

Generalization of form in visual pattern classification*

INGO RENTSCHLER,^{1,†} ERHARDT BARTH,¹ TERRY CAELLI,²
CHRISTOPH ZETZSCHE¹ and MARTIN JÜTTNER¹

¹*Institute of Medical Psychology, University of Munich, Goethestrasse 31, 80336 München, Germany*

²*Department of Computer Science, Curtin University of Technology, Perth, WA, Australia*

Received 12 August 1995; accepted 31 October 1995

Abstract—Human observers were trained to criterion in classifying compound Gabor signals with symmetry relationships, and were then tested with each of 18 blob-only versions of the learning set. Generalization to dark-only and light-only blob versions of the learning signals, as well as to dark-*and*-light blob versions was found to be excellent, thus implying virtually perfect generalization of the ability to classify mirror-image signals. The hypothesis that the learning signals are internally represented in terms of a ‘blob code’ with explicit labelling of contrast polarities was tested by predicting observed generalization behaviour in terms of various types of signal representations (pixelwise, Laplacian pyramid, curvature pyramid, ON/OFF, local maxima of Laplacian and curvature operators) and a minimum-distance rule. Most representations could explain generalization for dark-only and light-only blob patterns but not for the high-thresholded versions thereof. This led to the proposal of a structure-oriented blob-code. Whether such a code could be used in conjunction with simple classifiers or should be transformed into a propositional scheme of representation operated upon by a rule-based classification process remains an open question.

1. INTRODUCTION

In a series of experiments we have investigated to what extent human observers classify signals composed of compound Gabor waveforms according to evident physical signal representations (Fourier descriptors) and conventional pattern-classification strategies (Caelli *et al.*, 1987; Rentschler *et al.*, 1994; Jüttner and Rentschler, 1996; Rentschler *et al.*, in preparation). Our results have demonstrated, that *internal* representations used by human observers for such Gaussian modulated grating patterns do not always fit with obvious *physical* feature states. This is particularly true when symmetry relationships between patterns (Rentschler *et al.*, in preparation) or extrafoveal viewing conditions (Jüttner and Rentschler, 1995) are involved.

*A preliminary version of this paper was presented at the 16th Conference on Visual Perception (ECVP), held at the University of Edinburgh, UK, 25 August–29 August, 1993 (1993, *Perception* **22** (Suppl.): 46).

†E-mail: ingo@imp.med.uni-muenchen.de

To account for such deviations between human and formal classification performance, we adopted the view that behavioural data reflect the operation of a (formal) classifier on some internal signal representation rather than on the physical input signals *per se* (Rentschler *et al.*, 1994). Clearly, this concept hinges on the assumptions made for formally relating the physical and internal signal representations. So far, we have used linear filter operations for feature extraction (evenness and oddness pattern components), and we have assumed that the perceptual process of internal feature measurement gives rise to the occurrence of additive stochastic error signals. The corresponding random variables describe the additional degrees of freedom of bias and variance within the framework of a parametric Bayesian classifier. Estimates of internal feature states can then be obtained by using standard techniques of applied optimal estimation (Gauss, 1809; Gelb, 1974). Resulting class means or covariance matrices, the so-called *virtual prototypes*, provide insight into the structure and dimensionality of internal signal representations.

Virtual prototype analysis treats physical feature states as a given and models human classification behaviour via the variation of pattern proximity in feature space. The proximity can be measured as Euclidean distance in the case of the (least squares) Minimum Distance Classifier (e.g. Duda and Hart, 1973; see also Caelli *et al.*, 1987), or be defined in terms of the scatter of samples around class means (Mahalanobis distance; e.g. Duda and Hart, 1973, Chapter 2; see also Jüttner and Rentschler, 1996). This approach is consistent with Shepard's proposal of a monotonic relationship between distance in some multidimensional vector space and stimulus *similarity*, i.e. the principle of multidimensional scaling (Shepard, 1962a, b). The probabilistic version of this corresponds to the underlying principle of Thurstonian scaling (Thurstone, 1927).

However, the nature of our previous approach is linear in the sense that the feature representation of the signals is obtained from linear filtering, and linear superposition of the signal statistics and of the stochastic components of the process of internal feature measurement is being employed. Thus we are to expect, that there exist problems in human visual pattern recognition which are too complex to be solved in this way.

One such problem is that of mirror-image forms, which are perceptually more similar than they should be according to their distance in feature space. In foveal vision, mirror-image compound gratings (Rentschler and Treutwein, 1985) and compound Gabor signals (Caelli *et al.*, 1986; Rentschler *et al.*, 1988) are more difficult to distinguish than nonsymmetric signals, whereas they are virtually indistinguishable both in extrafoveal vision (Rentschler and Treutwein, 1985; Bennett and Banks, 1987) and in texture perception (Caelli *et al.*, 1986; Rentschler *et al.*, 1988). Similar results have been obtained with mirror-image edge-only patterns (Saarinen, 1987, 1988). For classifying mirror-image compound Gabor signals, observers rely almost entirely on evenness (bar type) features, whereas oddness features (left or right edge) can only be appreciated by learning more sophisticated neural processing strategies (Rentschler *et al.*, in preparation). The functional characteristics of the latter have not been revealed by applying virtual prototype analysis. We have argued, however,

that the discrimination and classification of mirror-symmetric signals cannot be understood in terms of linear systems analysis (Caelli *et al.*, 1986; Zetzsche *et al.*, in preparation), and this view is shared by Bennett and Banks (1991). We have further shown, that nonlinear transducer functions applied to multi-channel systems are likewise insufficient, and nonlinear interactions of channels, i.e. in scale space, are to be invoked (Zetzsche *et al.*, in preparation).

Thus it appears that there exist visual tasks which human observers solve by relying on internal representations more complex than those obtained from linear filtering. We examined the latter by training observers in classifying the samples of a learning set of signals with mutual symmetry relationships, and then testing their ability to generalize to 'blob' versions thereof. Our study was motivated by the observation, that in texture perception observers seem to encode compound Gabor signals in terms of the occurrence of specific blobs or parts of the patterns defined by size, shape, contrast parameters such as light and dark polarity, and (relative) position (see Caelli *et al.*, 1986). Related coding schemes have been considered by several authors (Koenderink, 1984; Watt and Morgan, 1985). If blob structures were indeed 'building blocks' of the visual quality of form, generalization to blob-only transformed patterns should be relatively easy. This possibility was under investigation here.

2. METHODS

2.1. Stimuli

Compound Gabor signals were generated on a 128×128 pixel format with a linear gray-level to luminance function. Intensity profiles were defined by

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

where L_0 determines the mean luminance, α the space constant of the Gaussian aperture, a the amplitude of the fundamental, b that of the third harmonic, and ϕ the phase angle of the latter. Thus the two-dimensional 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 ϕ . This allowed the use of a two-dimensional feature space with the Cartesian coordinates $\xi = b \cos \phi$ ('evenness') and $\eta = b \sin \phi$ ('oddness'). Due to the fixation of the fundamental waveform in cosine phase (symmetry with respect to the Gaussian window), a symmetric position of signals about the evenness axis implied mirror-symmetric modulation waveforms. Symmetric position about the oddness axis implied contrast inversion of the $3 f_0$ component.

The Gabor signals were displayed on a raster monitor linked to a digital image processing system (Barco TVM 3/3.2, P4 phosphor; Videograph, LSI 11/73). Space average luminance (DC) was kept constant at 60 cd m^{-2} . The stimulus patterns subtended a visual angle of 1.6° at a viewing distance of 165 cm, and the fundamental spatial frequency was 2.5 cyc deg^{-1} .

2.2. Set of learning signals

The set of learning signals consisted of 12 compound Gabor signals (Fig. 1), which were clustered in the evenness-oddness feature space in groups of three signals about the corners of a square (Fig. 2). This square was centred on the origin, with the sides of the square being parallel to the feature axes. The coordinates of the individual signals were chosen in such a way that for each signal there existed a mirror-image signal (inflection of the oddness coordinate) and a signal with inverted contrast of the

