



# Similarity-based models of human visual recognition

A. Unzicker \*, M. Jüttner, I. Rentschler

*Institute of Medical Psychology, University of Munich, Goethestr. 31, D-80336 München, Germany*

Received 24 December 1996; received in revised form 5 August 1997

---

## Abstract

Seven models of human visual recognition from cognitive psychology, visual psychophysics and connectionism were compared. They were used to predict psychophysical classification data obtained via supervised learning with parametrised grey-level patterns (compound Gabor signals). Four sets of learning patterns, as well as foveal and extrafoveal viewing conditions, were applied. Model performance was determined by comparing observed and predicted data with respect to root mean square deviation and to signal reconstruction via multidimensional scaling. Results show that a psychophysical theory of classification requires a similarity concept that is based both on physical signal description and on cognitive bias. The latter is less pronounced in foveal recognition, where all seven models performed almost equally well, but matters in extrafoveal recognition. Virtual prototype models (Rentschler et al. (1994), *Vision Research* 34, 669–687), which best accommodate stimulus- and observer-dependencies, are then of advantage. Concerning computational efficiency, a hyperBF model (Poggio and Girosi (1990), *Science* 247, 978) was much faster, and generalized signal detection models were much slower than the average. © 1998 Elsevier Science Ltd. All rights reserved.

*Keywords:* Classification; Recognition; Modeling; Supervised learning

---

## 1. Introduction

The process of recognition can be looked at in two ways: It consists either of assigning an object to a previously unknown class of objects or of identifying an object as a member of an already known class [1]. The first perspective is that of technical pattern recognition, where classifiers are designed as devices or processes that sort data into one of several categories or classes, e.g. [2–5]. The second perspective is prevailing in behaviour research and psychophysics. In the study of learning and memory, visual recognition is generally defined as a task, where the animal is shown a sample object and, after some interstimulus interval, both the sample and a novel object for choice (delayed matching to sample, DMS; e.g. [6]). Psychophysical studies of visual recognition usually require the subject to decide whether a test object is a mirror-image of the sample object [7], whether it is familiar or novel [8], or whether a test image is a view of a learned object or not [9,10].

As has been discussed by Edelman and Bülthoff [10], the latter psychophysical paradigm forces the subject to compare the stimulus with the representation of a specific object. Our previous work on visual recognition has been focused on precategorical classification learning, i.e. on the question of how a human subject, or an intelligent machine, acquires the ability to sort image data into previously unknown classes or categories. Related psychophysical tasks require the subject to compare the learning or test stimulus with the representation of the entire sets of occurrences of given categories. Hence our decision to adapt the technical methodologies of supervised learning and trainable classifiers to the conditions of human visual recognition [11–13].

Intuitively speaking, classes or categories consist of collections of objects that are grouped by similarity. Unfortunately, it is exceedingly difficult to cast the idea of similarity into a rigorous mathematical form. Watanabe ([1], p. 82) proved within the framework of mathematical logics that ‘any arbitrary two objects are equally similar’. To resolve this problem, he resorted to the assumption that, in order to allow a useful definition of similarity, some predicates must be more impor-

---

\* Corresponding author. Fax: +49 89 5996615; e-mail: sascha@imp.med.uni-muenchen.de

tant than others (value-oriented feature ponderation; [1], chapter 4). The situation is the same if, instead of similarity, the complementary concept of distance is considered. Further problems with the notion of similarity arise from the observation that similarity ratings by humans and animals for object pairs are asymmetric [14]. Human similarity judgements are also context-dependent [15], and right and left images of objects (mirror-image pairs) are more similar than are predicted on the basis of mathematical distance measures [14,16–18]. This led Mumford ([14], p. 2) to conclude that the idea of similarity ‘still defies mathematical modeling’.

Consistent with Watanabe’s argument, the problem of object similarity is alleviated if some sort of natural bias exists for certain features or attributes (see [19]). This might be the case if innate detector mechanisms, as are believed to exist in the field of early vision, are available. It is then possible to arrive at representational schemes on the basis of metric vector spaces, where similarity is measured as proximity, i.e. inverse distance. This idea underlies procedures of multidimensional scaling (MDS), where a monotonic relationship between distance in some vector space and the measure of pattern similarity is assumed [20,21]. Similarly, the so-called direct approach to pattern recognition [22] describes the successive steps of transformation of and feature extraction from input data in an image format. Classification is then achieved with respect to the proximity of input signals and class representatives in some feature space. The subsequent evaluation of similarity models of visual recognition will draw on this correspondence of direct object recognition strategies and MDS [23].

Another characteristic of the human judgement of similarity is its probabilistic nature. Given the same pair of stimuli, observers give different comparative judgements on successive occasions about the same pair of stimuli, A and B. Thurstone [24] concluded that the effect of sensory stimulation, the ‘discriminal process’, is a random variable, the value of which is subjected to statistical fluctuations. His scaling theory, or law of comparative judgement, uses the S.D. of the distribution of discriminational processes for measuring the psychological distance between A and B. By linking this concept of stimulus representation with the mathematical framework of the Gaussian theory of error [35], Thurstonian scaling provides a probabilistic measure of similarity.

The concept of probabilistic stimulus representation reappeared in signal detection theory (SDT), which, as Green and Swets ([25], chapter 3) state, differs from Thurstonian scaling more in emphasis than in content of the underlying theory of measurement. SDT went beyond Thurstone in that it incorporated a decision component able to disentangle factors of sensory sensi-

tivity and decision behaviour. Much like its predecessor, however, it essentially remained a one-dimensional approach, i.e. a theory restricted to stimulus variation along a 1D psychological continuum.

Over the years, a number of visual recognition or classification models have been proposed, mainly in the field of cognitive psychology [26,27]. These models differ with respect to their theoretical foundation and the mathematical formalism used. Moreover, they have been discussed mostly with respect to cognitive domains, and it is not clear to what extent they allow the consistent description of psychophysical data as well.

Given this situation, the present study evaluates seven representative models from the classification literature. They are all based on a certain notion of similarity as the basis of classification. They also assume a representational format, where a test stimulus can be represented in parameterised form, be it as a point in a multidimensional space or as a set of feature values. These models differ, however, in how representation and similarity are related to each other. Following Reed [28], we distinguish three types of classification models, namely distance models, probability models and a model of the connectionist type.

Distance models relate the interpoint distance between two stimuli to their degree of similarity as in MDS [20,21]. Pattern classes are represented either by class-specific prototypes [26] or by multiple exemplars, e.g. [29–32]. The set of distances between a given input signal and the class representatives is then used to decide about the class membership of the input signals. Both types of approaches, the prototype- and the exemplar-based ones, will be considered here.

The second type of model comprises probabilistic approaches. They assume a probabilistic stimulus representation, which is formally described by multidimensional likelihood distributions. The similarity between two given stimuli depends on the degree of overlap between their respective probability distributions. The most general of this class of approaches is general recognition theory (GRT) [33,34], which combines the notion of probabilistic stimulus representation with a deterministic decision mechanism. It is regarded by its authors as a multidimensional generalization of SDT.

Virtual prototype approaches [12,13,17] are also based on probabilistic stimulus representation, but model the internal stimulus representation as a result of measurement in the sense of applied optimal estimation, e.g. [35,36]. Feature vectors corresponding to input signals are thus combined with internal error vectors that are minimised through supervised learning. In their original form, probabilistic virtual prototypes (PVP) [12,37] assume multidimensional normal likelihood distributions. Generalised virtual prototypes (GVP), which are being considered here too, replace the Gaussian similarity functions of PVP by exponentials.

The third group of models, which just contains hyperBF networks as proposed by Poggio and Girosi [38], represents the connectionist perspective in our list of classification models. Here the problem of similarity-based classification is embedded into the framework of regularisation theory. Accordingly, the degree of class membership of a given input pattern depends on the extrapolation of the feature space representation of learned exemplar patterns of the respective class. So far this approach has been successfully applied to perceptual learning tasks such as vernier acuity learning [39] and to issues of 3D object recognition [38,40,41]. We will demonstrate here how hyperBF networks can be used to simulate human classification behaviour as well.

For evaluating computer vision models of recognition, two conditions should be met to make a comparison meaningful [42,43]. First, the comparison should be based on a sufficiently broad common data base. We meet this requirement by fitting every candidate model to 48 sets of psychophysical learning and classification data obtained in our laboratory under identical experimental conditions. These data were either collected under foveal viewing conditions or, to consider aspects of space variance of visual recognition as well, under conditions of extrafoveal viewing. Second, as far as possible, all candidate models were provided with the same number of free parameters to make their performance directly comparable. However, different models may require different numbers of parameters for an adequate specification. This holds true in particular for GRT, where the positioning of decision boundaries may be arbitrarily complex (cf. Section 2). To take account of this, we considered two implementations of GRT that differed in the way of segmenting feature space into decision regions.

In summary, for our comparison, the list of similarity-based models of visual recognition consisted of seven instances, namely one implementation of a prototype-based model (PPM), one implementation of an exemplar-based approach (GCM), two implementations of the virtual prototype approach (PVP and GYP), two implementations of general recognition theory (GRTI, GRTII), and one probabilistic network classifier (PNC), based on a hyperBF network with radial basis functions. These models were applied to predict 48 sets of psychophysical data on human visual recognition. Figures of merit were obtained in two ways, namely via root mean square (RMS) analysis of cumulative performance and via the application of MDS methodologies to recover residues between observer performance and model predictions.

