Human Neurobiol (1986) 5:37–47

Human Neurobiology
© Springer-Verlag 1986

On functional concepts for the explanation of visual pattern recognition

H. Glünder

Lehrstuhl für Nachrichtentechnik, Technische Universität München, Arcisstr. 21, D-8000 München 2, FRG

Received: 7 February 1985

Summary. An overview is given covering approximately 30 years of research in the field of functional concepts for the explanation of visual pattern recognition processes. Due to its important role, the Perceptron approach is briefly sketched. This concept, representing to a great extent the present opinion about pattern recognition in the field of visual sciences, is investigated with respect to its invariance properties and its structure. Fundamental facts about the nature of human perception (Gestalt-aspect) and about the processing structure of the visual system are used as qualification criteria. Alternatively, a recent concept that explains the high degree of invariance in visual perception by the evaluation of inner pattern relations is introduced and investigated with respect to its biological plausibility. It is also compared with the Perceptron concept.

Key words: Invariant features – Perceptrons – Relational concepts

Critical introduction

In the 1950s the need for automatic image data evaluation and pattern recognition, mainly for aerial reconnaissance purposes, increased considerably. Hence, while looking for new ideas, physicists and engineers became interested in mechanisms[1] involved in visual perception. Thus, one branch of the so-called Bionics research was founded. For the first time since the era of universal researchers, it was attempted to bring such different scientific branches as engineering sciences and technology together officially with biology and human sciences. Scientists from the latter group expected help from this collaboration in the explanation of biological phenomena and for the purpose of modelling biological systems (Biological Cybernetics). Today we know that the different interests of the two groups did not allow a stable interconnection and that, with some exceptions, the expected cooperation failed[2].

The first and perhaps the most important exception, besides the activities of the forerunners Rashevsky, Wiener, Pitts, McCulloch, v. Foerster and others around them, was the work of Rosenblatt (1957): a perception theory called Perceptron. This approach, at the time unique in its promises (e.g., an enormous degree of invariance was ascribed to this principle), was meant to serve engineers as well as biologists. Due to some shortcomings (cf. Minsky and Papert 1969) and to some unsuccessful applications the concept disappeared – at least temporarily – from the field of technical pattern recognition. Another reason for this failure was parallel processing, which is indispensable for the effectiveness of a Perceptron machine. Special purpose computers, however, were not attractive at a time when sequentially working general purpose computers entered the laboratories. Since the Perceptron concept is based on few but fundamental neuroanatomical and neurophysiological results (e.g., receptive fields, Hartline 1940; Kuffler 1953; Hubel and Wiesel 1959) it is no surprise that it still represents a common opinion in visual sciences of how pattern recognition occurs in mammals.

While engineers in the field of pattern recognition no longer cared very much about the biological plausibility of their algorithms (they tried to fit them to the capabilities of digital computers based on the classical principle of v. Neumann (see Fu 1968; Andrews 1972) some physiologists got acquainted with a well-known method used in the field of electrical engineering: the description of signals and systems in terms of (spatial) frequency components, and signal analysis by means of (spatial) frequency filters (Fourier analysis[3]). This suggestion for a certain *system description*[4], however, was converted into a *functional principle* that mainly originates from Campbell and Robson (1968). Perhaps as a consequence of the fact that this spectral filtering approach to pattern analysis is consistent with the Perceptron principle, it became even more accepted.

The lack of succeeding alternatives[5] for a functional description of biological structures performing two-dimensional pattern recognition can be ascribed to the following fact: in most cases it is easy to fit the specific and isolated results from neurobiological research to the well established and general framework of Perceptron-like concepts[6]. This

[1] That implies already a mechanistic view!

[2] It can be hoped that the present activities of bringing together the fields of human and machine vision are successful in the long run (e.g., Beck et al. 1983; Braddick and Sleigh 1983)

[3] A thorough and modern treatment of the two-dimensional Fourier transformation and its application to systems and signals is given by Gaskill (1978)

[4] Schade (1956) already used the Fourier calculus for the definition of the overall bandwidth of linear processing stages in early vision

[5] Surely there were some – but perhaps not at the right time in the right journal, e.g., Hoffman (1966)

[6] Especially if the concept itself influences the direction of further research (Taube 1961)

is especially true for detailed results that are produced without regard to their functional importance for the visual system. Particularly in the neurophysiological field of vision research the systems purpose is often neglected in favour of a somewhat purposeless structure: the Perceptron idea manifests a paradigm (in the sense originally introduced by Kuhn 1970). In a similar fashion most of the methods that are developed in the field of machine vision are a consequence of another paradigm: the principle of sequential v. Neumann-type computers.

There are logical reasons why the Perceptron concept cannot explain some evident performances of biological pattern recognition structures and some others only at unreasonably high expense. Thus, the time seems ripe for better concepts of similar extent, even at the risk of a sudden falsification, e.g., by completely revolutionary results from the biological side. Since verification in empirical sciences is not possible, profound falsification is essential for scientific progress (Popper 1972). In this article the basic idea of Perceptrons is explained and its inappropriateness for the explanation of visual perception is demonstrated. Next, a completely different, "relational"[7] approach is introduced, which is not yet fully developed but should be discussed. It is possible that such processing could be attractive even for technical applications. The present activity in proposing and building parallel working computers – some of which are well suited for the processing of image data (cf. Onoe et al. 1981) – supports this speculation.