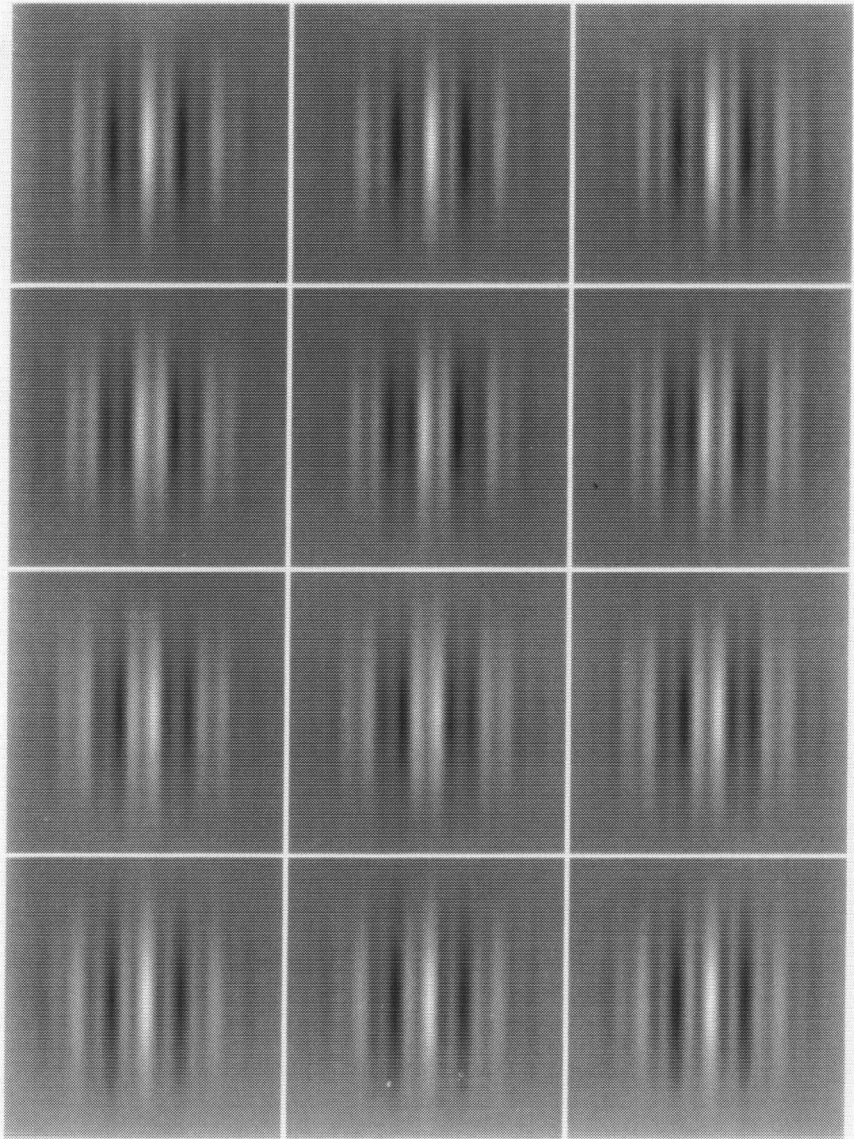


Figure 1. The learning set of 12 signals (originals).

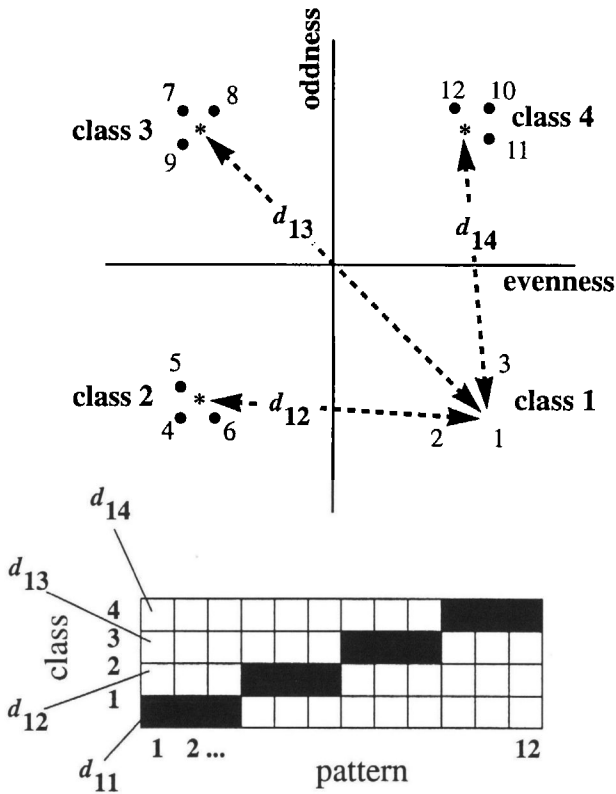


Figure 2. The learning set as defined in the evenness- and oddness-feature space.

third harmonic (inflection of the evenness coordinate). These latter two signals had the same Euclidean distance from the former. This implied that 8 of the 12 learning signals had the same lower image power, whereas four signals had the same slightly higher power.

2.3. Eighteen sets of test signals

Novel test signals were generated by means of gray-level modification of the learning set. The nonlinear image distortions consisted of dividing the range of 256 gray-levels into three regions, which were arranged symmetrically around DC. The width of the middle range was ± 10 , 20 or 40 gray-levels. Within one or two of the resulting regions, all gray-levels were set to 127 (i.e. DC), whereas they remained untouched elsewhere (Fig. 3). This resulted in $3 \times 6 = 18$ different sets of learning signals. Thus we obtained 18 *blob versions* for each of the 12 original patterns. The appearance of these blob-only patterns can be seen in Fig. 4, where the 18 blob versions corresponding to the first signal of the learning set are shown. The arrangement of the 6 lines corresponds to the six types of blob transformations shown in Fig. 3.

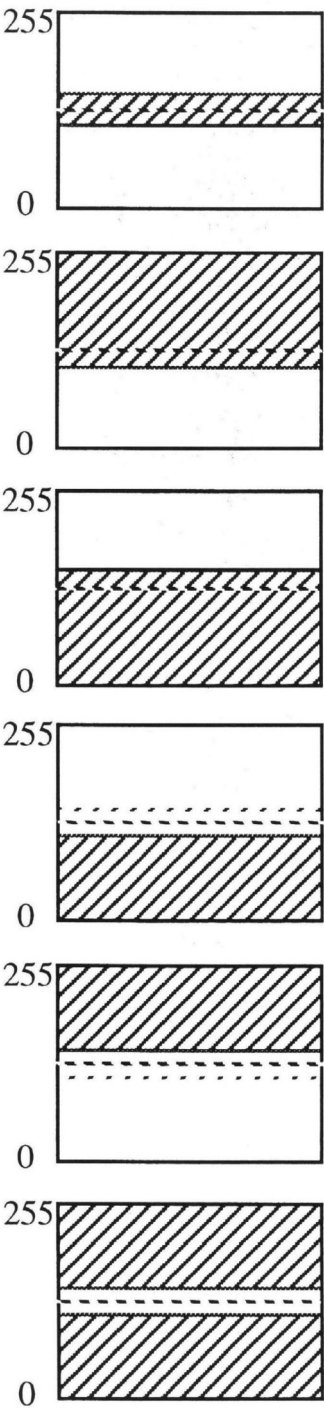


Figure 3. Definition of the 6 distortion types. The width of the gray-level interval around DC (i.e. level 127) was ± 10 , 20, and 40 leading to the 18 test signals shown in Fig. 4 (for details see text).

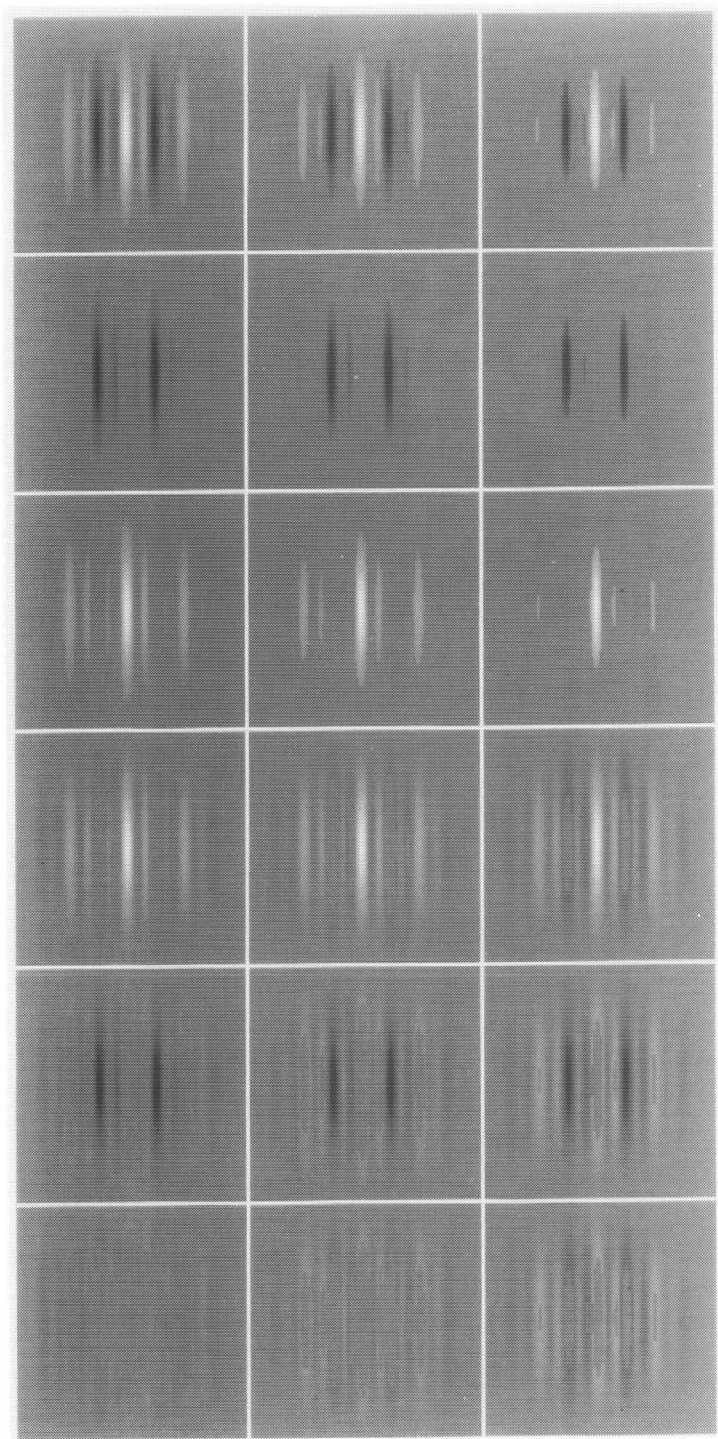


Figure 4. The 18 distorted versions of the first original (shown in Fig. 1 top left).