## 2. Experiments

### 2.1. General methods

Visual recognition is studied here as a type of classification learning at the precategorical level. This is to say that categories have to be learned for a set of learning signals during the experiment, whereas generalisation, i.e. the classification of novel test signals, is not considered.

A prerequisite of this approach is the use of unfamiliar stimulus patterns. We employ grey-level images (compound Gabor signals) that can be synthesised simply from their co-ordinate values in a 2D feature space of evenness and oddness dimensions. The choice of such features does not imply that they are necessary or even sufficient for human pattern recognition performance in general. What is important here is that evenness and oddness features allow us to uniquely define learning and test signals in the context of the present experiments, thus giving full control over within- and between-variances of signals clusters in feature space. They are also biologically plausible since pairs of even- and odd-symmetric filters can serve as basis functions of spatial vision when being centred at the same location of the visual field [44,45]. More realistic representations for visual recognition, may then be obtained by non-linearly combining the outputs of even- and odd-symmetric filters [46–48]. Yet, in principle, it is impossible to prove uniqueness of a given feature representation for solving a pattern recognition problem anyway [1].

A sufficiently broad psychophysical data base is obtained by having 16 observers learn the classification of four signal configurations that vary in the relationships of within- and between-variances of pattern classes. Because of this, two types of viewing conditions will be considered, namely foveal and extrafoveal stimulus presentation. The latter variation is of interest since it provides information as to the extent to which human visual pattern recognition is position-invariant, e.g. [49].

### 2.2. Stimuli

The compound Gabor signals (see Fig. 1) were generated in a  $128 \times 128$ , 8-bit pixel format with a linear grey-level-to-luminance function. Intensity profiles were defined by

$$G(x, y) = L_0 + \exp\left(-\frac{1}{\alpha^2}(x^2 + y^2)\right) \times (a \cos(2\pi f_0 x) + b \cos(2\pi 3f_0 x + \phi)), \quad (1)$$

where  $L_0$  determines the mean luminance,  $\alpha$  is the space constant of the Gaussian aperture,  $a$  is the amplitude of

the fundamental,  $b$  is that of the third harmonic, and  $\phi$  is the phase angle of the latter. Thus, the 2D images consisted of a fundamental cosine waveform and its third harmonic modulated by an isotropic Gaussian aperture which decayed to  $1/e$  in 32 pixels.

Signal variation was restricted to  $b$  and  $\phi$ . This allowed the use of a 2D feature space with the Cartesian coordinates

$$\xi = b \cos \phi \tag{2}$$

(evenness) and

$$\eta = b \sin \phi \tag{3}$$

(oddness). Four signal sets (Fig. 2), each consisting of 15 samples, were used in the experiments. The 15 signals of each set were grouped into three classes, forming three clusters in feature space, each containing five samples. The four sets of signals differed in the grouping of their samples and, therefore, in the degrees of within- and between-variance in feature space. The compound Gabor signals were displayed on a raster monitor (Lucius & Baer GBM 2310, P4 phosphor) linked to a digital image processing system (Matrox PIP 1024, mounted on a PC AT 486). Space average luminance (DC) was kept constant at 70 cd/m<sup>2</sup>.

The stimulus patterns subtended 1.7° at a viewing distance of 101 cm when seen foveally. The fundamental spatial frequency was 2.4 cpd. In 3° off-axis viewing conditions, the stimulus size was rescaled to 2.7° ac-

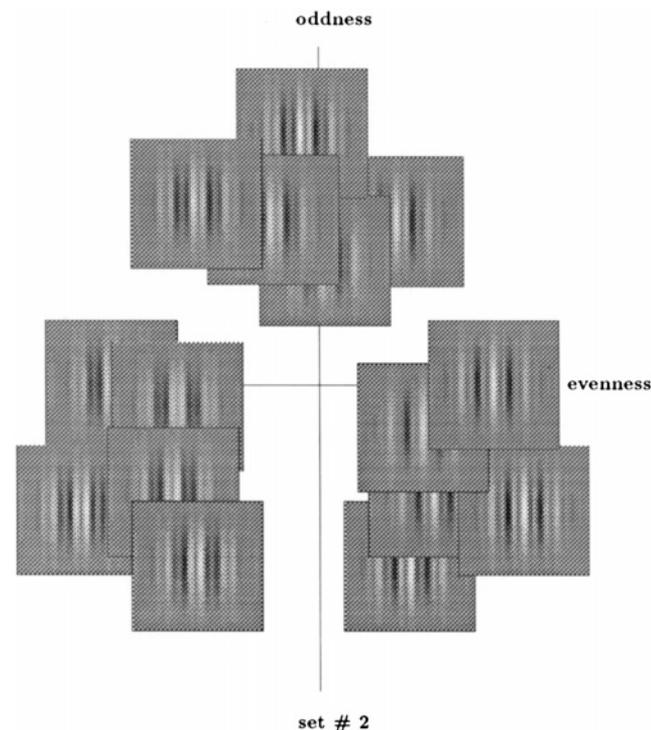


Fig. 1. Compound Gabor signals with  $f+3f$  spatial frequency components of stimulus configuration No. 2 (see also Fig. 3), plotted at the respective feature space coordinates.

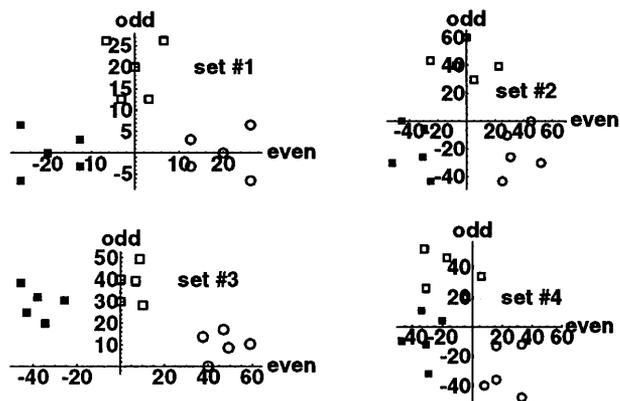


Fig. 2. Compound Gabor signals with  $f+3f$  spatial frequency components, in four different stimulus configurations. Circles: class No. 1, squares: class No. 2, filled squares: class No. 3.

ording to cortical magnification [50], i.e. by reducing the viewing distance by a factor of 1.6 (to 63 cm). Eccentricity was measured as the distance between the fixation point and the next edge of the stimulus pattern, a definition corresponding to a fixation 1.65° away from the left or right edge of the target.

The categorisation of the learning signals into three classes was still a feasible task, though in the extrafoveal learning condition, subjects frequently failed to reach the learning criterion of 100% correct in 40 sessions. Hence, signals like those shown in Fig. 1 can be regarded as ‘highly confusable’ rather than ‘fairly discriminable’.

### 2.3. Procedure of supervised learning and classification

The learning procedure consisted of a variable number of learning units. One learning unit contained three subsequent presentations, in random order, of the learning set with 200 ms exposure duration for each pattern. Following each presentation, a number specifying the class to which the pattern belonged was displayed for 1 s. The interval between the learning signal and the class number was 500 ms.

Each learning unit ended with a test of how well the subject was able to classify the 15 patterns. Only one exposure per sample was used here. The deliberately chosen learning criterion was reached if the observer had achieved an error-free classification. This procedure continued until the subject had met the learning criterion, or refused to continue due to an apparent inability to succeed with the task.

There were three viewing conditions: central (learning and testing in direct view), left (learning and testing when fixating a fixation target presented on the horizontal meridian 3° to the left of the centre of the stimulus patterns), and right (same but on the right side). Hence, in all conditions, learning and testing occurred at the same retinal location.

The procedure of supervised learning in foveal and extrafoveal view was applied to the four different sets of patterns depicted in Fig. 2. Three of these sets (i.e. sets No. 2–4) have been evaluated already in the previous study by Jüttner and Rentschler [13], who employed the same training scheme and the same viewing conditions as outlined above. This allowed us to re-analyse these data within the scope of the current work. Pattern set 1 in the present study was new, requiring additional experiments.

Altogether 16 observers participated in the experiments. They were divided into four groups of four subjects each. Each group was assigned to one set of patterns, which all subjects of a group had to learn under all three viewing conditions. Nine subjects started with the central classification task, before switching to the eccentric conditions left and right; the other seven performed the experiments in reverse order, starting with eccentric conditions. Viewing was always binocular.

### 2.4. Subjects

Data were obtained from 16 paid observers. All of them had never participated in any psychophysical experiment before. The age of the observers ranged between 20 and 30 years, 9 of them were female, 7 were male; all had normal or corrected to normal vision.

## 3. Models of classification

Here we investigate the extent to which similarity-based models of visual supervised learning, developed independently in the areas of psychometrics, visual psychophysics, and artificial neural nets, can be interpreted within the framework of a uniform mathematical structure. This is achieved by defining a likelihood function  $f_J(i)$  with the property that the probability of stimulus  $i$  being assigned to category  $J$  is given by

$$P(J | i) = \frac{f_J(i)}{\sum_k f_k(i)}, \tag{4}$$

where the sum is taken over the likelihood functions of all classes  $k$ . Eq. (4) is normalised in the sense that it keeps the predicted classification probabilities in the range 0–1. We will show in the following how the respective models can be rephrased in a way that conforms with Eq. (4).