Criteria for the assessment of pattern recognition concepts

This paper is concerned with explaining the importance of systems analysis in understanding the capabilities of some recognition concepts. Structural plausibilities (computing structure, i.e., hardware) and that of their specific mechanisms, as well as the relation between complexity and performance are discussed secondarily. The concepts will be investigated with respect to some elementary and obvious abilities demonstrated preferably by the human visual system. These are essentially those demonstrated in everyday life that allow us to perceive what is called the Gestalt of a pattern, i.e., characteristic properties such as symmetries, congruences, similarities etc. (For an explanation of the term "Gestalt" see Wertheimer 1923; Koffka 1935; Metzger 1975, or the more recent articles about the worth of Gestalt theory in a book edited by Beck 1982.) The degree of biological plausibility of processing structures is measured solely against some elementary neurobiological facts that have long been accepted: parallel processing configuration, neural signal representation as impulse rates, spatially and temporally integrating properties of neurons, average number of about 10^4 connections per neuron, layered cortex structure, retinotopic mapping from the receptors to the cortical input layer, etc. One reason for this limited choice can be seen in the systems view of this investigation, and another is given by the enormous complexity of the "object" under investigation in neurosciences, which results in a slow and pointilistic consolidation of its understanding.

In order to confine the topic even more we shall deal, biologically speaking, with the monocular (cyclopic), foveal (about one degree of visual angle), achromatic and suprathreshold vision without eye movements (but not in the sense of stabilized retinal images). Such heavy constraints, however, can make it difficult to find acceptable or even true explanations and theories. The confinement to the monocular and achromatic vision is possible if binocular and colour vision are assumed as conservative extensions (phylogeny) of a basic processing (cf. Livingstone and Hubel 1984). The restriction to foveal vision without eye movements is tolerable, or even plausible, if this type of recognition is seen as an underlying process of a sequential pattern analysis that involves scanning mechanisms (cf. Leibowitz and Post 1982). These can consist of comparative eye movements or "mental shifts" between points of "focal attention" (as Julesz and Bergen 1983 call it). The following investigations concern this local and rapid processing (down to about 50 ms) that takes place between the shifts.

Before we proceed to a more detailed and sometimes technical discussion about pattern recognition, a comment on the biological phenomenon called "receptive field" (Hubel and Wiesel 1962) shall be made: primarily it must be recognized as a consequence of the applied measuring technique. The fact alone that receptive fields can be measured, however, does not suffice for the conclusion that they characterize visual pattern processing. One could, for example, imagine parallel computing structures in which receptive fields can be measured whose purpose or function, however, cannot be deduced from these findings. The question arises whether receptive fields – at least of the complex and hypercomplex type – are of functional relevance for the processing of visual information?

Elements of recognizing structures

In this section basic components of pattern recognizing systems are introduced. The diagram of a general pattern recognition structure (Fig. 1) is used to explain what is commonly

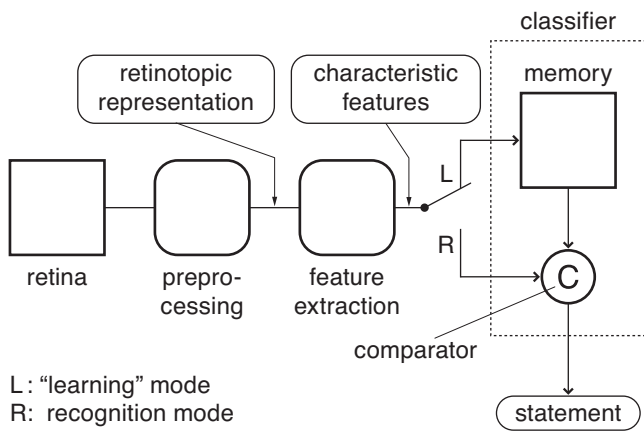

Fig. 1. General scheme of a pattern recognition system

[7] The term "relational" is not used in the same sense as in connection with syntactical and structural approaches to pattern description, as in the works of Barrow and Popplestone (1971), or Shapiro and Haralick (1982). Here, the term shall specify a procedure that leads to invariant features. These, however, can also be used for syntactical descriptions when applied at a higher processing level.

understood as pictorial pattern recognition. One should bear in mind that this scheme and its describing terms stem from engineering sciences. The structure proved useful for technical purposes, but must be judged critically when used in conjunction with biological systems.

The preprocessing stage performs operations such as compression and normalization of signal amplitudes, and contour extraction (Marr 1982; Hofer and Platzer 1978). It is typical for this unit that its output is a "retinotopic" pattern representation.

The feature extraction unit shall provide suitable input signals for the classifier section. They are either stored during the "learning" mode (switch in L-position), or compared with the stored data in the "recognition" mode (switch in R-position). In order to minimize the effort it is advantageous to use characteristic features instead of whole patterns for this purpose. Thus, irrelevant information is removed and significant properties of a pattern are expressed by a so-called feature vector.