2.4. Psychophysical procedure

The procedure of supervised learning and classification consisted of a variable number of learning units and employed the *learning set* of 12 original compound Gabor signals. The latter was partitioned into four classes of three signals each. One learning unit contained three subsequent presentations, in random order, of the learning set with a 200 ms exposure duration for each pattern. Following each presentation, a number was displayed for 1 s with an ISI of 500 ms, specifying the class to which the pattern belonged. Each learning unit ended with a test of how well the subject was able to classify the 12 learning signals. Only one exposure per sample was used here. Viewing was always binocular.

The learning procedure continued until the subject had achieved an error-free (100% correct) classification in one learning unit. Following the initial learning stage, the subject was then tested with the first test set. Each pattern of this test set occurred 50 times in random sequence and was classified according to the previously learned classes. Thereafter subjects were re-trained to 100% correct with the learning set, and were tested 50 times with each sample of the second test set. This procedure of alternating units of re-learning and testing was continued until the subjects had been tested with each of the 18 test sets.

To complete this program, a subject had to undergo an average number of 120 learning units, each consisting of $3 \times 12 = 36$ signal presentations with instruction plus 12 classification decisions, and a total 10 800 test decisions. The program was fulfilled during 20 to 30 experimental sessions, which were distributed over a period of about 3 months.

2.5. Subjects

Six paid subjects participated in the experiments, 3 females (G.A., P.E. and R.B.) and 3 males (J.N., H.S. and M.A.). Their ages ranged between 20 and 35 yr, and they had normal or corrected-to-normal vision. Three subjects (J.N., H.S. and R.B.) had participated in the earlier experiments on supervised learning and classification of mirror-image signals reported by Rentschler *et al.* (in preparation), whereas the remaining three had never participated in psychophysical experiments before.

3. RESULTS

Generalization results for the 18 test sets are shown in Fig. 5 with the relative classification frequencies ('classification probabilities') being plotted against the numbers of the twelve test images. Only cumulative data over the group of six observers are shown, although the latter were not entirely consistent in their judgments. The reason for this restriction is our interest in more general aspects of generalization behaviour, against which details of individual characteristics will be neglected. However, individual classification strategies are being considered in another study focusing on the possible applicability of recognition-by-parts strategies from machine intelligence

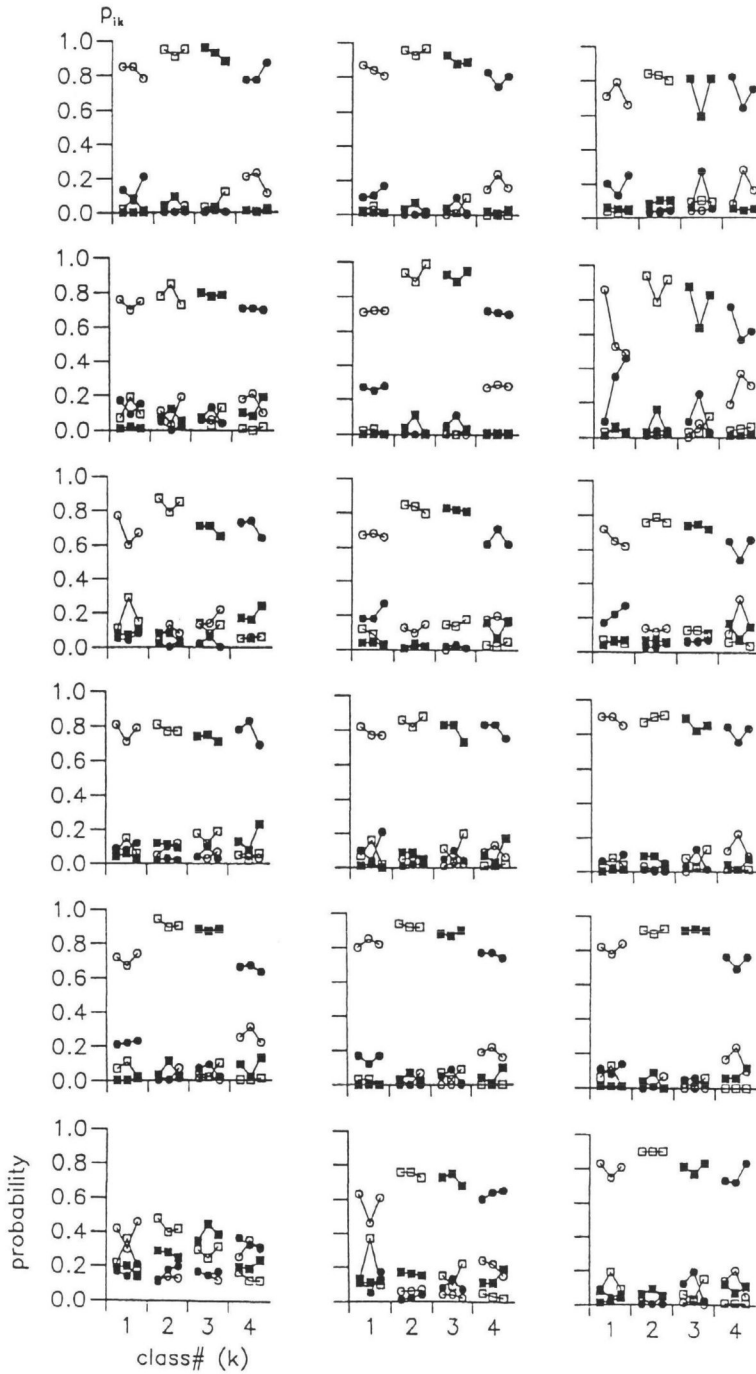


Figure 5. Psychophysically measured generalization performance for the 18 sets of test signals pooled over six observers.

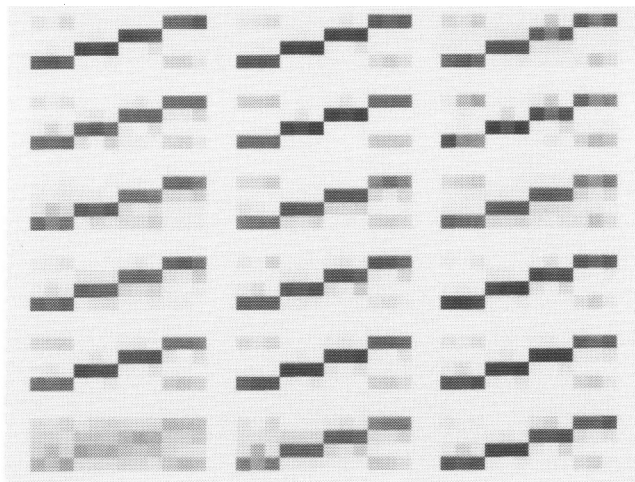


Figure 6. Same results as shown in Fig. 5 but plotted as grey-level matrices as defined in Fig. 2.

(Jüttner *et al.*, 1996). The data on supervised learning of the original signal configuration (learning set) are not shown either. We simply state that subjects were to classify a given set of test signals once they had reached a criterion of 100% correct in classifying the learning set.

For judging the observed generalization performance of our six observers, it is important to note that the 18 types of blob transformations enacted on the learning signals are pairwise complementary with respect to the light and dark contrast polarities. The three blob transformations characterized in rows 1 of Fig. 3 and Fig. 4 *removed* narrow ranges of gray-levels around DC, whereas those of rows 6 exclusively *preserved* the same ranges. Thus it can be said, that, in an informal sense, these transformations removed or preserved 'zero-crossings' (although a low-resolution Laplacian operator might have the same zero crossings when applied to the patterns in row 1). The three transformations characterized in rows 2 removed the *upper ranges* of gray-levels (light image components) plus some range around DC. Those of rows 3 removed the *lower ranges* of gray-levels (dark image components) plus the same range around DC. The three transformations of rows 5 differ from those of rows 2, in that they preserved the gray-level ranges around DC. The same difference exists between the transformations of rows 3 and 4 of Fig. 3 and Fig. 4. Briefly, both pairs of rows 2 and 5, as well as 3 and 4, correspond to light and dark image blob transformations, respectively, with the transformations 4 and 5 preserving the 'background' of gray levels around DC, i.e. the range of zero-crossings.

The overall conclusion from the psychophysical data shown in the same sort of array in Figs 5 and 6 is, that the subjects did surprisingly well in classifying the test signals although the latter resulted from rather drastic image distortions. Scrutiny of the classification data, however, reveals some interesting characteristics of classification behaviour.

A comparison of the respective classification performances corresponding to the 'no zero-crossings' and the 'zero-crossings only' conditions (rows 1 and 6 of Fig. 5)

suggests that the processing of light and dark image components (blobs) is particularly important for the generalization of classification performance. However, it is also clear, that the latter decreased the more image gray-levels are lost from the range around DC (Fig. 5, row 1, rightmost graph, and row 6, leftmost graph).

The test patterns employed for obtaining the data shown in rows 2 or 3 of Fig. 5 contained dark or light blob configurations only (see rows 2 and 3 of Fig. 4). The resulting classification performance is good, with a somewhat better performance for test patterns with dark (row 2) blobs. The classification data shown in rows 4 and 5 are again equivalent with respect to contrast polarity but are generally somewhat better than those depicted in rows 2 and 3. The latter superiority obviously stems from the additional image information provided by preserving the gray-level ranges of zero-crossings in the test signals.