### 3.1. Generalised context model (GCM)

The generalised context model [32,51], which generalises the context model of Medin and Schaffer [31], makes explicit use of the concept of similarity as con-

ceived of by Luce [52]. Here, the likelihood function of a signal  $i$  being a member of class  $J$  assumes the form of a weighted sum of the similarities  $\eta_{ij}$  between that signal and all members  $j$  of class  $J$ :  $f_J(i)$  corresponds to the numerator of equation (1) of [53]:

$$f_J(i) = \beta_J \sum_{j \in J} \eta_{ij}, \tag{5}$$

where the sum is taken over all signals of class  $J$ . The bias parameter  $\beta_J$  with  $0 \leq \beta_J \leq 1$  and  $\sum_j \beta_j = 1$  describes the preference of a subject for category  $J$ . The similarity  $\eta_{ij}$  of two signals  $i$  and  $j$  is given by

$$\eta_{ij} = e^{-d_{ij}^p} \tag{6}$$

with the so-called similarity exponent  $p$ . The (psychometric) distance  $d_{ij}$  is expressed as a ‘weighted’  $L_r$ -Norm,

$$d_{ij} = c \left( \sum_{k=1}^n w_k (x_{ik} - x_{jk})^r \right)^{1/r}. \tag{7}$$

A plot of the three likelihood functions  $f_J(i)$  is shown in Fig. 3. The signal  $i$  is characterised by the coordinates  $x_{ik}$  of its feature vector, i.e.  $\mathbf{x}_i = \{x_{i1}, \dots, x_{in}\}$ . Along each feature dimension  $k$  (see below) the distance is weighted by a factor  $w_k$  with  $0 \leq w_k \leq 1$  and  $\sum_k w_k = 1$ . The overall distance scale is further adjusted by the parameter  $c$  while the exponent  $r$  determines the type of  $L_r$ -metric:  $r = 1$  corresponds to the city-block metric,  $r = 2$  to the Euclidean, and the maximum metric is obtained for  $r \rightarrow \infty$ .

It is worthwhile to note that the similarity functions of GCM are symmetric by definition,  $\eta_{ij} = \eta_{ji}$ . As a consequence, the classification errors of assigning a signal of class  $I$  to class  $J$ , and vice versa, are being treated as basically equal, with the bias parameters  $\beta_J$  allowing for some asymmetry.

In the present implementation of GCM, the feature vectors were two-dimensional (evenness and oddness coordinates), four free parameters were admitted:  $c$ , the ratio  $w_1/w_2$ , and two degrees of freedom for the three biases  $\beta_J$  with the constraint  $\sum_j \beta_j = 1$ . The parameters  $r$

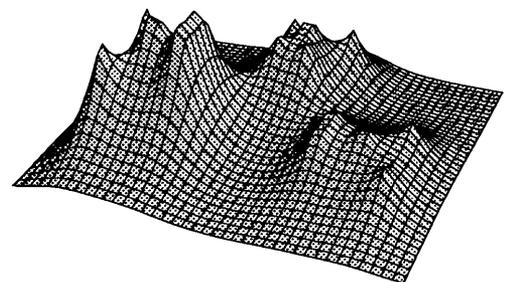


Fig. 3. Sum of the three likelihood functions  $f_1(i), f_2(i), f_3(i)$ , corresponding to the denominator of Eq. (4) with exemplary parameters of the GCM. Note the dimension weighting (steeper gradients in evenness-direction); on the other hand, the height of the three hills corresponds to the respective bias towards a class.

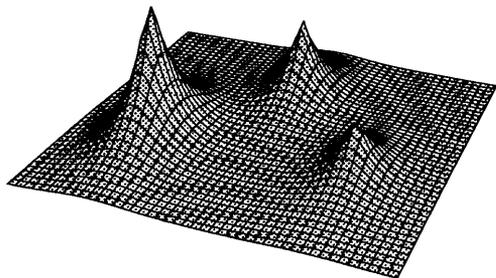


Fig. 4. Sum of likelihood functions for the PPM. Note the same parameters bias and weight like in the GCM.

(order of  $L_r$ -Norm) and  $p$  were fixed at the values of 2 (Euclidean metric) and 1, respectively. The latter assumption is justified by the results of Kahana and Bennett [53], who varied the exponents as a free parameter and obtained nearly optimal results for the above values. The free parameters were varied to minimise the difference between observed and predicted data.

### 3.2. Probabilistic prototype model (PPM)

The probabilistic prototype model by Nosofsky [32] and its precursor by Reed [26], to whose notion of probabilistic classification models we refer here, use the same exponential decay functions for the probability densities as the GCM. However, unlike the GCM, the exponential decay functions are centred at the class prototypes, i.e. the mean vectors of the category (or class) exemplars (see Fig. 4). The notion of a class prototype requires one ‘mental image’ or ‘class model’ for a set of signals forming a category in the categorisation process. The prototype is not necessarily a member of the set of signals of a given class but represents the latter best (see also the theory of vector quantisation [54] and Section 3.5). Formally, PPM is obtained by replacing the likelihood function Eq. (5) by

$$f_j(i) = \beta_j e^{-d_{ij}^p} \quad (8)$$

with the distance function  $d_{ij}$  being identical to that of Eq. (7) for  $p = 1$ .

By employing three densities for evaluating the class conditional classification probabilities in Eq. (4), the experimental data were fitted by varying the four free parameters of the model.

### 3.3. Probabilistic virtual prototypes (PVP)

The general classification form Eq. (4) is equivalent to Bayes theorem with equal a priori probabilities if its left-hand side is interpreted as the probability of class  $J$  being given once occurrence  $I$  has been observed (a posteriori probability). The right-hand side of Eq. (4) is then the likelihood of  $J$  divided by the sum of the

likelihood functions of all classes (the normalisation factor). To obtain a model of human classification behaviour, Rentschler and co-workers [12,17,37] assumed that the class conditional densities are internally measured with statistically independent additional errors. In the sense of applied optimal estimation [35,36], the internal measurement errors are varied to allow a least-squares fit between observed and predicted classification probabilities.

By choosing multivariate normal distributions as class conditional densities, the likelihood function (a plot of the three functions  $f_j(i)$  is given in Fig. 5) writes

$$F_j(i) = \frac{1}{2\pi \sqrt{|\mathbf{C}_j|}} \exp(-d) \quad (9)$$

with

$$d = \frac{1}{2}(\mathbf{x} - \mu_j)^T \mathbf{C}_j^{-1} (\mathbf{x} - \mu_j) \quad (10)$$

and

$$\mathbf{x}' = \mathbf{x} + \mathbf{e} \quad (11)$$

to define the process of internal measurement ( $\mathbf{x}$  is signal vector in physical feature space,  $\mathbf{C}_j$  is the covariance matrix,  $\mathbf{x}'$  is the signal vector in internal representation, and  $\mathbf{e}$  is the error vector of internal measurement). The bias model of classification behaviour yields

$$\mu'_j = \mu_j + \mathbf{e}_j \quad (12)$$

The latter mean pattern or prototype vectors are called virtual prototypes (see [12] for details), since the observer behaves as if the signal classes were distributed according to Eq. (12). Note that Eq. (12) assumes an internal shift of the otherwise undistorted class conditional densities, whereas GCM and PPM vary (in the sense of variance models) the width of the similarity function. Thus, PVP model the degrees of freedom of the internal measurement by varying the ‘perceptual dimensionality’ [13] of the internal representations underlying pattern recognition. The closer two virtual prototypes get together, the higher the respective misclassification.

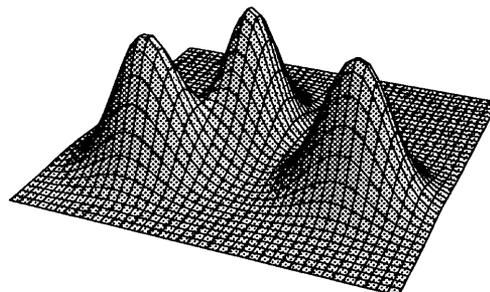


Fig. 5. Sum of likelihood functions for PVP.

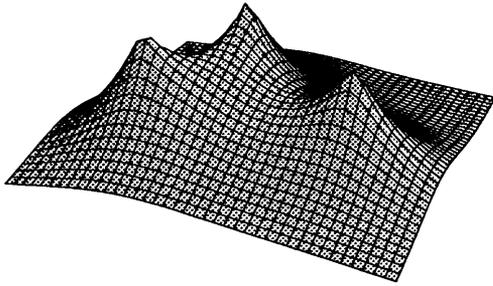


Fig. 6. Sum of likelihood functions for the GVP

PVP has been implemented again for the 2D feature coordinates of pattern evenness and oddness. Since, in case of the bias model, its solutions are invariant against translation of the virtual prototype configuration [13], the number of free parameters of the computational problem is reduced from six to four. PVP is based on the assumption that learning, governed by neural plasticity, can be characterised by assuming adaptive encoding at the level of internal representations. One advantage of this perspective is that it allows the description of the dynamical properties of the learning process by successively evaluating the concurrent prototypes in running temporal windows [12,13,55].

### 3.4. Generalised virtual prototypes (GVP)

This model is obtained from the PVP by alleviating the restriction to multivariate normal density functions. Formally, it can be expressed by substituting the likelihood equation (Eq. (9)) by

$$f_j(i) = \frac{1}{2\pi \sqrt{|C_j|}} \exp(-d^p), \quad (13)$$

with a general decay exponent  $p$ . Fig. 6 shows the sum of the functions  $f_j(i)$ . Since  $d$  is quadratic in  $x$  (Eq. (10)), we obtain for the particular choice  $p = 1/2$  an exponential decay that corresponds to an exponential decay for the similarity function like in the GCM and PPM considered above. The motivation for implementing the virtual prototype model with such an exponential likelihood function is twofold: First, Kahana and Bennett [53] obtained a remarkably good data fit with a similarly modified version of prototype model. Second, Shepard [56] has argued from a more theoretical perspective that the probability for stimulus generalisation decays exponentially with stimulus distance, and this regularity may be mathematically derived from universal principles of natural kinds and probabilistic geometry.