The classifier performs the above-mentioned comparison based on these feature vectors. This operation should lead to decisions about the membership of patterns to certain pattern classes. The quality of classification, i.e., the ability to distinguish between patterns belonging to different classes, strongly depends on the choice of the features. If they are known it is possible to specify the performance of the so-called optimum classifier (for details see Sebestyen 1962; Fukunaga 1972). It is not possible, however, to specify in a mathematical sense the optimal features for a certain recognition task. Hence, for the purpose of working out concepts explaining visual pattern processing, the best suited features with respect to the following criteria must be found: the recognition task, presented by general biological and behavioural theories; the potential processing power of the visual system, studied by experimental psychologists and psychophysicists; its processing structure and mechanisms, investigated in the fields of neuroanatomy and neurophysiology.

Invariant recognition in technical and biological systems

While reading the previous description doubts about the separation into functional blocks may have arisen. Particularly the classifier stage should be critically investigated. Until now, technical pattern recognition dealt with problems that are characterized by finite, typically small numbers of pattern classes (as in the case of character recognition). Under such circumstances there is some hope of finding acceptably "separating" features by heuristic search. In the case of visual perception, however, the number of classes is not limited (if we leave out of consideration the finite extent and resolution of the fovea). Another significant difference exists in the magnitude of classes, in other words, the number of representations belonging to a class. Due to their well defined tasks technical systems rarely need to be invariant against many different forms of pattern presentations. These are defined by the type (quality) of transformation that relates the individual pattern to its prototype. Even if this invariance should be necessary, its extent is usually limited to small deviations (quantity) from a prototype representation. For a living creature, however, invariant recognition to the utmost extent might be vital for its survival.

There are four fundamental approaches to invariant pattern recognition[8]:

1. A classifier is used that is tolerant against feature variations. This can be achieved by storing "blurred" features which, in the end, will yield a reduced quality of classification. Furthermore, this method does not enable statements about the actual variance (quality as well as quantity) between a pattern and its prototype; although it represents important information, especially for biological systems. For example, it is a key feature for the specification of the meaning of a percept, preferably in a contextual situation. Figure 2 shows a simple example in which the letter "H", as a whole, is used as a feature. For the purpose of restricted rotational invariant recognition (±45°) a "blurred" feature is formed (Fig. 2c shows the binary template). This may lead to false classifications, e.g., for the letters shown in Figure 2d.

2. The feature vectors are stored separately for every pattern representation. Although this principle allows the most detailed statements of classification, it obviously requires a storage capacity that grows inflationary with the number of pattern classes (Bremermann 1971): the degree of invariance depends on the effort. Figure 3 presents five patterns belonging to the class "H". If these are stored as features they will allow very detailed statements of classification, e.g., prototype (a), inclined by 45° to the right (b), broad version (c), small version (d) and slim version (e).

3. The idea may arise to adapt either the presented feature vectors to the stored ones, i.e., to transform them until

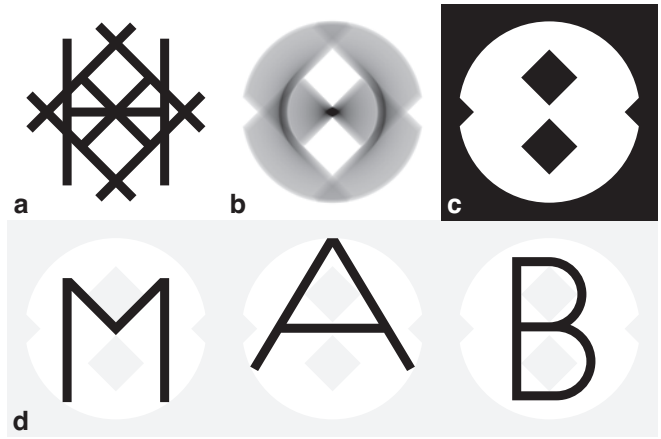

Fig. 2a–d. Pattern "H" superimposed in three angular positions (**a**), the "blurred" feature for ±45° rotation invariant recognition (**b**), its binary template (**c**), and a selection of three patterns that are supposed to be classified as "H" (**d**)

Fig. 3a–e. A selection of representatives for the pattern class "H"

[8] The demand for invariance represents the difference between code deciphering and real pattern recognition

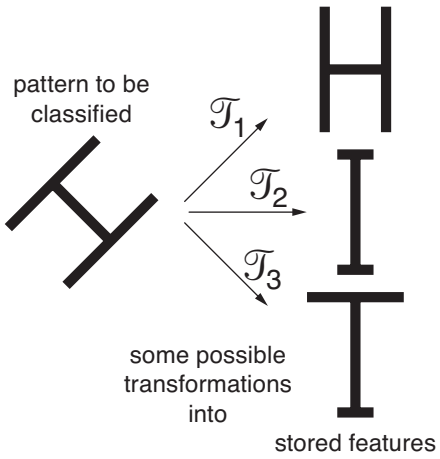

Fig. 4. Invariant recognition by a transformational approach: a pattern is made to fit stored features. Without restrictions this may lead to false classifications

the best match is reached, or vice versa (Bremermann 1971; Marko 1973). In order to maintain the quality of classification these transformations must be confined in their extent. However, the criteria for these restrictions are at least pattern dependent. Figure 4 depicts an example to characterize the problem.