Furthermore, there is virtually no confusion of mirror-image *test* signals, as would be expected from the observers' difficulty to learn the symmetry relationships present within the configuration of learning signals (Rentschler *et al.*, in preparation; see also Introduction). A weakness of the generalization to test patterns with symmetry relationships would show up in a specific pattern of generalization probabilities (see Fig. 7), which cannot be discerned in Fig. 6. This reveals a particularly interesting aspect of the generalization process. Our observers 'rectified' via learning the anisotropy of perceptual similarities between mirror-image and non-mirror-image learning signals of same Euclidean distance in feature space. Surprisingly enough, this enabled them to correctly classify even mirror-image test signals. Whether this should be considered as a *strong form of pattern generalization* or rather constitutes, psychologically speaking, a case of concept (of symmetry) formation is not easy to say.

Summarizing we note three surprising properties of the set of generalization data plotted in Figs 5 and 6. First, generalization performance to blob-only versions of the learning signals was, with a few exceptions, excellent. Second, generalization performance was equally good with test patterns with bipolar blob structure (dark and light blobs), with dark blobs only, and with light blobs only. Third, the ability of classifying mirror-image signals was acquired through learning and could be generalized to the test signals.

As such, these results are not sufficient for drawing direct conclusions about the nature of the visual representations involved. However, the first result is clearly compatible with our hypothesis that compound Gabor signals are encoded in terms of their blob structure. The second result implies that dark *or* light blob pattern components

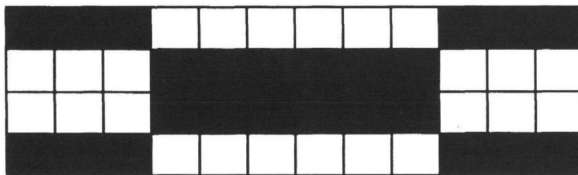


Figure 7. Schematic illustration of the distance pattern that would result in a representation in which the symmetric patterns are closer.

are sufficient for the classification of compound Gabor signals. Hence it is compatible with the hypothesis that the learning and test patterns are internally represented with 'labelled' contrast polarity of blobs. Yet it could also be that the patterns are contained in two different subsystems subserving the representation of the dark and light structures, respectively. This stronger view receives support from the introspective reports of our subjects and from individual classification data, which show that subjects tend to rely for generalization on only one polarity (cf. also Treisman and Gelade, 1980). Neither labelled nor separate ON/OFF representations can be explained within the theoretical framework of linear systems (e.g. Gaskill, 1978), as will be shown in the following section. The third result suggests that generalization is obtained from the same combination of pattern representation and classification rule that has been build up via learning of the original pattern configuration.

Taken together, we have reason to assume that the classification of compound Gabor signals in a general context cannot be modelled in terms of the virtual prototype approach in its present linear form: one factor is the conjecture that such signals may be internally encoded in terms of blob structures; another is that it also seems that dark and light blob structures of compound Gabor signals are separately encoded. In the following section we discuss ways of dealing with these difficulties. The inability of linear systems to capture symmetry relationships between patterns and pattern classes, however, will be dealt with in a separate study (Zetzsche *et al.*, in preparation).

4. COMPUTATIONAL ASPECTS OF PATTERN GENERALIZATION

4.1. Classification and generalization criteria

A straightforward assumption about pattern classification *and* generalization is that learning and test patterns can be represented in one and the same vector space. The basic but certainly not sufficient criterion for classification is then that learning patterns are separable, that is, can be distinguished at all. If generalization is possible, it can further be inferred that the test patterns are located within the class boundaries established through learning (*Generalization Criterion G*) though the discriminant function may be arbitrarily complex (nonlinear).

In technical pattern recognition, the use of *linear* discriminant functions (sets of hyperplanes) is common, and forms the basis of minimum-distance-based classification schemes (e.g. Duda and Hart, 1973). This restriction (*Generalization Criterion G1*) delimits the range of suitable pattern configurations considerably, but does not imply an *intra-class clustering* of learning and test patterns. The distance between test and learning signals of a given class may be much larger than both the intra- and the inter-class distances of the learning patterns, provided the former is less than the between (inter-class) distances of the test signals of a given class and the learning signals of all other classes. The simplest model for the description of our data would thus consist of a single visual representation followed by a minimum distance classifier. Provided a suitable representation has been found, such a model would account both for the fact

that the test signals appear perceptually quite remote from (dissimilar to) the learning set, and for the result that they can be classified correctly.

In our view, however, an even stronger restriction (*Generalization Criterion G2*) appears reasonable as a basis of human generalization capabilities, namely that the learning and test patterns of each class are conjointly *clustered*. This implies that the intra-class distances between learning and test patterns are comparable to the intra-class distances between the individual learning signals of the class, while the inter-class distances to all other patterns are substantially greater. Simply stated, the existence of a visual representation wherein the test and the learning patterns are more similar to each other than they are to any pattern of the other classes is here assumed. The separability of classes via linear discriminant functions is then obvious. However, it is also clear that the global perceptual dissimilarity between learning and test patterns must then be accounted for by assuming a further representation (or metric) which makes this property explicit.

To explore the extent, to which such restrictions apply, it seems reasonable to begin with the computation of *distances* between all the patterns involved and consider relationships between distances. Various distance measures are possible, but with respect to the following it is irrelevant whether Euclidean distances, correlations, or angles (correlations normalized by vector length) between pattern vectors are considered.

We can then proceed to consider how different types of pattern representations (metric feature spaces) conform with the criteria established above. From a technical viewpoint the consideration of Euclidean distances in signal space is a first step. The results of corresponding computations are shown in Fig. 8 (left), where for each of the 18 sets of test patterns the Euclidean distances between the 12 test signals and the original pattern classes (mean pattern vectors or prototypes) are plotted. A gray-level code is used to visualize the distances, whereby increasing darkness indicates decreasing distance, that is, proximity between test patterns and learned prototypes.

A comparison of behavioural classification data and such simulated distances is not trivial. In particular, a full specification of the classification scheme and of the noise sources within the system would be necessary in order to enable a direct comparison (see Rentschler *et al.*, 1994). To keep the computations manageable we have restricted ourselves to two kinds of first-order approximations. First, we computed distance matrices like the ones shown in Fig. 8 for all representations. The values in these matrices do not directly correspond to experimentally obtained percent-correct values (as illustrated in Fig. 6) but are related to the latter as follows: a strong diagonal structure in the distance matrices (as in some matrices in Fig. 8, left) indicates *clustering*, and thereby, *linear separability* (G2, and thereby G1, will be satisfied). This is compatible with the occurrence of strong diagonals in the percent-correct matrices (Fig. 6), indicating almost perfect classification performance of the subjects. It should be noted, however, that this relation is not one-to-one, that is, perfect linear separability may well be possible *without* the emergence of such dominant diagonal structures. To take this into account, we have also computed the correct classifications obtained with a linear classifier being optimal with respect to the learning patterns. The results of these simulations will be discussed in the following section.

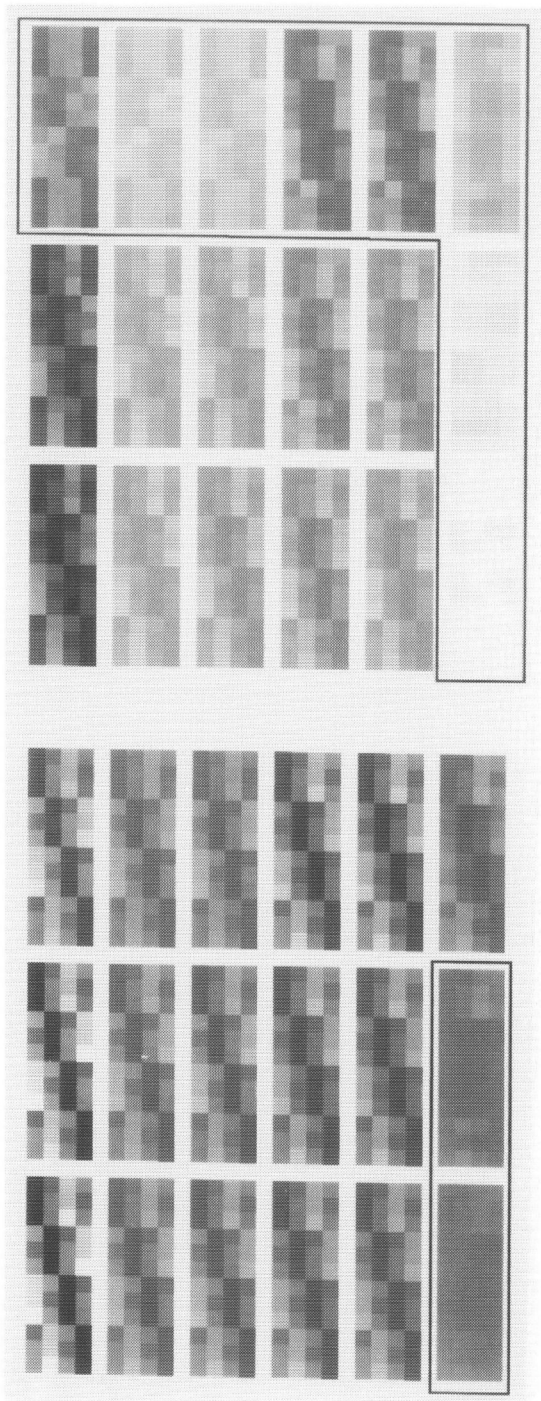


Figure 8. Euclidean distances between original classes (prototypes, mean vectors) and the 18 sets of test signals. The gray-values shown, normalized to 256 levels across all 18 blob versions, are proportional to the distances (the darker the closer). The arrangement of the plotted distances is as follows. The rows of each block contain the 12 distances between the 12 distorted patterns and one of the original classes—see Fig. 2. The resulting 12 by 4 matrices are computed for the 18 distortion types and displayed with the same arrangement as the corresponding patterns in Fig. 4. The results shown on the right are obtained by shifting the 12 signals relative to the prototypes of the 4 classes (mean vectors) such that the plotted values are distances which are minimal with respect to the shift. The minimum-distance rule would imply that each pattern has to be closer to the correct prototype than to all other prototypes (G1). The blob versions which do not yield 100% correct classification according to this criterion are marked.