### 3.5. General recognition theory (GRT)

General recognition theory (GRT), as proposed by Ashby and co-workers [27,33,34], assumes that cate-

gories are associated with regions in feature space. According to this concept, a stimulus is not memorised directly but the stimulus plus superimposed noise is stored at the representation level. Classification errors occur when the (Gaussian) noise is large enough to remove the signal from the correct region in feature space. Since this assumption resembles the signal and noise paradigm of signal detection theory, GRT can be seen as a multidimensional extension of it. A formulation of GRT that conforms with Eq. (4) can be arrived at by defining the likelihood function  $f_j(i)$  with which a signal  $i$  is classified to category  $J$  as

$$f_j(i) = \int_{G_j} \exp(-m^2) dx, \quad (14)$$

where  $G_j$  is the feature space region associated with category  $J$ . The exponent  $m^2$  is given by

$$m^2 = (\mathbf{x} - \mathbf{x}_i)^T \mathbf{C}_j^{-1} (\mathbf{x} - \mathbf{x}_i), \quad (15)$$

with signal coordinates  $\mathbf{x}_i$  and the noise  $(\mathbf{x} - \mathbf{x}_i)$ .

A visualisation of the integral Eq. (14) is given in Fig. 7. Note that this figure is not directly comparable with Figs. 3–6, since it shows only the integrand of Eq. (14). The integral over this discontinuous function equals the value  $f_j(i)$  in a given point  $i$ . Whereas in all other models only the values of functions at certain points are involved, GRT requires a numerical integration over the entire feature space. For this reason it is much slower in computation than all other models. This implies that GRT is under no condition formally equivalent to the GCM, as claimed by Nosofsky and Smith [57].

Moreover, GRT follows a different concept of signal processing. In the previous models the representation of an input signal was deterministic and the decision process supposed to be probabilistic, whereas GRT as-

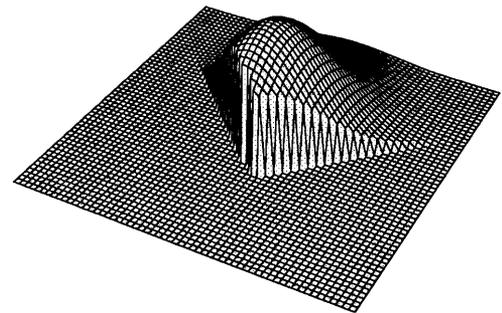


Fig. 7. Probabilistic distribution of the representation in GRT. Categories are associated with regions  $G_j$  in feature space. Gaussian noise is added to an incoming signal  $i$ , so the vector signal + noise may be outside the correct domain  $G_j$ . The part of the Gaussian cut by the (here linear) decision bounds determine the probability with which a signal  $i$  is assigned to a given category  $J$ . In the present example, the integral over the remaining part of the Gaussian yields the predicted probability for a signal  $i$  placed at the centre of the Gaussian to be correctly classified.

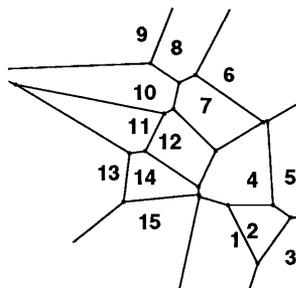


Fig. 8. Voronoi tessellation for the 15 signals of stimulus configuration No. 4. Every signal represents a 'basin of attraction' from which it is likely to be chosen in the classification task. All points inside a given basin are closer to the centre point than to the centre points of the neighbouring basins. For the present version of GRT, a tessellation of three regions only is employed.

sumes the opposite: classification errors are due to the occurrence of noise already at the level of the internal representation; whereas decision rules are deterministic. In brief, GRT simulates 'external noise'; whereas the other models focus on 'internal noise'.

The implementation we are proposing here is different from previous models of GRT [34,53,58,59]. Due to the preset low number of the free parameters, we had to restrict ourselves on the one hand to linear decision bounds. A realisation of a general linear classifier without any constraints for decision bounds would require here, for three categories, eight free parameters. On the other hand, an implementation with intersecting straight lines as considered by Kahana and Bennett [53], or quadratic bounds [27], is possible for an even number of categories only and puts rather arbitrary constraints on the decision bounds. For this reason, we sought a generally applicable method that keeps the number of free parameters limited.

The problem of finding appropriate segmentation occurs frequently in speech and image coding, and in tasks of data compression. A rather successful approach to such problems is offered by vector quantisation techniques, which are used both for storage and transmission of data [54,60]. The idea of vector quantisation is to assign a given set or distribution of input signals to a number of classes, and then to represent any signal just by the class to which it belongs. A common way to obtain linear decision bounds in feature space is the Voronoi tessellation algorithm ([61], p. 225). We propose to use this approach in combination with GRT. This means that regions are assigned to points in feature space using a simple nearest-neighbour criterion. To compute the Voronoi tessellations, we used the Mathematica package 'Computational geometry'.

Fig. 8 gives an example of a two-dimensional Voronoi tessellation for the single signals of set No. 3. In our present work, we investigated two versions of GRT, which differed by the choice of free parameters. In both models, a Voronoi tessellation, given by the

class prototypes (see Fig. 9) yielded the linear decision bounds for the three regions corresponding to categories.

In the first version (GRT1), the separating three lines could only be rotated about their centre (see Fig. 9). The remaining three parameters were occupied by width, eccentricity and orientation of the Gaussian shown in Fig. 7. In the second implementation (GRT2), the lines could be rotated about their centre (one parameter) and be shifted both in evenness- or oddness-directions (two parameters), whereas the width only of the Gaussian was varied.

### 3.6. Probabilistic network classifier (PNC)

In recent years hybrid networks containing layers for both supervised and unsupervised learning became popular [62–64]. In this context, radial basis functions [38] defined in feature space—as employed before by Poggio and Edelman [65] and by Bühlhoff and Edelman [66]—are analogous to similarity-based non-network models. Such a network has been used to simulate human classification behaviour.

The PNC is based on a hyperBF algorithm. HyperBF networks were shown to be equivalent to regularisation algorithms [38]. The prediction of human classification behaviour is a similar task. From observed classification data with a given set of learning signals, one has to predict classification behaviour for the whole feature space. This is equivalent to an interpolation problem of scattered data. In order to choose an appropriate function, which approximates a given set of data, one has to compromise between smoothness and closeness to the data. This compromise is controlled by a regularisation parameter (see [38]). In the GCM and PPM models considered here, a role similar to the regularisation parameter is played by the overall distance parameter (see Eq. (7)). To keep the number of free parameters limited, we allowed three hidden neurons (RBFs) only, corresponding to the number of categories. With these prerequisites, we could develop the likelihood function needed for Eq. (4).

A schematic description of the PNC is shown in Fig. 10. The first (hidden) layer of the network—the feature space layer in our case—used the standard competitive

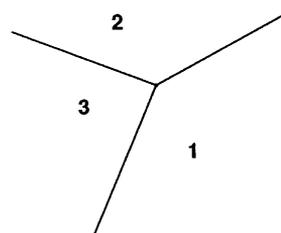


Fig. 9. Voronoi tessellation used in the GRT models, here for the prototypes of configuration No. 4.

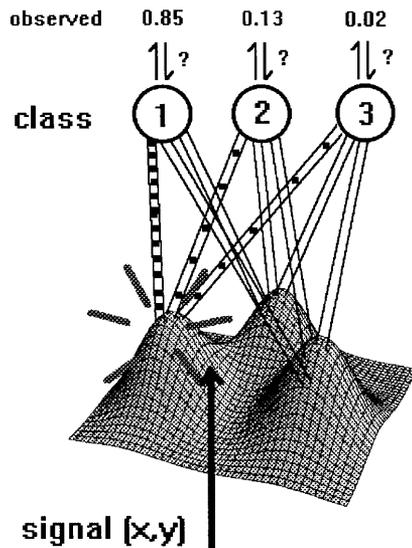


Fig. 10. PNC, a signal  $(x, y)$  (evenness, oddness) activates the hidden neuron layer consisting of RBFs; the RBF-neurons ‘fire’ across the synaptic weights to the continuous-valued output layer that yields the estimated probability of classification.

learning rule (see [61], chapter 9.1) to determine class prototypes by an unsupervised learning procedure. According to this rule, only one of the hidden layer units (the winner) can fire at a time. For each input stimulus, the network finds the winner  $J^*$  among the class representatives and then updates the weights  $w_{J^*i}$  for the winning unit only to make the  $w_{J^*}$  vector closer to the current input signal  $s$ . The updating rule is ([61], equation (9.1))

$$\Delta w_{J^*i} = \eta(s_i - w_{J^*i}), \quad (16)$$

which moves  $w_{J^*}$  directly towards  $s$ . After training, the three  $w_J$  values obtained in this way are taken as class prototypes and the radial basis functions (RBFs) are placed there. These RBFs are now regarded as ‘receptive fields’ in feature space, their height is indicating the sensitivity to a signal  $(x, y)$  in feature space.

In the following supervised learning procedure an input signal  $i$  activates the hidden neuron layer consisting of RBFs; the RBF-neurons ‘fire’ across the synaptic weights (now from hidden to output layer)  $\omega_{ij}$  to the continuous-valued output layer. The value of the output neurons yields an a priori probability for the signal  $i$  to be recognised as a member of category  $J$ , or, a first-estimate of the desired likelihood function  $f_J(i)$ .