4. The most convenient approach is the extraction of invariant features. Classification based on such features presents no principal difficulty, provided they contain sufficient information! The problem, though, is finding a proper method for their extraction. As with method 1, additional variance information is required.

Due to the fore-mentioned restrictions the first three methods are adequate for technical tasks and actually are successfully applied. With respect to the previously made claims, however, they appear highly inappropriate for the explanation of visual recognition processes. Furthermore, Foster and Mason (1979) showed, based on results from sophisticated psychophysical experiments, that method 3 is unlikely to apply to visual pattern recognition.

Classification, supposing learnt patterns (*re*-cognition), represents just one special performance of biological pattern processing. A more general and perhaps basic ability is that of understanding and describing unlearnt patterns (Narasimhan 1964 recognizes it as the real pattern recognition problem). This can be performed by analyzing their structural composition, and by intra- and inter-modal association. Thus, features of biological relevance should also be suited for these purposes. The examples in Figure 5 illustrate some consequences of this demand. The pattern in Figure 5a may be called "H-shaped", which means a context-free statement of analogy (intra-modal association). Obviously, such statements are extremely invariant. The fact that pattern description and understanding do not necessarily require a classifier (in the mechanistic sense of Fig. 1) leads to the conclusion that these abilities are based on invariant features! This means that invariant recognition, achieved by suitable classifiers (methods 1 to 3), cannot explain these invariant results from visual processing. In contrast, the pattern at the top of Figure 5b is recognized as the letter "H". This is a contextual statement (based on our knowledge of characters and writing) which is less abstract and thus less invariant.

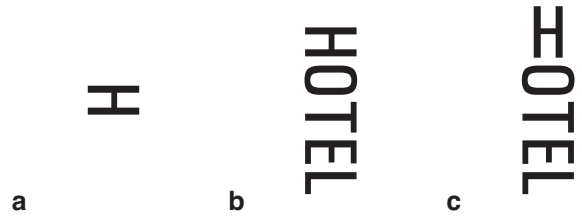

Fig. 5a–c. Pattern description by context-free analogies: "H-shaped" (**a**) and contextual classification: the symbol at the top belongs to the class of characters "H" (**b**). Degree of contextual influence: "H" or "H-shaped"? (**c**)

The degree of contextual influence can be estimated from the example in Figure 5c.

If we ask for the categories of human visual perception (e.g., in the sense of Kant's philosophy), we must suspect them of being invariant properties that are consistent with Gestalt theory. In order to formulate functional concepts for the explanation of visual pattern recognition processes, methods for their computation must be found. This, however, appears rather difficult: most of the published approaches deliver only partly invariant features (mostly shift invariant) or they lack biological plausibility. Two of the latter type shall be mentioned because they represent typical examples of the transfer of technical methods to biological systems without regarding the special needs and constraints of organisms. Casasent and Psaltis (1977) use – for technical purposes – a combination of Mellin and Fourier transformations to attain a shift, size and rotation invariant "pattern representation". Since then, several scientists proposed this method as a solution of the invariance problem in visual pattern recognition (Schwartz 1984 who presents an extensive survey; Reitboeck and Altmann 1984). A representation showing the same invariances can also be obtained by following a suggestion of Gerrissen (1982, 1984) that was investigated by Kröse (1985). It is based on a pattern's auto correlation function given in polar coordinates with a logarithmic radial coordinate. The auto correlation function of this intermediate representation has the desired properties. Besides the problem of their computation within the visual system (especially the mapping of the important foveal field

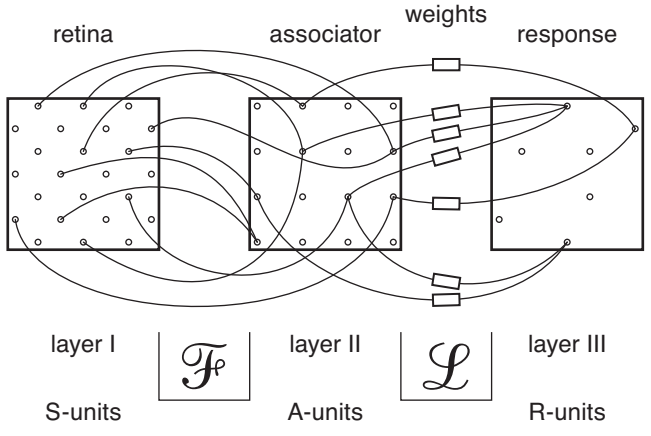

Fig. 6. Basic Perceptron structure

of view) it remains unclear why these representations are helpful for pattern descriptions in the sense of Gestalt psychology. Substantially, these features offer only one biologically relevant property: a high degree of invariance.

Perceptron-like concepts

Pattern recognition following the Perceptron approach (Rosenblatt 1962) corresponds well to the concept of a hierarchy of receptive fields, which is widely accepted in the visual sciences. It is essentially a kind of template matching, i.e., the stored features are the patterns themselves, or more precisely, parts of them. The disadvantages of such features were treated in the last section.