Already the simple computation of Euclidean distances seems to be a satisfactory model, since the diagonal structures in Fig. 8 (left) compare reasonably well with the behavioural data (Fig. 6), even with respect to the strong clustering criterion G2. Thus it is also possible to satisfy G1, i.e. to find hyperplanes in signal space, which delimit the four classes defined by the clustering of the 12 original patterns in such a way, that all but two blob versions (indicated in Fig. 8, left, by means of a frame) can be classified perfectly.

The direct visual comparison, however, of test and prototype images via zero-shift cross-correlation is an unlikely assumption for the assessment of pattern similarity. The *full* cross-correlation information is needed to represent human performance in image matching (alignment; Caelli and Rentschler, 1986), and this suggests the use of *maximum* correlation (minimum distance with respect to shift-operations) as a behaviourally more realistic distance measure (Fig. 8, right). Switching to this measure changes the situation, in that the patterns shown in the third column and those shown in the sixth row of Fig. 4 are clearly no longer consistent with the proposed strong generalization criterion G2. Evaluation according to the minimum distance rule reveals, that there is even a considerable violation of the weaker criterion G1. The greatest violation occurs for the dark-blob patterns in column 3, row 2, which are classified with 58% correct only. The experimental data (Fig. 6, the means are shown in the last column of Table 1) show no such steep decrease in performance (73% correct are obtained for this distortion type). We conclude that the simple cross-correlation model accounts for basic aspects, like the good generalization of the moderately thresholded dark-only and light-only blob patterns. It does not predict, however, behavioural performance for the extremely thresholded blob versions, which require the use of thin line elements for correct classification (right column of Fig. 4).

A further assumption, specific to the present context, would be that the visual system performs operations which are similar to the operations by which the test sets themselves were generated. Suppose that we chose one of the distortion types shown in Figs 3 and 4 as a model for the preprocessing of all the patterns. It can then be seen that a proper subset of the set of test patterns could be mapped on identical 'representations', such that the strong criterion G2 is satisfied. Our simulations show, however, that the results of such operations depend critically on the precise value of the gray-level thresholds used. The implementation of this type of algorithm in biological vision systems is also not plausible.

4.2. Visual representations

We thus proceed to consider biologically more realistic visual representations. Based on standard assumptions about the function of early visual pathways, we have chosen two representative types of isotropic image representations since there exist no differences in orientation between the patterns employed here. The first one is a Laplacian pyramid (LAP) representation (Burt and Adelson, 1983), which corresponds to the simple centre-surround mechanisms of visual neurophysiology (see Fig. 12 in the Appendix). The second one is a 2D pyramid based on the clipped-eigenvalues (CEV) operator as introduced in our previous work (Zetsche and Barth, 1990a, b;

Barth *et al.*, 1993a). It provides a model of the function of end-stopped neurons. A brief description of the latter formalism, which again uses different levels of resolution, is found in the Appendix (see also Fig. 12).

Based on these representations we added two further computational elements of visual processing. The first one is the ON/OFF separation as motivated in Section 4.3 which leads to the LON/LOFF, CON/COFF representations as described in Fig. 10 (the ON/OFF representations and later the MAX/MIN representations are used in combination with a switching circuit for choosing the appropriate representation with respect to the polarity of the input). The second one is related to the concepts of gain control and local phase (a ratio of even and odd measures) implemented here as the CEVG and LAPG operators (with the ON/OFF versions CONG/COFFG, LONG/LOFFG) defined in the Appendix. The 'G' refers to 'gradient' since the difference between the CEV and CEVG (and equally between LAP and LAPG) operators is that the former were divided by the gradient of image intensity. In the present context such an operation reduces the differences between blobs of different size and contrast.

Furthermore, we considered representations, where the local maxima of the Laplacian and of the CEV-operator are extracted (so-called MAXLAP and MAXCEV representations which will be described in more detail in Section 4.3). To obtain a distance measure for patterns in these representations, the positions of the maxima in scale space were correlated (after slight low-pass-filtering).

For all the representations listed above, we computed and plotted the distance patterns in the same way as shown in Fig. 8 (right), that is, under the assumption of maximum cross-correlation or minimum distance (to reduce computational complexity, we computed the full transforms of the 128 by 128 pixel images but then extracted only the mid row of the transforms before computing the distances). An overview of the results is given in Table 1 for the 18 test sets (in the arrangement of Fig. 4) and the representations specified. The numbers in the table denote the percentage of patterns of a given test set (of 12 patterns), which would be classified correctly according to the minimum-distance rule. In the last column the mean behavioural data are included, again with the caveat, that they cannot be compared directly with the simulation results (see Section 4.1).

The analysis of distance structures summarized in Table 1 failed to yield unequivocal results. In general, most representations can account for the basic effects, such as the good generalization of the dark-only and light-only blob patterns with low and medium thresholds (Fig. 4, left and mid columns). Furthermore, some difficulties observed with the pixelwise representation and Euclidean distances, such as the pronounced drop-off of correct classification for the extremely thresholded test patterns (see Section 2.3 and Fig. 4, right column) could be ameliorated by employing biologically more plausible representations. However, some representations, especially MAXCEV/MINCEV, seem to perform worse than the pixelwise representation according to the minimum-distance rule used here for the simulations.

Still, many representations predict better classification for light-blobs only (row 3) than dark-blobs only (row 2) whereas the behavioural performance is rather better for dark blobs, and, in addition, it is higher for the test set in row 6, column 3, than

Table 1.

Simulated classification results for different representations and mean behavioural performance (in % correct)

Pixelwise	LAP			CEV			LON/LOFF			CON/COFF		
100 100 66	100	100	58	100	100	83	100	100	83	100	100	91
100 100 58	100	100	50	100	83	58	83	83	16	91	83	50
100 100 66	100	100	66	91	91	83	83	83	83	91	91	66
100 100 83	100	100	83	100	100	83	100	100	100	91	91	100
100 100 83	100	100	83	100	100	66	83	83	100	91	91	83
0 0 0	0	0	0	25	16	16	33	0	25	25	25	33

LONG/ LOFFG	CONG/ COFFG			MAXLAP/ MINLAP			MAXCEV/ MINCEV			Behavioural performance		
100 100 100	100	100	100	100	100	50	100	91	41	87	87	75
66 66 8	83	75	33	83	91	58	50	50	25	75	82	73
100 83 66	100	100	58	75	41	58	66	50	50	72	74	69
100 100 100	100	100	100	83	91	100	83	75	66	76	81	85
66 50 66	83	83	66	66	75	100	66	66	75	78	84	85
41 16 8	33	25	25	8	8	66	33	25	58	39	66	81

for both the light-only and dark-only blobs. Further, many representations show a greater variability across test pattern types than was observed with the behavioural results. Whether the above arguments can be used to judge the usefulness of the representations, depends on the full statistical model used for the decision process. We conclude that the present predictions neither provide clear hints as to the nature of the visual representation(s) underlying generalization, nor answer the question as to the decision strategies employed by our observers (minimum-distance rule or others).

4.3. Towards more abstract codes

The previous results are less surprising, given the possibility that both our assumptions on the nature of visual representations and of decision rules (classifier types) underlying the generalization of form are too simplistic to accommodate all aspects of human performance. Yet it seems reasonable to assume that the formidable power of biological pattern recognition originates predominantly from the brain's ability to construct and modify internal representations via learning of pattern structure while relatively simple decision rules are being maintained. The fact, that expertise in sophisticated recognition tasks (e.g. face recognition with foreign ethnicities, radiological diagnosis, plant classification) depends on long learning periods but thereafter allows rapid decision making, seems to support this view. The above-mentioned 'rectification' of the internal representation of mirror-symmetric patterns through learning and the resulting ease of generalization to blob-only versions thereof may be another example of this. Furthermore, this notion follows classical theories of perceptual learning (Gibson, 1969), according to which the learning process involves the development of a structural pattern representation in terms of distinctive features and feature relations. Our further treatment of the generalization problem will thus emphasize the issue of

pattern *structure* as revealed in different representations, while the formal nature of decision behaviour will not be further pursued.

Accordingly, we shall now proceed to reassess the representations under consideration. In particular, we will consider basic neurophysiological and computational aspects, and will refer to the introspective reports of our subjects concerning their *recognition strategies*. Based on the earlier hypothesis, that the subjects' 'internal code' consisted of a relatively abstract description of the patterns, seemingly involving light and dark blob elements with explicit positional relationships ('blob algebra', see Caelli *et al.*, 1986, for an analysis in the context of texture perception), we suggest that the location of blob features in scale space (the MAXCEV representation) provides a suitable representational basis for characterizing generalization performance.

That the treelike configuration of 2D points in scale space could capture the essential image structure was first suggested by Zetsche *et al.* (1993). Subsequent experimental and theoretical investigations of this hypothesis revealed that the maxima of curvature do indeed yield a consistent tree-structure in scale space (Barth *et al.*, 1993b, c). The algorithm, which we propose for computing multi-scale blob features, bears similarities with coding schemes proposed earlier by several authors (Koenderink, 1984; Watt and Morgan, 1985). We propose that the stability of blob features across scale space is the computational basis for determining the similarity of blob structures between learned classes and test patterns. Here the argument for the need of a stability analysis in scale space is essentially the same as that used by Rosenfeld and Thurston (1971) in the context of edge detection, and by Marr and Hildreth (1980) for evaluating the location of zero-crossings, namely, the positional uncertainty associated with the determination of such descriptors at only one level of resolution. The usefulness of 'parsing scale space' for stability analysis in the field of object recognition has been demonstrated by Bischof and Caelli (1988).