The output layer is now trained with the Hebbian learning rule (see [61], equation (9.38)), according to which the synapses between contemporarily firing neurons are strengthened. This changes the weights  $\omega_{ij}$  as follows:

$$\Delta \omega_{ij} = \epsilon(h_i - p_i)R_j, \quad (17)$$

where  $h_i$  and  $p_i$  are the observed and predicted classification probabilities and  $R_j$  is the activity of the hidden

neuron, the value of the RBF. Of course,  $\omega_{ij}$  are generally not symmetric. Tuned by the learning rate  $\epsilon$ , the network converges quickly.

The  $p_i$  corresponds to the  $P(J|i)$  in Eq. (4), so we obtain for the likelihood function

$$f_J(i) = \sum_{j \in J} R_j w_{ij}. \quad (18)$$

Instead of running an optimisation algorithm for the (nine) parameters  $\omega_{ij}$ , we can now simply train the rapidly converging network.

To avoid confusion, two points should be emphasised here. First, the model predictions were not used to assign learning patterns to specific classes but to simulate the classification matrix of a given subject. Thus, the neural net was trained with the confusion matrices and not with the input patterns. Similar training procedures have also been employed by studies on object recognition [10,41,65]. Second, the predictions of the PNC could be further improved by more densely paving the feature space or by choosing other (not necessarily radially symmetric) basis functions.

#### 4. Model evaluation

The purpose of similarity-based models is to reveal hidden structures of psychological data. By applying such techniques, complex confusion matrices like the one shown in Table 1 can be summarised and efficiently displayed, thus yielding a deeper understanding of the underlying recognition processes.

The raw data consisted for each subject and learning condition of a  $15 \times 3$  matrix with the relative classification frequency of each of the 15 signals being assigned to three categories (see Table 1 for an example).

Table 1  
Confusion matrix of subject MF

Class	Signal	Class 1	Class 2	Class 3
1	1	0.750	0.225	0.025
1	2	0.800	0.175	0.025
1	3	0.825	0.175	0.000
1	4	0.700	0.300	0.000
1	5	0.725	0.200	0.075
2	6	0.200	0.650	0.150
2	7	0.075	0.625	0.300
2	8	0.200	0.625	0.175
2	9	0.125	0.600	0.275
2	10	0.225	0.625	0.150
3	11	0.025	0.375	0.600
3	12	0.075	0.275	0.650
3	13	0.025	0.325	0.650
3	14	0.025	0.225	0.750
3	15	0.000	0.275	0.725

Relative classification frequencies as numerical data. For example, signal No. 7, belonging to class No. 2, is correctly classified with relative frequency 0.625.

The seven candidate models were fitted to the individual classification data obtained for each pattern set and viewing condition. Model performance was assessed in two ways. First, the free parameters of each approach were computed by minimising the RMS error, defined by

$$\text{RMS} = \sqrt{(n_c n_s - 1)^{-1} \sum_{ij} (P_{ij} - P'_{ij})^2}, \quad (19)$$

where  $n_c$  is the number of classes and  $n_s$  is the number of signals per class.  $P_{ij}$  denotes the model-predicted classification probability for assigning stimulus  $i$  into class  $j$ , and  $P'_{ij}$  refers to the corresponding observed classification frequency. To solve this minimisation problem, we applied a modified version of the downhill simplex algorithm [67], which was reiterated with a gradually adapting starting point to improve the optimisation.

For a visualisation of the model performance, we employed a multidimensional scaling (MDS) technique proposed by Cutzu and Edelman [23]. To apply this method, for each signal pair  $(i, j)$  (rows of the classification matrix, see Table 1) the Euclidean distance  $d_{ij}$  between the classification vectors was computed, i.e.

$$d_{ij} = \sqrt{\sum_k (p_{ik} - p_{jk})^2}. \quad (20)$$

The resulting distance matrix was then submitted to MDS yielding a geometrical 2D solution. To quantify the congruence between two given MDS solutions, the configurations of reconstructed feature states were fit to each other by means of a Procrustes transformation (a combination of scaling, rotation, reflection and translation, see [68]). The residual distance remaining after the transformation served as a measure of pairwise congruence between configurations.

## 5. Results

### 5.1. Root mean square values (RMS)

To arrive at a statistically reliable comparison of model predictions, the RMS data were pooled across observers and a three-way analysis of variance (ANOVA) was performed with model, pattern set and viewing condition as factors. The ANOVA yielded significant main effects of pattern set ( $F(3,280) = 27.12$ ,  $P < 0.001$ ) and viewing condition ( $F(1,280) = 11.86$ ,  $P < 0.001$ ), whereas the effect of model was only close to significance ( $F(6,280) = 1.86$ ,  $P = 0.09$ ). From the interaction terms only that of model  $\times$  pattern set was significant ( $F(18,280) = 1.93$ ,  $P = 0.02$ ).

The effects of pattern set and viewing condition are illustrated in Fig. 11 which shows the respective mean RMS values, averaged across models. First, we observe

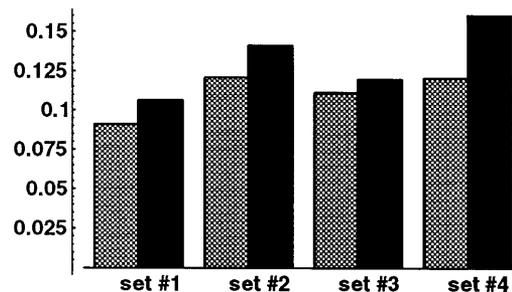


Fig. 11. RMS values averaged across models for foveal (grey) and extrafoveal (black) viewing conditions, four sets of learning patterns.

that RMS values are consistently higher for extrafoveal learning than for foveal learning. In addition, there is a consistent ranking of RMS values in both viewing conditions: RMS increases on the order of pattern sets 1, 3, 2, 4, for both foveal and extrafoveal learning. A further one-way ANOVA revealed that this ranking is significant only with respect to the most extreme differences, i.e. set 1 versus set 2 and 4 for foveal learning and set 1 and set 3 versus set 2 and set 4 for extrafoveal learning. Nevertheless, these results suggest the existence of an underlying effect of the structure of the individual pattern sets. Indeed, the pattern configurations (set No. 2 and No. 4, see Fig. 2) vary considerably in within-class variance, i.e. in the spread of the individual samples of a given class. In fact, the mean variance across classes in the four pattern sets correlates with the corresponding RMS values. In the case of foveal learning, the correlation coefficient is 0.87, whereas for extrafoveal learning it reaches 0.94. We conclude that the degree of ‘fuzziness’ of the signal classes in feature space has a pronounced effect on the psychometrics of classification.

As to the individual performance of the models in our comparison, we recall that the initial ANOVA yielded no global effect of model type. However, the fact that the main effect of the factor model and the interaction term model  $\times$  pattern set were both either significant or close to significance led us to reanalyse the performance of the individual models in the four pattern sets. Fig. 12 shows the RMS values for the seven models, grouped according to pattern set and viewing condition. To evaluate the significance of the observed RMS deviations of the various candidate models, separate two-way ANOVAs with model and pattern set as factors were computed for both viewing conditions, and linear deviation contrasts were constructed. Significant contrast parameters would then indicate, depending on their sign, a relative decrease or increase in performance of a particular model.

For foveal viewing, none of the models showed a significant global effect across the four pattern sets. The only significant local effect was the increase in performance of the PVP model in set 3 ( $t(77) = -3.60$ ,

$P < 0.001$ ). For extrafoveal viewing, there was a global advantage of the GVP model ( $t(203) = -2.11$ ,  $P < 0.04$ ). Furthermore, the performance decrease of the PVP model for pattern set 2 ( $t(203) = 2.95$ ,  $P < 0.04$ ) and its increase for set 3 ( $t(203) = -3.00$ ,  $P < 0.03$ ) were significant.

## 5.2. MDS analysis

As outlined above, the analysis of the foveal classification data in terms of RMS values did not establish significant performance differences between the candidate models. Such a result could indicate that all tested approaches are essentially equivalent in describing foveal classification behaviour. Alternatively, it could just signify a lack of sensitivity of the RMS analysis concerning the detection of subtle differences. As a further attempt to differentiate between the candidate models, we subjected the foveal classification data to further analysis in terms of MDS. Here, the basic idea is to transform, for a given set of learning stimuli, both the pattern of model-predicted classification probabilities and the experimental classification frequencies into 2D geometrical configurations. These MDS-derived configurations can be compared either with each other or with the signal configuration as defined in physical feature space.

As an example consider Fig. 13. This plot shows the configuration of pattern set 4 (small symbols, cf. Fig. 2, bottom right) in superposition with the MDS solution derived from the experimental classification data in foveal view (large symbols). To obtain the latter, the

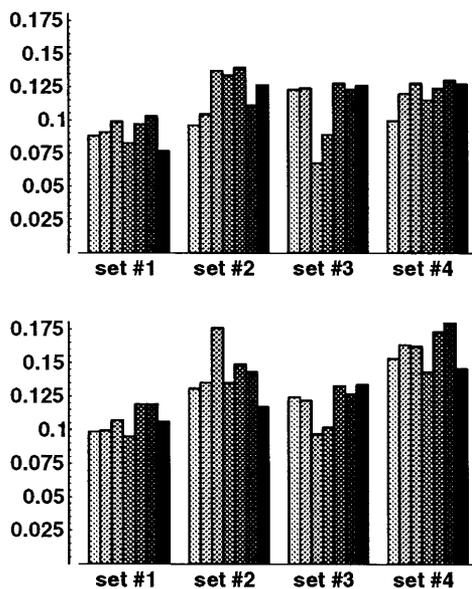


Fig. 12. RMS values grouped according to pattern set and model (each cluster, set 1–4, from left to right: GCM, PPM, PVP, GVP, GRT1, GRT2, PNC). Top: foveal viewing condition, bottom: extrafoveal viewing condition.