Perceptrons are usually described as systems of interconnected processing layers each containing a set of computing units (Fig. 6)[9]. The input pattern is presented to so-called sensory units (S-units) in layer I. This pattern representation is converted into an activity configuration in layer II according to a fixed rule $\mathcal{F}$ describing the connections between the S- and the A-units. The original Perceptron uses a random scheme $\mathcal{F}$. The A-units in layer II perform a threshold operation on the sum of their input signals (nonlinearity). These units become active if the threshold is exceeded. The signal transmission between layers II and III is similar to the one between layers I and II, except that rule $\mathcal{L}$ can be modified: each interconnection can be altered individually thus influencing the signal transmission to the R-units (weighting). For a given ("learnt") scheme $\mathcal{L}$ active R-units indicate decisions for particular pattern classes. The computing mechanism of these units is the same as that of the A-units. Figure 6 shows a basic three-layer Perceptron. There exist, however, more elaborate forms, consisting of several associator-layers that serve for a nonlinear combination of signals. This can result in features that allow better classification.

The input signal of an A-unit can be described as the sum of the products between each signal value (e.g., the intensity of a discrete image point) and the weight of its transmitting connection. Therefore, a set of weighted connections constitutes a spatial weighting function that acts as a template. The mathematical expression for such input signals is the cross correlation coefficient c_i. It is a measure for the match of a pattern $p(x,y)$ and a template $t_i(x,y)$. (The subscript i indicates the shape of the template.)

$$c_i = \iint_S p(x,y)\cdot t_i(x,y)\,\mathrm{d}x\mathrm{d}y \qquad (S\text{: area of layer I})$$

According to Figure 1 an active A-unit stands for the presence of one particular feature of a pattern, e.g., a line element at a certain location with a specific orientation. Now, in order to become independent of the particular location of such a feature detector, it appears logical to apply it at every position within the input layer. The resulting cross correlation function $c_i(\Delta x,\Delta y)$ contains all correlation coefficients. They are represented as a function of the relative translational positions Δx and Δy of the template type i, with respect to the origin of the pattern function.

$$c_i(\Delta x,\Delta y) = \iint_S p(x,y)\cdot t_i(x-\Delta x, y-\Delta y)\,\mathrm{d}x\mathrm{d}y$$

The corresponding processing structure is called homogeneous or space invariant. It allows a much better theoretical treatment of Perceptrons than the originally proposed random scheme $\mathcal{F}$, without introducing essential restrictions (Marko and Giebel 1970; Marko 1974; Platzer 1975; Fukushima and Miyake 1982). Although the processing properties do not vary within the layer, the location of the result still depends on the position of the stimulus. Therefore, the features are *not* shift invariant. As a consequence, any subsequent processing must also be performed by space invariant structures. This would not be necessary if invariant features were used.

The second stage of the Perceptron in Figure 6 can be understood as a classifier. During the phase of "learning", the weights of the network $\mathcal{L}$ are altered until each R-unit responds exclusively to patterns of one class. It is beyond the scope of this article to answer the questions whether it is possible to reach this state and which "learning strategies" are most promising for this purpose. Nevertheless, assuming a stable state is achieved, then the couplings between A- and R-units can, again, be interpreted as a set of templates.

It follows that the Perceptron concept permits only template matching operations. In spite of that the first stage, in which the features are determined and thus must be recognized as the bottleneck of the whole system, offers an often cited but somewhat trivial possibility of gaining invariant features (e.g., Fukushima and Miyake 1982): each type of template must be applied to the input in any desired geometrically transformed configuration: shifted (already introduced for the space invariant processing), rotated, scaled, reflected, etc. The resulting parallel systems constitute a transformation invariant processor. In order now to obtain invariant features, all correlation coefficients stemming from the same type of template must be combined in a nonlinear manner. In the case of space invariant processing, the corresponding shift invariant features are attained by spatially integrating the thresholded cross correlation functions. Obviously, these features are mere integral values, one for each set of equally shaped templates. They allow, at best, the counting of the corresponding shape elements contained in a pattern, but no statements about their positions with respect to each other. Such extreme disproportion between the effort (number of templates) and the information content of the features (it is not at all sufficient for the characterization of patterns) is rather unlikely for biological systems.

It can be concluded from the above discussion that there is no formation of useful invariant features within a Perceptron. Therefore, the postulated invariance of recognition must be achieved by other means. Independent of the invariance of the features, invariant classification can be attained in the second stage using either of the following mechanisms:

1. The network $\mathcal{L}$ permits the "learning" of blurred templates.

2. It is possible to "learn" separate templates for every representation of a pattern or, more exact, of its feature vector, provided there are sufficient interconnections.

[9] An excellent introduction to the function and structure of Perceptrons together with some simple examples demonstrating the performance of this concept is given by Singh (1966)

Both methods were treated in the previous section (methods 1 and 2) and were found to be unsuited for the explanation of visual recognition capabilities.

At this point a general remark on template-matching cross correlations should be made: because the correlation coefficients are evaluated by threshold operations, i.e., by decisions about the degree of match, it is indispensable to prevent a pattern of high intensity from surpassing the threshold of an inadequate template (intensity versus shape "cross-talk"). A common remedy against this effect is the normalization of the correlation coefficient to the integral pattern intensity (mathematically a division). Although it is not impossible that multiplications and divisions with variables could be carried out in neural circuits, it seems unlikely when considering the required speed and accuracy[10].