The determination of local maxima at multiple levels of resolution has the advantage that it can account for the more abstract notion of blob algebra in the sense that predicates such as the 'existence of a large blob in the centre and a smaller one at the right of it' can be formulated consistently, the size of a blob being encoded by its position in scale space. For these reasons, we computed the locations of the local maxima of the clipped-eigenvalues (CEV) operator for a sequence of different resolutions, i.e. 16 scales spanning two octaves (the same frequency range as in Fig. 12 with a denser sampling for better illustration). These locations of the blobs were plotted for a horizontal cut at half-height through the learning and test images as a function of scale.

The plots of the maxima across scale (Fig. 9, MAXCEV representation) reveal a relative invariance with respect to light-blob operations (Fig. 9, from left to right), thus providing a basis for the proposed requirement of *similarity of form* assumed to underlie generalization performance. This invariance is closely related to the strong generalization criterion G2, which requires a sufficient similarity between the representations of the corresponding training and test patterns. Yet the MAXCEV representation still allows the discrimination between patterns of different classes (Fig. 9, from top to bottom). The minima (MINCEV representation) display the same type of stability for dark-blob operations. This suggests, that the two representations can

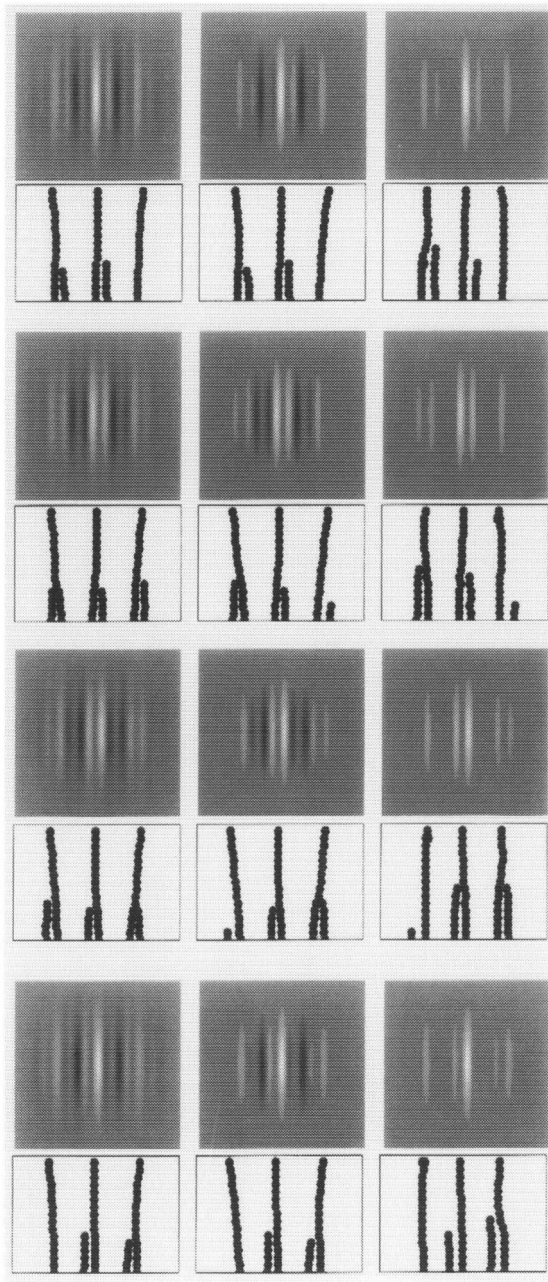


Figure 9. Locations of bright blobs plotted as a function of scale are shown below the corresponding images. The arrangement is as follows. From top to bottom the pattern number varies (the first patterns of each class are shown, i.e. patterns 1, 4, 7, and 10—see Fig. 2), and from left to right the distortion type (mid patterns of rows 1 and 3 in Fig. 4). The locations of the blobs were obtained in two steps. First, we computed the 2D-representations as shown in Fig. 12 on the right (the same range of scales was used but with a finer sampling of 16 scales). Second, the middle row of the images were extracted and those locations were selected for plotting, for which left and right neighbours were smaller (local maxima).

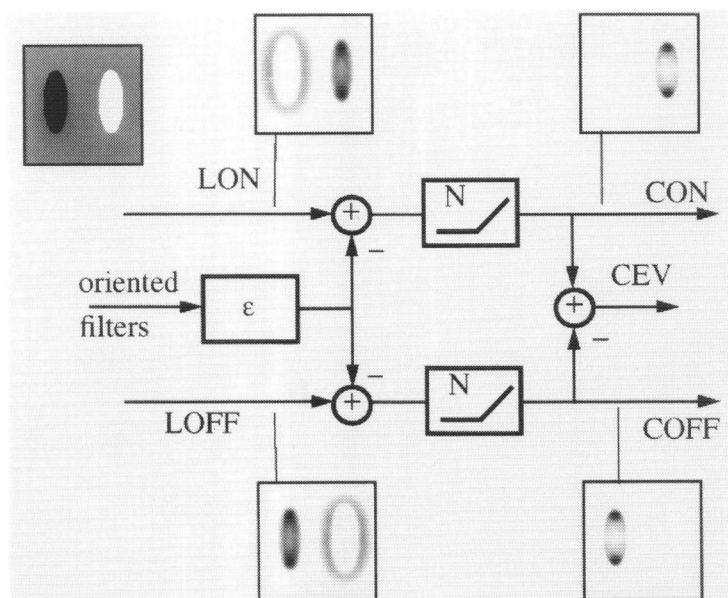


Figure 10. The Laplacian and the CEV operator (for details see Appendix). 'LON' and 'LOFF' denote the ON and OFF components as obtained by one-way rectification of the Laplacian. The eccentricity ε (defined in Eqn (A2)) results from nonlinear inhibition between oriented filters. The resulting 'CON' and 'COFF' components of the CEV operator differ in two ways from the 'LON' and 'LOFF' components. First, they are 2D-operators, i.e. they equal zero for straight patterns, and respond only to curved features like corners and line ends. Even more important for the current context, they differ with respect to the ON/OFF separation as follows. A 'CON' operator responds only to light-on-dark features, whereas a 'LON' operator will respond also to dark-on-light features as illustrated for the image shown top left.

be employed for obtaining a pattern recognition scheme with the required degrees of invariance (generalization) and differentiation (discrimination between classes).

Similar results, but with a reduced invariance to blob operations, can be obtained by using the Laplacian for generating representations in scale space (MAXLAP representation). This approach has the general problem of insufficient ON/OFF-separation, which leads to 'false responses' in the MAXLAP representation of dark patterns and in the MINLAP representation of bright patterns (see Fig. 10). The use of the 2D operator is sufficient to avoid this problem but we do not know yet whether it is also necessary here. However, it seems that there exists a relationship between clean ON/OFF separation and the suppression of straight patterns, which 2D operators yield by definition. An alternative nonlinear method for the construction of separate ON and OFF representations would be a model of the MIRAGE-type as proposed by Watt and Morgan (1985), which starts with a set of spatial filters, the outputs of which are combined in such a way that negative and positive response components are carried separately. Whether such a model can also avoid the emergence of 'spurious' blobs remains to be investigated.

According to Fig. 9 the MAXCEV representation has the essential features proposed here as a basis of successful generalization: sufficient similarity within classes, as well as sufficient dissimilarity between classes. The principle illustrated in Fig. 9 has, of