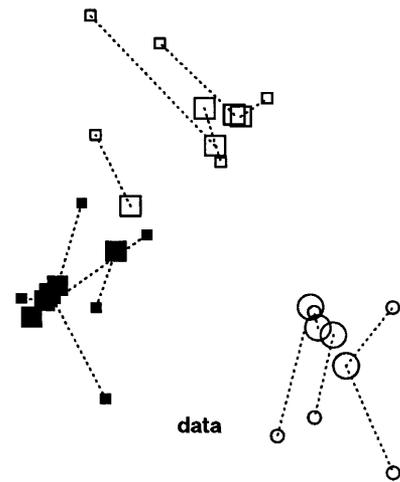


Fig. 13. Physical and predicted feature states of learning signals for observer behaviour (data). Small symbols: objective feature coordinates of learning patterns. Large symbols: feature states reconstructed via MDS from observer behaviour. Data pooled over four observers for pattern configuration 4.

data of each subject were first MDS-analysed. The resulting configurations were then optimally Procrustes-transformed in order to fit the physical signal configuration. The dashed lines in Fig. 13 indicate the mean displacement vectors across subjects for each signal. As it is evident from the plot, samples belonging to the same class tend to be clustered much more densely in the MDS-derived configurations than in the underlying physical feature space. This observation also holds for the data of the other pattern sets not shown here. It indicates that due to the training procedure observers tend to develop compact class concepts where within-class variations appear to be suppressed relative to differences between different classes.

From the perspective of modelling, the question at issue is as to what extent the configuration of pattern similarity evident in the experimental data is reflected in the predictions of the candidate models. Fig. 14 shows, again for pattern set 1, the MDS solutions derived from the model-predicted classification probabilities (large symbols). Note that the MDS configuration of the empirical data (small symbols in Fig. 14) served as target configuration to which the model predictions were Procrustes-transformed. The general impression from Fig. 14 is that all seven candidate models replicate the similarity pattern derived from the experimental data quite well. To further quantify the extent of congruence between model-predicted and data-based MDS solutions, we computed the distance between corresponding configurations. Fig. 15 shows the congruence values for the seven models, grouped according to pattern set. A comparison of Fig. 15 and Fig. 12 (top) suggests that both RMS values and MDS congruence measures generally follow a similar overall pattern although the variation across pattern sets seems to be

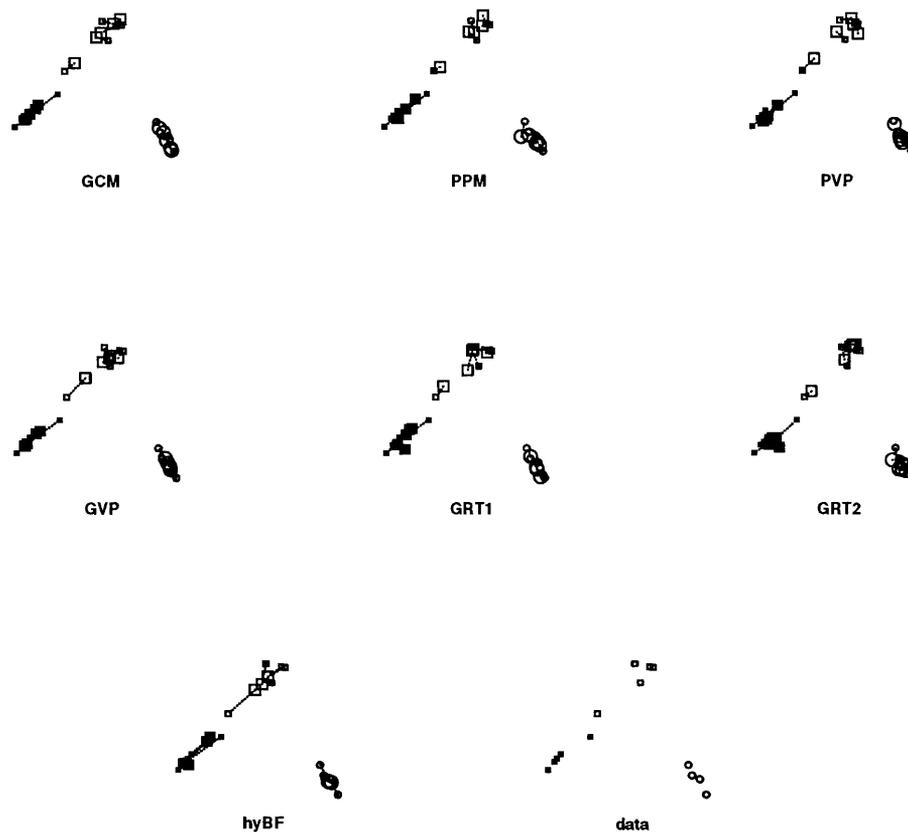


Fig. 14. Observed and predicted feature states of learning signals. Small symbols: feature states reconstructed via MDS; large symbols: feature coordinates reconstructed via MDS from model predictions (GCM, PPM, PBC, GBC, GRT1, GRT2, hyBF). Bottom right panel: observed data reconstructed via MDS only (data). Data pooled over four observers for pattern configuration 4.

larger in the MDS case. This was confirmed by a two-way ANOVA of the MDS distances with pattern set and model as parameters. The analysis only yielded a significant global effect of pattern set ( $F(3,111) = 38.6$ ,  $P < 0.001$ ), whereas that of model ( $F(6,111) = 1.03$ ,  $P = 0.41$ ) and pattern set  $\times$  model ( $F(18,111) = 0.81$ ,  $P = 0.688$ ) were non-significant. Furthermore, linear deviation contrasts indicated as the only significant local effect a performance increase of the PVP model for pattern set 3 ( $t(77) = -2.43$ ,

$P < 0.03$ ), thus confirming the effect previously obtained in the RMS analysis of the data.

In summary, the analysis of the data in terms of both RMS and MDS suggests an equivalence of the candidate approaches concerning the modelling of foveal classification behaviour. In contrast, there is a tendency of divergence in model performance with respect to extrafoveally acquired class concepts. We will elaborate on these findings and how they may relate to the formal structure of the individual models in the discussion.

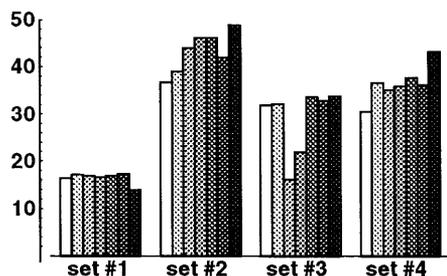


Fig. 15. Distances between model-predicted and data-based MDS solutions, grouped according to pattern set and model (each cluster, set 1–4, from left to right: GCM, PPM, PVP, GVP, GRT1, GRT2, PNC). Data pooled over 4 subjects for each set.

### 5.3. Resource consumption

Differences in computation time obviously influence the degree of complexity to which theoretical concepts underlying the respective models can be realised. To our knowledge, for example, higher-dimensional implementation of GRT with quadratic decision bounds does not yet exist. To complete the description of our comparison, the computational time needed by the models is given in Fig. 16. This display reveals that, among the seven models under consideration, the two versions of GRT are extremely slow and hyperBF is very fast.

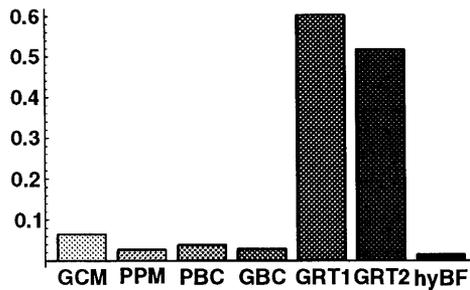


Fig. 16. Computational time required by minimisation algorithms of the seven models considered here (arbitrary units).

## 6. Discussion

The goal of this study was to compare seven similarity-based models of visual classification or categorisation. Four of these models, the exemplar-based GCM, the prototype-based PPM, and the two implementations of the multidimensional signal detection model GRT, have been developed in the context of cognitive psychology. The two types of modified Bayes classifiers, the PVP and the GVP, have been designed to analyse psychophysical learning and classification data, whereas PNC is a neural net structure of the hyperBF type predicting classification performance. We have endeavoured to formulate these models in such a manner to bring them in line with the predictions of the PVP and GVP. The reason for this is the fact that the latter models most directly reflect the structure of the (parametric) probabilistic Bayes classifier, which is standard in technical pattern recognition. The main result of our study is that the current models show more or less equivalent performance in predicting human visual recognition as observed in a number of psychophysical tasks of supervised learning and classification.

### 6.1. Similarity of similarity-based models of visual recognition

The finding of a quasi-equivalence of classification models has been obtained by comparing human performance and model predictions in two different ways. One was the analysis of RMS between observed and predicted data. The other one was based on the fact that all sets of learning and test patterns were arranged in feature space as conspicuous configurations. It was thus possible to recover these configurations from the psychophysical classification data and from the classification matrices produced by the various models by using multidimensional scaling (MDS; see [23]). This allowed the visualisation and the numerical evaluation of deviations between reconstructed feature states from observed and from predicted data. The results from both types of analyses, RMS and MDS, were largely consistent.