Neural pattern processing according to the Perceptron principle can be imagined as follows: the retinal pattern or its preprocessed retinotopic representation at the input layer of the cortex is linearly transmitted to the next layer by synaptic connections to "A-type neurons". The fixed weighting is achieved by the properties of these synaptic transmissions. The integration and threshold operations are performed by the neurons. Their output signals, i.e., their own activities, are subsequently processed in a corresponding manner, following the principle of convergence, until the "R-layer" is reached. It consists of "grandmother-type" neurons. These neurons, and those of possible layers in between, show some degree of plasticity in their synaptic input transmission. This variability for the purpose of learning is similar to the synaptic mechanisms proposed by Hebb (1949). It can be stated that, though the Perceptron structure corresponds well to the concept of a hierarchy of receptive fields (simple, complex, hypercomplex, etc.) the corresponding processing principle does little to explain the perceptive power of the visual system.

Pattern recognition and spatial frequency

After this rather cursory treatment it should be possible to understand why it is obvious to describe such Perceptron-like systems, at least partially, in terms of spatial frequencies: sinusoidal functions are eigenfunctions of linear space invariant systems. Their processing by such systems results merely in altered amplitudes and phases, i.e., they remain sinusoidal functions. The characterization of a linear *system* by its attenuation and phase lag, both as a function of spatial frequency, is sometimes advantageous (cf. Schade 1956). The expansion of pictorial *signals* into sinusoidal functions proves useful for certain technical tasks, e.g., texture analysis (Lendaris and Stanley 1970; Lukes 1977; Platzer and Glünder 1979; and others), image transmission and coding (e.g., Pratt 1979). It is rather obscure, however, why this mathematical method became such a standard functional concept in visual sciences. Specific linear space invariant systems, namely frequency filters or sets of these, can perform a spatial frequency analysis on signals. Speaking in terms of Perceptron systems, the appropriate templates for this purpose are bipolar sinusoidal gratings of different frequency, orientation, and phase (frequency analysis via correlation). The cross correlation coefficients, from extended gratings, are ideal measures for a signal description in the frequency domain (Fourier coefficients). Depending on the spatial extent of these templates (window), their retinal distribution, and the subsequently applied grouping mechanisms, either a global or local frequency description of a pattern can result (Fourier features). While the first is no longer considered seriously in visual sciences, the latter is quite popular in the fields of automatic image analysis (Platzer 1976; Jacobson and Wechsler 1982; Glünder and Bamler 1983) and vision research (Glezer and Cooperman 1977; Daugman 1984; and others). The computation of cross correlation functions using spatially restricted gratings is essentially the same as the frequency filtering of a pattern with filters of bandwidths (frequency channels) that are inversely proportional to the extent of the spatial window functions (for details see Gaskill 1978). Thus, local or global "frequency excerpts", i.e., bandpass filtered versions of a pattern can be calculated[11]. These may be useful for some sort of preprocessing (Marr 1982).

Frequency analysis – local or global – as a serious functional concept in vision is doubtful: neither the space invariance, to the required extent, of the processing (e.g., between neural layers) nor its linearity for high contrast patterns[12] are physiologically and psychophysically evident. Furthermore, it is not obvious why Fourier features should lead to better suited pattern descriptions than others. If they actually play an essential functional role in vision, their relevance, e.g., with respect to the findings of Gestalt psychology, and the subsequent processing mechanisms must be explained. In conclusion, a speculation on the reasons that led to the popularity of Fourier features is presented: if a system is linear, or can be approximated as linear, it can be described by its transmission properties of sinusoidal functions (see above). Since the visual system was suspected to be linear it was tested with – mostly – sinusoidal gratings. A prominent finding was that the system is quite sensitive to the specific periodicities of such patterns. It was then argued: the system is able to distinguish patterns of different periodic structure thus it performs frequency analysis. This conclusion, however, is not valid because the discrimination of

[10] Neural signals are coded as impulse rates (Poisson process) and thus are quite noisy. It follows, that for a certain computational accuracy (typically a few percent) the required integration time depends on the impulse rate which is, however, limited to less than 1000 imp/s. The processing time per elementary operation, e.g., normalization, is presumably not much longer than 10 ms. Therefore, logarithmic signal conversions and especially back-conversions for the purpose of multiplying, would require a lot of parallel processors for the computation of each product, if the signal to noise ratio and the signal dynamic is to be maintained. In order to reduce the effort and to avoid conversion losses, other operations must be considered (e.g., Barnea and Silverman 1972).

[11] Gaussian windows are plausible from a signal theoretical point of view. The corresponding feature detectors – called 2D Gabor functions – resemble certain receptive field profiles. These profiles correspond well to those "Gabor patterns" that yield best detection in psychophysical experiments (cf. Caelli and Moraglia 1985). They consist, however, of only two or three cycles per Gaussian halfwidth. Thus, it can be suspected that their *spectral* relevance for the *function* of the visual processing is not very high.

[12] Most psychophysical investigations are performed using low contrast patterns (mostly gratings) for which nearly every nonlinear system appears to be linear

pattern periodicities can also be achieved by other techniques. An example is introduced in the next section.