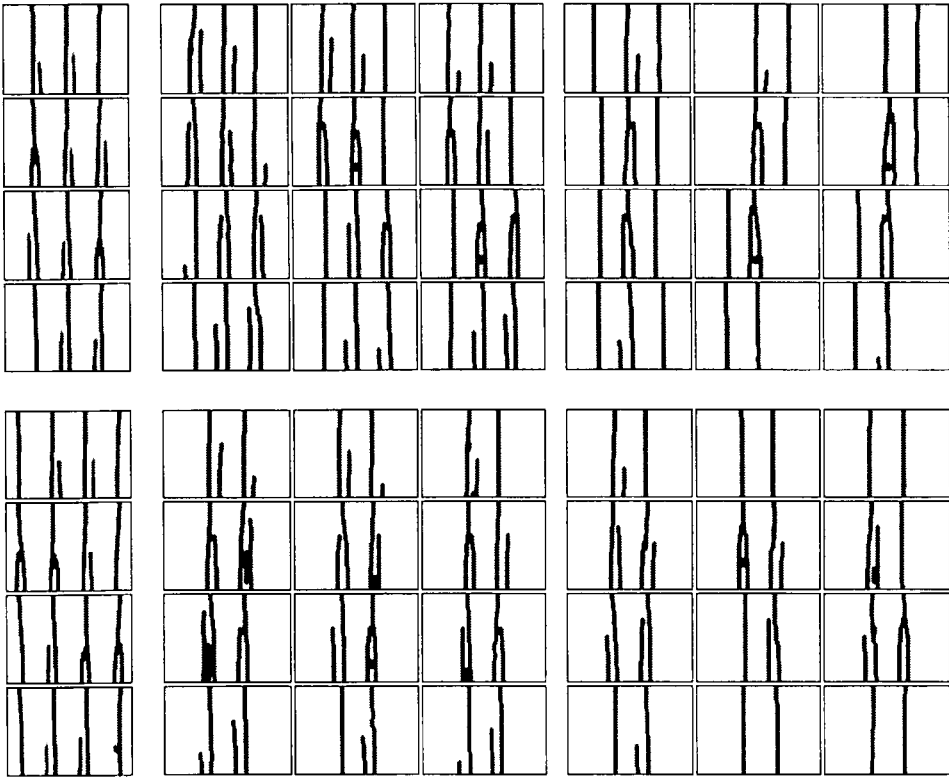


Figure 11. MAXCEV representations (upper half) and MINCEV representations (lower half) for complete sets of 12 test patterns (arranged as in Fig. 1) and for two blob types with medium and extreme thresholds (columns 2 and 3 in Fig. 4). The left column shows the 4 prototypes (class means of the original patterns). The upper 3 by 4 blocks show the MAXCEV representations of the light-only blob patterns with medium threshold (left block, corresponding to row 3 column 2 in Fig. 4) and extreme threshold (right block, corresponding to row 3 column 2 in Fig. 4). The lower two blocks show the MINCEV representations of the dark-only blob patterns, again with medium and extreme thresholds. The locations of maxima are here not illustrated by plotting circles (as in the previous two figures) but by setting pixels to one or zero, depending on whether they correspond to maxima or not, and then low-pass filtering the resulting binary images (the images shown here have been used to obtain the distance measures in Table 1). The first column of the upper left 3 by 4 block shows the same maxima as the right column in Fig. 9.

course, its limits. Figure 11 shows the MAXCEV and the MINCEV representations for all 12 patterns of a given blob type including the most critical blob operations for which the similarity of the representations decreases. Nevertheless, we assume that there still is sufficient similarity within classes, as well as sufficient dissimilarity across classes to allow for a better agreement between data and predictions than obtained by the other representations in Table 1.

However, the above mentioned features cannot be taken advantage of by applying a simple correlation and minimum-distance rule (see again Table 1). One possible way of taking account of this difficulty is to develop a notion of similarity based on the *explicit structural properties* of the patterns. Accordingly, future research should

investigate more abstract codes that can be derived from the MINCEV/MAXCEV representation. Such codes may, for example, explicitly represent both attributes of individual blobs (e.g. position or size) and relational attributes of pairs of blobs (e.g. distance or size ratio). It may then be possible that relatively simple classifier structures acting on this representation could account for the experimental data.

Alternatively, such structure-oriented codes may be used for a propositional scheme of representation operated upon by a rule-based classification process. Indeed, Jüttner *et al.* (1996) demonstrated that such an approach could predict basic aspects of both the individual and group performance in learning and generalizing the patterns in the present experiment. Their study, however, is based on the decomposition of patterns into parts defined in the image domain, thus avoiding assumptions about stimulus encoding in terms of known filter properties of the visual system. Hence the present study and that by Jüttner *et al.* (1996) are complementary in the sense that their emphasis is more on aspects of early vision and cognition, respectively. Which aspects of these two concepts have to be combined to provide a more realistic scheme of visual classification and generalization is a question for future research.

5. DISCUSSION

We examined the extent to which human observers are able to generalize classification performance learned with compound Gabor signals that are pairwise mirror-symmetric to each other. The test signals used were blob-only (gray-scale thresholded) versions of the learning signals. Subjects were trained to criterion in classifying the learning signals, and then tested with the learning sets one at a time. We found that generalization was possible to a remarkably large extent. Without further learning, our subjects showed excellent classification capabilities for bipolar blob-only versions of the learning signals (i.e. patterns containing both light *and* dark blob information), as well as for test patterns with light blobs *or* dark blobs only. Likewise, the ability to recognize symmetry relationships was readily generalized.

These psychophysical findings suggest that the learning signals are internally represented in a way which is similar to the distorted test signals. Briefly, the signals seem to be internally represented in terms of a 'blob code', as has been conjectured by Caelli *et al.* (1986) in the context of texture discrimination. Furthermore, the observed equivalence for generalization performance of dark-only and light-only blob codes, taken together with the introspective reports of our subjects, suggests that labelled or even separate representations exist for the two types of contrast polarities.

Assuming separate encoding processes for dark and light pattern components is consistent with neurophysiological results concerning the existence of two functionally separate neural subsystems. Indeed, a 'darkness' (OFF) system and a 'lightness' (ON) system have been discussed by neurophysiologists in an attempt to find correlates for the perception of darkness and brightness (for reviews see Fiorentini *et al.*, 1990; Schiller, 1992). These channels have been related previously to the visual tasks of detecting and discriminating light increments and decrements. Our results suggest

that ON- and OFF-type subsystems may be functional even at the level of spatial visual representations underlying pattern recognition, i.e. cognitive functions.

As a first step in our computational analysis we investigated whether our data can be described within the standard framework for technical pattern recognition. Since distances in feature space are the essential values required for the application of minimum-distance classifier schemes, we started with the evaluation of distance measures with respect to the image pixel configurations and with respect to different visual representations (Laplacian and CEV operators including ON/OFF separation and gradient normalization). The results of our simulations showed that all representations can basically capture the generalization of the light-only and dark-only blob patterns but fail to predict human performance for the extreme blob operations. We concluded that both the assumptions about representation and classifier are insufficient for adequately characterizing human strategies of pattern classification and generalization.

This situation led us to reconsider the structural properties of possible representations with respect to basic neurophysiological and computational aspects as well as the introspective reports of our subjects. We found that, mainly due to the consistent separation of ON and OFF signals, the MINCEV/MAXCEV representation best matches the notion of an ideal blob-code which accommodates the proposed requirement of *pattern similarity* assumed to underlie the generalization of form. Thus it is important to note that the suggested representation is introduced not in an ad-hoc fashion for this special context but in accordance with the views of several authors regarding general aspects of human pattern recognition (Marr and Hildreth, 1980; Watt and Morgan, 1985; Bischof and Caelli, 1988; Zetsche *et al.*, 1993).

However, the ability to establish *pattern similarities* must be complemented by *pattern discrimination*, since subjects can distinguish between learning and the test patterns. This complementarity is already explicit in Pavlovian conditioning. On the one hand, a subject will give a conditioned response to the conditioned stimulus but, to some extent, to similar stimuli as well. On the other hand, a subject can also be trained to discriminate stimuli by pairing the conditioned and the unconditioned stimuli, while similar stimuli are presented unpaired (see Brown and Herrnstein, 1975, Chapter 3). Similarly, Marr and Nishihara (1978, p. 272) list 'stability and sensitivity' among the criteria for judging the effectiveness of shape recognition systems: 'To be useful for recognition, the degree of similarity between two shapes must be reflected in their descriptions but at the same time even subtle differences must be expressible. These opposing conditions can be satisfied only if it is possible to decouple stable information that captures more general and less varying properties of shape, from information that is sensitive to the finer distinctions between shapes.' A possible neurophysiological basis of this functional complementarity has recently been found by Fujita *et al.* (1992) in the columnar structure of the infero-temporal cortex of the macaque. In this context, we suggest that the structure of maxima across scale captures the 'stable information' whereas the full representation allows for 'finer distinctions' between learning and test patterns.

Further, we have argued that the locations of maxima in scale space are an adequate basis for more abstract codes which would involve the extraction of more explicit

structural properties of such representations. Whether the resulting representations can then already be handled by relatively simple classifier schemes, or whether they require a more complicated subsequent elaboration of a propositional scheme operated upon by a rule-based classification process, as suggested by Jüttner *et al.* (1996), remains a topic for future research.

Acknowledgement

The study has been supported by a grant from the Deutsche Forschungsgemeinschaft (DFG-Re 337/7) to I.R. and C.Z.

APPENDIX

1. Representations of blob-patterns are shown in Fig. 12 (Laplacian pyramid on the left and 2D-pyramid on the right).

2. The clipped-eigenvalues operator is defined as follows. The dot-responsive cell model as introduced by Zetzsche and Barth (1990a) starts from the determinant of the Hessian, which can be written as

$$G = \frac{1}{4}(l_{xx} + l_{yy})^2 - \frac{1}{4}(l_{xx} - l_{yy})^2 - l_{xy}^2 = (\nabla^2 l)^2 - \varepsilon^2,$$

where $l(x, y)$ denotes image intensity at position (x, y) .

The Laplacian and eccentricity ε determine the eigenvalues of the Hessian:

$$d_{1,2} = \nabla^2 l \pm \varepsilon.$$

The clipped-eigenvalues operator is then defined as

$$\text{CEV} = d_2^+ - d_1^-, \quad (\text{A1})$$

where the clipping operation d^\pm takes the positive and negative values respectively (one-way rectification). Note that CEV equals the Laplacian for isotropic patches where $\varepsilon = 0$. The eccentricity ε can be written as

$$\varepsilon = \frac{1}{4}(l_{xx} - l_{yy})^2 + \frac{1}{4}(l_{uu} - l_{vv})^2, \quad (\text{A2})$$

where the coordinates (u, v) are rotated relative to (x, y) by 45° . Thus, the eccentricity can be seen as resulting from four oriented filters l_{xx} , l_{yy} , l_{uu} , l_{vv} with vertical, horizontal, and oblique orientations—see Fig. 10. The operator CEV has been described and generalized in Zetzsche and Barth (1990b) and Barth *et al.* (1993a), and their relationships to other curvature estimates have been discussed there. The multi-scale version of CEV is obtained by blurring the intensity $l(x, y)$ with a Gaussian kernel. Equivalently, one could use blurred convolution kernels for computing the derivatives.