Two conditions met by our study are supportive of our conclusion as to the similarity of the seven classification models considered here. Our comparison of model predictions was based on a sufficiently broad basis of psychophysical data obtained in our own laboratory under identical experimental conditions. Furthermore, all candidate models were provided, as far as possible, with the same numbers of free parameters to make their direct comparison meaningful.

The apparent similarity of categorisation models found here is reminiscent of the one reported by Cohen and Massaro [69], who compared a fuzzy logical model of perception (FLMP), a two-layer connectionist model (CMP), a model derived from SDT, a linear integration model (LIM), and a model based on MDS as employed to the analysis of a two-categories visual classification task and a four-categories task of phonological classification. These authors concluded that predictions of classification probabilities in some categorisation tasks are not sufficient to distinguish among models. We, however, believe that the surprising result of the quasi-equivalence of the models at issue reflects the existence of structural similarities between them.

In Section 3 we have shown that it is possible to formulate the seven models under consideration in a way that conforms with the view of pattern recognition as statistical decision making ([1], chapter 8). Eq. (4) is obtained from the Bayes formula in the case of equal prior probabilities, which is equivalent to using likelihoods instead of Bayesian probabilities. As has been pointed out, individual models differ in the way the likelihood functions are implemented but the structural variability of classification models thus generated is constrained by the general form of Eq. (4).

The common mathematical structure, imposed on the classification models by their respective authors, would yield a convergence of model performance both with very low and very high numbers  $n$  of free parameters. The differences of model structures would be blurred completely for the cases of  $n = 1$  and  $n = 45$  (number of data points or feature states of learning sets), with the latter case consisting of a trivial data fit. To achieve a reasonable compromise between these extreme cases, we imposed the additional constraint onto the models of fixing  $n$  at the relatively low value of 4. This restriction allowed a reasonable comparison of model predictions, with the individual characteristics of the model structures still being preserved. The latter fact is evident from the likelihood functions  $f_j(i)$  shown in Figs. 3–6.

The constraints imposed on the classification models considered here evidently caused some sort of first-order equivalence regarding the data description. This became clear especially in foveal classification tasks where a close correspondence was found between the feature space representations of learning sets and the perceived interclass similarity as manifest in the MDS solution (e.g. Fig. 14).

For extrafoveal classification tasks, however, this correspondence is less convincing and the emerging divergence between physical and internal representations allows second-order structural differences between the individual models to become manifest. As will be argued below, it is the different extent to which the various classification models allow to disentangle the respective effects of physical signal representation and cognitive bias that underlies the existence of such second-order effects of model performance.

## 6.2. Space variance of visual recognition

In Biology, it is often assumed that visual recognition is largely shift- or space-invariant, i.e. independent of object location, [70]. Concerning human vision, a more elaborate version of this view is the hypothesis that an image ‘can be made equally visible everywhere in the visual field by scaling its size’ (cortical magnification theory; [71], p. 56). The limitations of this concept are obvious from the existence of a large number of counterexamples (see [72,73]). A possible explanation of this discrepancy is the existence of structural differences between internal pattern representations in foveal and extrafoveal view as revealed by visual classification learning [13].

All models under consideration in the present study performed consistently better for foveal than for extrafoveal viewing conditions. Since they can all be regarded as distance-related in the sense that the degree of between-class similarity is a monotonous function of between-class distance, this suggests that for foveal viewing the configuration of the signals in the physical feature space is vertical to the internal representations of the signals underlying the mental classification process.

For extrafoveal viewing the internal class representations appear to be distorted in the sense that they no longer adequately capture the physical signal configuration. This distortion is no specific deficit in the processing of the oddness feature dimension as it has been proposed earlier [53]. Indeed, none of the models employing explicit dimensional weight factors for the evenness and oddness dimension (GCM and PPM) performed any better in fitting the data. The superiority of the virtual prototype approaches (PVP and GVP) for this condition (see Figs. 11 and 12) allows some conclusions as to the nature of internal representations in extrafoveal vision. This is particularly true for pattern set 3. Fig. 17, for example, shows a virtual prototype solution obtained by minimising the RMS for the PVP implementation. The configuration of the virtual prototypes is degenerated to a nearly collinear configuration. According to the PVP and GVP concept, virtual prototypes are regarded as biased representations of their physical counterparts. Due to the class specificity, all

within-class distances remain unchanged, and so do the within-class variances. As a direct consequence, the relative between-class distances, i.e. the interclass distances normalised by the within-class variances, change as well. In other words, the virtual prototype concept introduces a notion of perceptual distance or similarity, which is the product of an observer-dependent (subjective) component (the class-specific biases) and a stimulus-dependent (objective) component (the within-class variances, defined by the distribution of the physical signals). The interpretation of the result shown in Fig. 17 therefore, is two-fold: First, the perceptual class distances are much smaller than the physical ones. Second, while the original configuration (No. 3 in Fig. 2) extends in two feature dimensions, the corresponding internal representation is degenerated to one dimension.

Obviously, the phenomenon of reduced perceptual dimensionality, which we have discussed elsewhere [12,13] cannot be accounted for by the alternative approaches PPM, GCM and GRT in their present implementations. In case of PPM and GCM this is due to the fact that their free parameters, i.e. the dimensional weights and the overall distance parameter, do affect between-class distance and within-class variance simultaneously, thus leaving all perceptual (relative) between-class distances unaltered. This might explain the relatively poor performance of these models in case of pattern set 3, where the divergency between perceptual and physical class distances becomes most pronounced. In case of the GRT approach, the explanation is more difficult. This model could account for an adjustment of the perceptual distance by varying the within-class variance with constant between-class distance. However, GRT strongly relies on an appropriate choice of the decision boundaries. While these boundaries may be arbitrarily complex in principle, they were constrained in order to restrict the number of the degrees of freedom. We sought to take account of this difficulty by testing GRT with two different solutions of the segmentation problem within the decision space but results indicate that this model is less suitable for a parsimonious description of classification performance in the domain of grey-level patterns.

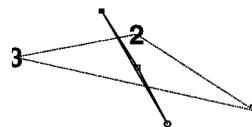


Fig. 17. Example (subject TM) for a virtual ‘prototype’ triangle, a simplified representation of the likelihood functions in Fig. 5. The dotted lines represent the ‘physical’ triangle of the signal set 3 in Fig. 2. The solid triangle connects the centers of the Gaussians in Fig. 5. Circles: class No. 1, squares: class No. 2, filled squares: class No. 3.

### 6.3. Conclusions

We compared seven models of human visual recognition from cognitive psychology, visual psychophysics, and connectionism. They were employed to predict the assignment of parametrised grey-level patterns to previously unknown classes of objects. Related tasks depend on the simultaneous ponderation of multiple feature dimensions, thus constituting pattern recognition proper [1]. To establish a sufficiently broad data base, we measured classification performance for four different learning sets of patterns and in two viewing conditions (foveal versus extrafoveal performance), thus considering also aspects of space variance of recognition.

For relating input signals to class concepts (examples, prototypes, or conditional densities), their similarity is measured either in terms of distance in a metric vector space (distance models), of overlap of class conditional densities (probabilistic models), or within the framework of regularisation theory (hyperBF model). For comparison, the models were implemented, as far as possible, with no more than four free parameters. Moreover, our approach was restricted to supervised classification learning, thus ignoring performance with novel test signals, i.e. generalisation. The latter has been dealt with elsewhere [18,74].

We showed that

- a psychophysical theory of classification requires a similarity concept that is based both on physical signal description and on cognitive bias;
- the cognitive bias is less pronounced in foveal recognition, where all seven models, according to RMS and MDS criteria, performed almost equally well;
- the cognitive bias is more important in extrafoveal recognition. Virtual prototype models (PVP, GYP), which best accommodate stimulus- and observer-dependencies, are of advantage.

We contend that structural ‘similarities of similarity models of visual recognition’ [69] underlie the equivalence of model predictions reported here.

These similarities are due to two types of constraints imposed onto these models. First, they all conform to the view of pattern recognition as statistical decision making (see [1], chapter 8). Second, the present comparison is based on the restriction to four free parameters per model.

Another aspect of classification has been excluded from the present study too, namely that of learning dynamics. This issue has been ignored largely in the cognitive and psychophysical classification literature, and where attempts have been made to fit models to successive blocks of learning data [12,13,55] they fail to capture aspects of dynamic memory, e.g. [75]. By contrast, such characteristics are implicit in connectionist classification models, which were represented in this

study by the PNC model. Whether or not non-connectionist approaches, such as the PVP or the GYP, can be modified to accommodate dynamic memory functions is an issue of ongoing research.

### Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, grant 337/10-2 to I.R. This study was part of a Ph.D. project of Alexander Umzicker at the Faculty for Medicine, University of Munich. We are grateful to the helpful suggestions of an anonymous reviewer.

