A. Turpin. Variability components of standard automated perimetry and frequency-doubling technology perimetry. *Invest Ophthalmol Vis Sci*, 42(6):1404–10, 2001.

J. W. Van Der Steeg, P. Steures, M. J. C. Eijkemans, J. D. F. Habbema, P. M. M. Bossuyt, P. G. A. Hompes, F. Van Der Veen, and B. W. J. Mol. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? *BJOG: An International Journal of Obstetrics and Gynaecology*, 113(7):825–831, 2006.

D. D. Stettler, A. Das, J. Bennett, and C. D. Gilbert. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36 (4):739–50, 2002.

D. D. Stettler, H. Yamahachi, W. Li, W. Denk, and C. D. Gilbert. Axons and synaptic boutons are highly dynamic in adult visual cortex. *Neuron*, 49(6):877–87, 2006.

O. Steward. Anatomical and biochemical plasticity of neurons: regenerative growth of axons, sprouting, pruning, and denervation supersensitivity. In Michael Selzer, Stephanie Clarke, Leonardo Cohen, Pamela Duncan, and Fred Gage, editors, *Neural Repair and Rehabilitation*. Cambridge University Press, Cambridge, 2006.

J. A. Swets and R. M. Pickett. *Evaluations of Diagnostic Systems: Methods From Signal Detection Theory*. Academic Press, New York, 1982.

P. N. Tan, M. Steinbach, and V. Kumar. *Introduction to data mining*. Pearson Addison-Wesley, 2006.

M. L. Tant, F. W. Cornelissen, A. C. Kooijman, and W. H. Brouwer. Hemianopic visual field defects elicit hemianopic scanning. *Vision Res*, 42(10): 1339–48, 2002.

T. J. Teyler, J. P. Hamm, W. C. Clapp, B. W. Johnson, M. C. Corballis, and I. J. Kirk. Long-term potentiation of human visual evoked responses. *Eur J Neurosci*, 21(7):2045–50, 2005.

B. Thuraisingham. *Data Mining: Technologies, Techniques, Tools, and Trends*. CRC Press, Inc., Boca Raton, FL, USA, 1998. ISBN 0849318157.

Tobii. Tobii 50 series product description. Technical report, Tobii Technology, Sweden, 2004.

R. B. Tootell, N. K. Hadjikhani, W. Vanduffel, A. K. Liu, J. D. Mendola, M. I. Sereno, and A. M. Dale. Functional analysis of primary visual cortex (v1) in humans. *Proc Natl Acad Sci U S A*, 95(3):811–7, 1998.

S. Trauzettel-Klosinski and J. Reinhard. The vertical field border in hemianopia and its significance for fixation and reading. *Invest Ophthalmol Vis. Sci.*, 39(11):2177–86, 1998.

F.A. Ubaudi. Development of a treatment outcome prediction model for childhood leukaemia. Technical report, Bioinformatic Research Centre, Univsersity of Technology, 2005.

D. Van Essen. Organization of visual areas in macaque and human cerebral cortex. In L.M. Chalupa and J.S. Werner, editors, *The Visual Neurosciences*, volume 1, pages 507–521. MIT Press, 2004.

D. C. Van Essen, W. T. Newsome, and J. H. Maunsell. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res*, 24(5):429–48, 1984.

R. VanRullen and T. Dong. Attention and scintillation. *Vision Res*, 43(21): 2191–6, 2003.

K. Veropoulos. *Machine learning approaches to medical decision making.* PhD thesis, University of Bristol, 2001.

J. Vesanto. Som-based data visualization methods. *Intelligent-Data-Analysis*, 3:111–26, 1999.

X. Wang, D. Nauck, M. Spott, and R. Kruse. Intelligent data analysis with fuzzy decision trees. In *Soft Computing for Information Mining (Workshop Proceedings 27th German Conference on Artificial Intelligence)*, Ulm, 2004.

T. Watanabe, Sr. Nanez, J. E., S. Koyama, I. Mukai, J. Liederman, and Y. Sasaki. Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nat Neurosci*, 5(10):1003–9, 2002.

X. Wi, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1):1–37, 2007.

I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann, 1999.

A. Y. Wong, R. M. Dodge, and L. A. Remington. Comparing threshold visual fields between the dicon tks 4000 automated perimeter and the humphrey field analyzer. *J Am Optom Assoc*, 66(11):706–11, 1995.

J. Wusteman. Explanation-based learning as justification of knowledge. *Expert Systems*, 9(3):139–148, 1992.

J. C. Wyatt and D. G. Altman. Commentary: Prognostic models: clinically useful or quickly forgotten? *British Medical Journal*, 311(7019), 1995.

X. Zhang, S. Kedar, M. J. Lynn, N. J. Newman, and V. Biousse. Natural history of homonymous hemianopia. *Neurology*, 66(6):901–5, 2006.

J. Zihl. Visual scanning behavior in patients with homonymous hemianopia. *Neuropsychologia*, 33(3):287–303, 1995.

J. Zihl and D. von Cramon. Visual field recovery from scotoma in patients with postgeniculate damage. a review of 55 cases. *Brain*, 108 (Pt 2):335–65, 1985.