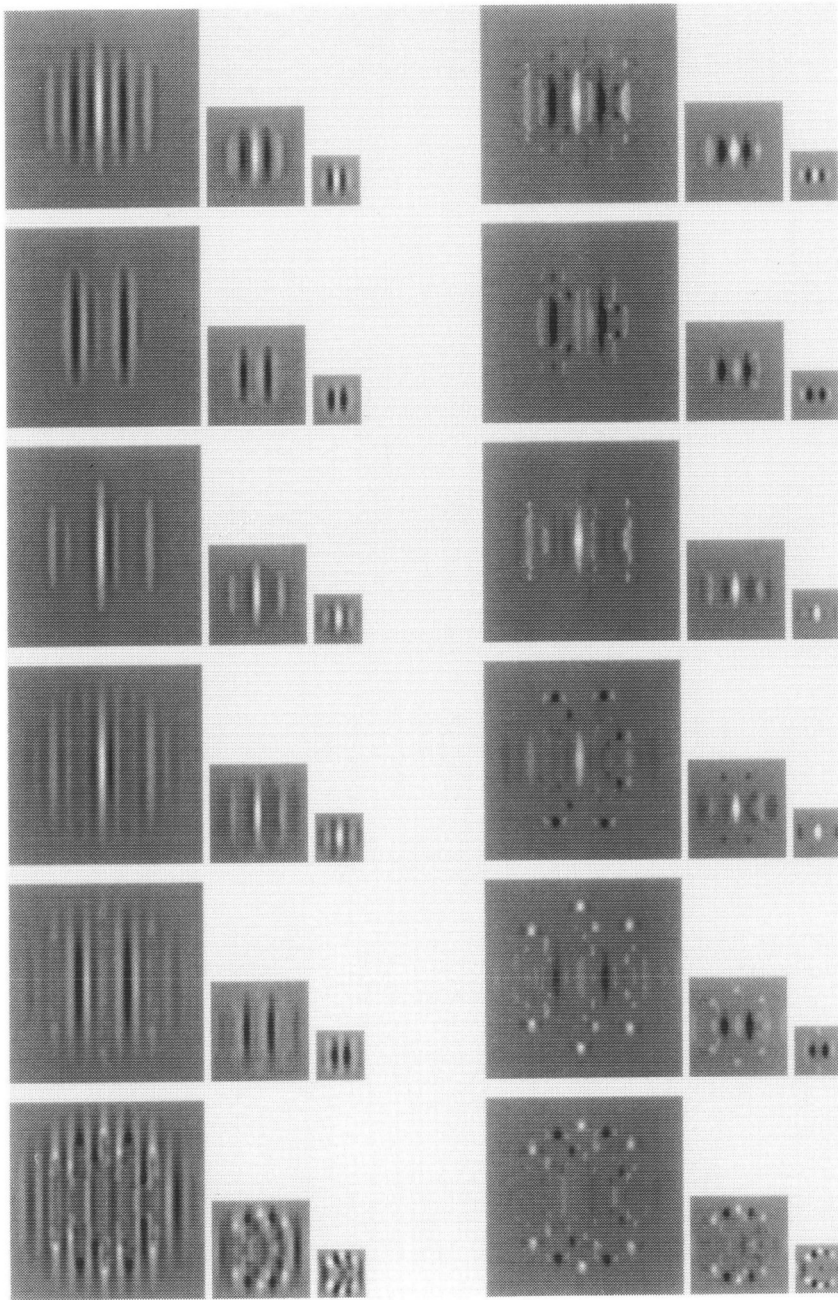


Figure 12. Laplacian-pyramid (left) and 2D-pyramid (clipped-eigenvalues operator as defined by Eqn (A1)) representations (right) on three different scales for 6 distorted versions of the first original pattern (these distortions are shown in the mid column of Fig. 4). The representations include a threshold of 5% of the maximum and are the basis for the computation of the predicted classification results shown in Table 1 and the local maxima shown in Fig. 9. Note that the blobs extracted by the CEV operator, as opposed to the Laplacian, correspond to those which an observer extracts by inspection and verbal description. This is due to the LON/LOFF versus CON/COFF difference shown in Fig. 10.

3. The CEVG and LAPG operators are defined as

$$\text{CEVG} = \frac{\text{CEV}}{l_x^2 + l_y^2 + c}$$

and

$$\text{LAPG} = \frac{\text{LAP}}{l_x^2 + l_y^2 + c},$$

with the value of the constant c chosen as 10% of the maximum of $l_x^2 + l_y^2$.

REFERENCES

- Barth, E., Caelli, T. and Zetsche, C. (1993a). Image encoding, labelling and reconstruction from differential geometry. *CVGIP: Graphical Models and Image Processing* **55**, 428–446.
- Barth, E., Ferraro, M. and Zetsche, C. (1993b). On the structure of 2D-curvature scale space. *Perception* **22** (Suppl.): 21.
- Barth, E., Zetsche, C., Ferraro, M. and Rentschler, I. (1993c). Fractal properties from 2D-curvature on multiple scales. In: *Geometric Methods in Computer Vision II*. B. Vemuri (Ed.). Proceedings SPIE **2031**, pp. 87–99.
- Bennett, P. J. and Banks, M. S. (1987). Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision. *Nature* **326**, 873–876.
- Bennett, P. J. and Banks, M. S. (1991). The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Res.* **31**(10), 1759–1786.
- Bischof, W. F. and Caelli, T. (1988). Parsing scale-space and spatial stability analysis. *CVGIP* **42**, 192–205.
- Brown, R. and Herrnstein, R. J. (1975). *Psychology*. Methuen and Co Ltd, London.
- Burt, P. J. and Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **COM 31**(4), 532–540.
- Caelli, T. M. and Rentschler, I. (1986). Cross-correlation model for pattern acuity. *J. Opt. Soc. Am. A* **3**, 1948–1956.
- Caelli, T. M., Hübner, M. and Rentschler, I. (1986). On the discrimination of micropatterns and textures. *Human Neurobiology* **5**, 129–136.
- Caelli, T., Rentschler, I. and Scheidler, W. (1987). Visual pattern recognition in humans. I. Evidence for the existence of adaptive filters. *Biol. Cybern.* **57**, 233–240.
- Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. Wiley, New York.
- Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P. H. and Thomas, J. P. (1990). The perception of brightness and darkness. Relations to neuronal receptive fields. In: *Visual Perception. The Neurophysiological Foundations*, Ch. 7. L. Spillmann and J. S. Werner (Eds). Academic Press, New York.
- Fujita, I., Tanaka, K. and Cheng, M. I. K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346.
- Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics*. Wiley, New York.
- Gauss, C. F. (1809). *Theoria Motus*. English translation (1963): *Theory of the Motion of the Heavenly Bodies about the Sun in the Conic Sections*. Dover, New York.
- Gelb, A. (1974). *Applied Optimal Estimation*. MIT Press, Cambridge, MA.
- Gibson, E. (1969). *Principles of Perceptual Learning and Development*. Appleton, New York.
- Jüttner, M. and Rentschler, I. (1996). Reduced perceptual dimensionality in extrafoveal vision. *Vision Res.* **36**(7), 1007–1022.
- Jüttner, M., Caelli, T. and Rentschler, I. (1996). Recognition-by-parts: A computational approach to human learning and generalization of shapes. *Biol. Cybern.* (in press).
- Koenderink, J. J. (1984). The structure of images. *Biol. Cybern.* **50**, 363–370.
- Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organisation of three-dimensional shapes. *Proc. R. Soc. London, Ser. B* **200**, 269–294.

- Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proc. Royal Soc. London, Ser. B* **207**, 187–217.
- Rentschler, I. and Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature* **313**, 308–310.
- Rentschler, I., Hübner, M. and Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Res.* **28**(2), 279–291.
- Rentschler, I., Jüttner, M. and Caelli, T. (1994). Probabilistic analysis of human supervised learning and classification. *Vision Res.* **34**, 669–687.
- Rosenfeld, A. and Thurston, M. (1971). Edge and curve detection for visual scene analysis. *IEEE Trans. Comput.* **C-20**, 562–569.
- Saarinen, J. (1987). Perception of positional relationships between line segments in eccentric vision. *Perception* **16**, 583–591.
- Saarinen, J. (1988). The effect of exposure duration on the analysis of spatial structure in eccentric vision. *Spatial Vision* **3**, 1–7.
- Schiller, P. H. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences* **15**(3), 86–92.
- Shepard, R. N. (1962a). The analysis of proximities: multidimensional scaling with a unknown distance function. i. *Psychometrika* **27**, 125–140.
- Shepard, R. N. (1962b). The analysis of proximities: multidimensional scaling with a unknown distance function. ii. *Psychometrika* **27**, 219–246.
- Thurstone, L. L. (1927). Psychophysical analysis. *Am. J. Psychol.* **38**, 368–389.
- Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology* **12**, 97–136.
- Watt, R. J. and Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Res.* **25**, 1661–1674.
- Zetzsche, C. and Barth, E. (1990a). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Res.* **30**, 1111–1117.
- Zetzsche, C. and Barth, E. (1990b). Image surface predicates and the neural encoding of two-dimensional signal variation. In: *Human Vision and Electronic Imaging: Models, Methods, and Applications*. B. Rogowitz (Ed.). Proceedings SPIE 1249, pp. 160–177.
- Zetzsche, C., Barth, E. and Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding. In: *Digital Images and Human Vision*. A. B. Watson (Ed.). MIT Press, Cambridge, MA, pp. 107–138.