### References

- [1] Watanabe S. Pattern Recognition. New York: Wiley, 1985.
- [2] Duda RO, Hart PE. Pattern Classification and Scene Analysis. New York: Wiley, 1973.
- [3] Ahmed N, Rao K. Orthogonal Transforms for Digital Signal Processing. New York: Springer, 1975.
- [4] Pratt WK. Digital Image Processing. New York: Wiley, 1978.
- [5] Sklansky J, Wassel GN. Pattern Classifiers and Trainable Machines. New York: Springer, 1981.
- [6] Dudai Y. The Neurobiology of Memory. Oxford: Oxford University Press, 1989.
- [7] Koriat A, Norman J. Mental rotation and visual familiarity. *Percept Psychophys* 1985;37:429–39.
- [8] Rock I, DiVita J. A case of viewer-centered object perception. *Cogn Psychol* 1987;19:280–3.
- [9] Edelman S, Poggio T. Bringing the grandmother back into the picture: A memory-based view of object recognition. *Int J Pattern Recog* 1992;6:37.
- [10] Edelman S, Bülthoff H. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vis Res* 1992;32(12):2385–400.
- [11] Caelli T, Rentschler I, Scheidler W. Visual pattern recognition in humans. I. Evidence for adaptive filtering. *Biol Cybern* 1987;57:233–40.
- [12] Rentschler I, Jüttner M, Caelli T. Probabilistic analysis of human supervised learning and classification. *Vis Res* 1994;34:669–87.
- [13] Jüttner M, Rentschler I. Reduced perceptual dimensionality in extrafoveal vision. *Vis Res* 1996;36(7):1007–22.
- [14] Mumford D. Mathematical Theories of Shape. *Proc SPIE 1570 Geometric Methods in Computer Vision*, 1991:2–10.
- [15] Tversky A. Features of similarity. *Psychol Rev* 1977;84:327–52.
- [16] Rentschler I, Treutwein B. Loss of spatial phase relationships in extrafoveal vision. *Nature* 1985;313:308–10.
- [17] Rentschler I, Caelli T. Detection and recognition. In: Haken H, Stadler M, editors. *Synergetics of Cognition*. New York: Springer, 1990:233–48.
- [18] Rentschler I, Barth E, Caelli T, Zetsche C, Jüttner M. Generalization of form in visual pattern classification. *Spatial Vis* 1996;10:59–85.
- [19] Edelman S. Representation, similarity, and the chorus of prototypes. *Minds Mach* 1995;5:15–46.
- [20] Shepard RN. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 1962;27:125–40.

- [21] Shepard RN. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* 1962;27:219–46.
- [22] Caelli T, Bischof W. Machine learning paradigms for pattern recognition and image understanding. *Spatial Vis* 1996;10:87–103.
- [23] Cutzu F, Edelman S. Faithful representation of similarities among three-dimensional shapes in human vision. *Proc Natl Acad Sci* 1996;93:12046–50.
- [24] Thurstone LL. Psychological analysis. *Am J Psychol* 1927;38:368–89.
- [25] Green DM, Swets JA. Signal Detection Theory and Psychophysics. New York: Krieger, 1974.
- [26] Reed SK. Pattern recognition and categorization. *Cogn Psychol* 1972;3:346–82.
- [27] Ashby FG. Multidimensional Models of Perception and Cognition. Hillsdale: Erlbaum, 1992.
- [28] Reed SK. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973.
- [29] Brooks L. Nonanalytic concept formation and memory for instances. In: Rosch E, Loyd B, editors. *Cognition and Categorization*. Hillsdale: Erlbaum, 1978.
- [30] Estes W. Array models for category learning. *Cogn Psychol* 1986;18:500–49.
- [31] Medin D, Schaffer M. Context theory of classification learning. *Psychol Rev* 1978;85:207–38.
- [32] Nosofsky RM. Attention, similarity, and the identification categorization relationship. *J Exp Psychol General* 1986;115:39–57.
- [33] Ashby FG, Perrin NA. Toward a unified theory of similarity recognition. *Psychol Rev* 1988;95:124–50.
- [34] Ashby FG, Gott J. Decision rules in the perception and categorization of multidimensional stimuli. *J Exp Psychol Learning Mem Cogn* 1988;14:33–53.
- [35] Gauss C. *Teoria motus*. English translation (1963): Theory of the motion of the heavenly bodies about the sun in conic sections. New York: Dover, 1809.
- [36] Gelb A. *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.
- [37] Rentschler I, Jüttner M, Caelli T. Ideal observers for supervised learning and classification. In: Steyer R, Wender K, Widaman K, editors. *Psychometric Methodology*. Stuttgart: Gustav: Fischer, 1993:2554–8.
- [38] Poggio T, Girosi F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* 1990;247(4945):978.
- [39] Poggio T, Fahle M, Edelman S. Fast perceptual learning in visual hyperacuity. *Science* 1992;256:1018–21.
- [40] Maruyama M, Teraoka T, Abe S. Recognition of 3D flexible objects by GRBF. *Biol Cybern* 1994;70:377–85.
- [41] Logothetis N, Pauls J, Bülthoff H, Poggio T. View-dependent object recognition by monkeys. *Curr Biol* 1994;4(5):401–13.
- [42] Jain R, Binford T. Dialogue-Ignorance, myopia, and naivety in computer vision systems. *CVGIP: Image Understand* 1991;53:112–7.
- [43] Snyder MA. Reply. A commentary on the paper by Jain and Binford. *CVGIP: Image Understand* 1991;53:118–9.
- [44] Robson J. Receptive fields: neural representation of the spatial and intensity attributes of the visual image. In: Carterette E, Friedman MP, editors. *Handbook of Perception*. New York: Academic Press, 1975:81–116.
- [45] Pollen DA, Ronner SF. Phase relations between adjacent simple cells in the visual cortex. *Science* 1981;212:1409–11.
- [46] Burr DC, Morrone MC, Spinelli D. Evidence for edge and bar detectors in human vision. *Vis Res* 1989;29:419–31.
- [47] Zeevi YY, Porat M. Image representation by localized phase. *Proc SPIE 1199: Vis Commun Image Process*, 1989:1512–7.
- [48] Wegmann B, Zetsche C. Visual system based polar quantization of local amplitude and local phase of orientation filter outputs. In: *Human Vision and Electronic Imaging: Models, Methods, and Applications*. Proc SPIE 1249, 1990, pp. 607–613.
- [49] Nazir TA, O'Regan JK. Some results on translation invariance in the human visual system. *Spatial Vis* 1990;5:1–19.
- [50] Rovamo J, Virsu V. An estimation and application of the human cortical magnification factor. *Exp Brain Res* 1979;37:495–510.
- [51] Nosofsky RM. Choice, similarity, and the context theory of classification. *J Exp Psychol Learning Mem Cogn* 1984;10:104–14.
- [52] Luce RD. Detection and recognition. In: Luce RD, Bush RR, Galanter E, editors. *Handbook of Mathematical Psychology*. New York: Wiley, 1963:103–89.
- [53] Kahana M, Bennett PJ. Classification and perceived similarity of compound gratings that differ in relative spatial phase. *Percept Psychophys* 1994;55(6):642–56.
- [54] Gray R. Vector quantization. *IEEE Mag* 1984;4:4–28.
- [55] Unzicker A, Jüttner M, Rentschler I. Modeling the dynamics of visual classification learning. *Math Soc Sci*, 1998 (to appear).
- [56] Shepard RN. Toward a universal law of generalization for psychological science. *Science* 1987;237:1317–23.
- [57] Nosofsky RM, Smith J. Similarity, identification, and categorization: comment on Ashby and Lee, 1990. *J Exp Psychol General* 1991;121(2):237–45.
- [58] Ashby FG, Maddox WT. Integrating information from separable psychological dimensions. *J Exp Psychol Hum Percept Perform* 1990;16:598–612.
- [59] Ashby FG, Maddox WT. Complex decision rules in categorization: Contrasting novice and experienced performance. *J Exp Psychol Hum Percept Perform* 1992;18:50–71.
- [60] Nasrabadi NM, King RA. Image coding using vector quantization: a review. *IEEE Trans Commun* 1988;36:957–71.
- [61] Hertz J, Krogh A, Palmer G. *Introduction to the Theory of Neural Computing*. Reading, MA: Addison-Wesley, 1990.
- [62] Hecht-Nielsen R. Counterpropagation networks. *Appl Opt* 1987;26:4979–84.
- [63] Huang WY, Lippman RP. Neural net and traditional classifiers. In: Anderson DZ, editor. *Neural Information Processing Systems*. New York: American Institute of Physics, 1988:387–96.
- [64] Specht DF. Probabilistic neural networks. *Neural Networks* 1990;3:109–18.
- [65] Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. *Nature* 1990;343(6255):263.
- [66] Bülthoff H, Edelman S. Evaluating object recognition theories by computer graphics psychophysics. In: Glaser D, Poggio T, editors. *Exploring Brain Functions: Models in Neuroscience*. New York: Wiley, 1992.
- [67] Nelder J, Mead R. A simplex method for function minimization. *Comput J* 1965;7:308–13.
- [68] Borg I, Lingoes J. *Multidimensional Similarity Structure Analysis*. New York: Springer, 1987.
- [69] Cohen M, Massaro D. On the similarity of categorization models. In: Ashby FG, editor. *Multidimensional Models of Perception and Cognition*. New York: Erlbaum, 1992:395–447.
- [70] Dill M, Wolf R, Heisenberg M. Visual pattern recognition in *Drosophila* involves retinotopic matching. *Nature* 1993;365:751–3.
- [71] Virsu V, Rovamo J, Nasanen R. Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature* 1978;271:54–6.

- [72] Strasburger H, Rentschler I, Harvey L. Cortical magnification theory fails to predict visual recognition. *Eur J Neurosci* 1994;6:1583–8.
- [73] Azzopardi P, Cowey A. The overrepresentation of the fovea and adjacent retina in the striate cortex and the dorsal lateral geniculate nucleus of the macaque monkey. *Neuroscience* 1996;72:627–39.
- [74] Jüttner M, Caelli T, Reuschles I. Recognition-by-parts: a computational approach to human learning and generalization of shapes. *Biol Cyber* 1996;74:521–35.
- [75] Cover TM, Wagner TJ. Topics in statistical pattern recognition. In: Fu KS, editor. *Digital Pattern Recognition*. New York: Springer, 1976:15–46.