Relational concepts

"The information for the constant dimension of an object is normally carried by invariant relations in an optic array"
J. J. Gibson (1966)

This quotation typifies a common opinion among well known Gestaltists and expresses the underlying principle of the relational concept[13]. Descriptions of inner pattern relations can hardly be obtained by comparing a pattern with fixed templates but rather by comparing it, or parts of it, with itself. A mathematical operation that performs this is the auto correlation function $c_a(\Delta x, \Delta y)$.

$$c_a(\Delta x, \Delta y) = \iint_S p(x,y) \cdot p(x - \Delta x, y - \Delta y)\, \mathrm{d}x\mathrm{d}y$$

This function characterizes a pattern by its inner coherence with respect to rigid translations (shift congruence). It indicates pattern periodicities in a more general way than Fourier descriptions do. Figure 7 illustrates this statement. It shows two patterns, each composed of three identical squares that are diagonally lined up. The pattern in Figure 7a leads to a distinct percept of a threefold discontinuity whereas the pattern in Figure 7b appears significantly more continuous. (This effect is best observed for foveal presentation!) The corresponding auto correlation functions (thresholded) clearly reflect these properties (Figs. 7c, d). For convenience, the central parts of the Fourier power spectra of the original patterns are displayed in the Figure 7e, f. Although the power spectra contain the same information as the (unthresholded) auto correlation functions, they do not represent the mentioned properties as clearly. Rather, they remain hidden in the spectral fine structure while the information about the orientation of the squares dominates these transforms. Both pattern representations are shift invariant: because the auto correlation function depends on the *relative* translations Δx and Δy it is independent of the pattern's absolute position. The power spectrum (loss of phase) is the Fourier transform of the auto correlation function and shift invariant as well.

However important information about shift symmetry may be for the description of a pattern, this alone does not suffice. Thus, equally significant properties such as angle and size relations, and axial symmetries must also be considered as features. They can be evaluated analogously to the auto correlation function by the computation of "generalized auto comparison functions".

Figure 8 shows a simplified diagram of a relational processor that extracts such features. It was introduced by members of the image sciences and pattern recognition group at the Institut für Nachrichtentechnik (Glünder et al. 1984). In order to increase the significance of these correlation results, line-like input patterns, not necessarily binary ones, are presumed (see Lowenthal and Belvaux 1967). They can be provided by a suitable preprocessing stage. The transformation unit $\mathcal{T}$ produces geometrically transformed versions of the input pattern which are compared with the original, untransformed pattern. Contrary to the Perceptron and its fixed set of templates, here fixed transformation rules are applied. Of primary interest are the following geometric transformations: rigid translations, rotations, changes in scale, reflections and combinations of all four. (Temporal changes of a pattern, such as motion, deformation, etc. can be revealed by its comparison with delayed versions.) The comparator circuit C is defined by its mathematical operation: the product that is used in the case of classical correlation, the squared difference, and the absolute value of the difference can serve for this purpose. The latter is favoured because of its easy "neural implementation" requiring only elementary properties of neurons: excitation, inhibition, and summation (Rashevsky 1960; Glünder et al. 1984). The spatially integrated comparison results (comparison coefficient $d_{\mathcal{T}}$) can now be represented as functions of their transformation variables $d(\ldots)$. They constitute relational feature functions that indicate inner pattern relations.

$$d_{\mathcal{T}} = \iint_S \left| p(x,y) - \mathcal{T}\{p(x,y)\} \right| \mathrm{d}x\mathrm{d}y$$

Compared to the product, the difference measure has the additional advantage that the best match is indicated by a zero output independent from the pattern intensity. No match results in a coefficient of twice the spatial integral value of the input pattern. Thus, there is, in general, no need for normalization.

Next, the invariance properties of relational feature functions shall be investigated. Their evaluation, as measures of similarity between the input pattern and its transformed versions, results in a general invariance property: a relational feature function is always invariant with respect to its underlying transformation. Unfortunately, the angle and size features, $d(\varphi, x_f, y_f)$ and $d(m, x_f, y_f)$, are bound to the fixed points (x_f, y_f) of their transformations (centres of rotation or zoom) and thus they are shift variant. This dependency can be eliminated by considering only those fixed points that result in best matches for each value of the transformation parameters φ (angle)[14] or m (scale factor) respectively. Both feature functions $\hat{d}(\varphi)$ and $\hat{d}(m)$ are now shift, rotation, scale and reflection invariant. They express rotational symmetries and geometric similarities respectively. It is noteworthy that the number and positions – absolute and relative – of the optimal fixed points are powerful features as well. In order to achieve useful results from scale transformations the *locally* energy-preserving scaling must be applied. The effect of the scale dependent integral intensity on the comparison coefficient can be easily compensated for. The variance problem, associated with the calculation of the axial symmetry feature, can be solved similarly: the same degree of invariance can be reached if the position and orientation of the reflection axes are eliminated.

[13] Another statement of this kind is found in the introduction to Rosenblatt's book but without any conceptual consequence for the subsequent 565 pages

[14] A similar procedure is used by Hsu et al. (1982) to find a suitable centre for the expansion of a pattern into circular harmonic functions

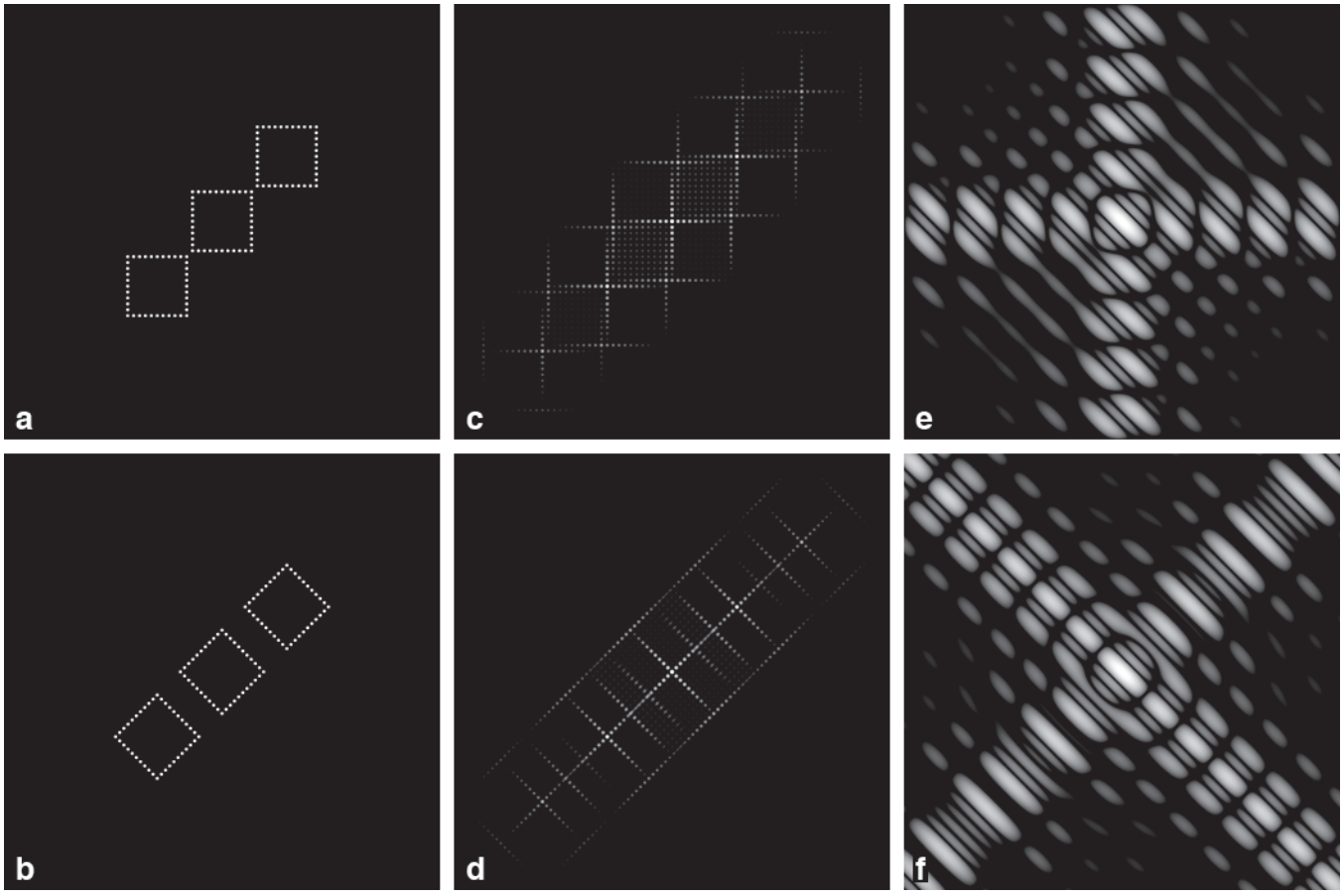

Fig. 7a-f. Discontinuity (**a**) and continuity (**b**) in perception that are clearly reflected by the auto correlation functions (**c, d**) but are hidden in the fine structure of the power spectra (**e, f**)

Compared to one-dimensional feature functions, combination of these offer far more descriptive power, without restricting the invariance. The feature $\hat{d}(\varphi,m)$, for example, expresses size relationships between pattern elements that differ in their orientations and vice versa.

Finally the peculiarity of the translational feature function $d(\Delta x,\Delta y)$ is pointed out. This seemingly substantial feature (Fig. 7) turns out to be merely shift invariant. This is a consequence of its reference to a coordinate system (orientation variance) and to the graduation of this system (size variance). Suggestions exist for derived one-dimensional features that are either scale or rotation invariant (e.g., the chord functions introduced by Moore and Parker 1974). A desirable feature, however, that indicates the translational distance between similar pattern elements (rotation invariance) with respect to their dimensions (size invariance) could not yet be found. Results from psychophysical experiments with textures (Julesz and Bergen 1983) revealed that under certain conditions the translational relationship between pattern elements is in fact ignored. One of these conditions is the rotationally randomized presentation of the texture elements which demands for the full rotational invariance of the observer. Although texture perception may be based on totally different mechanisms, these findings give rise to the conjecture that a translational feature, perhaps due to its restricted invariance property, is not involved in this task.

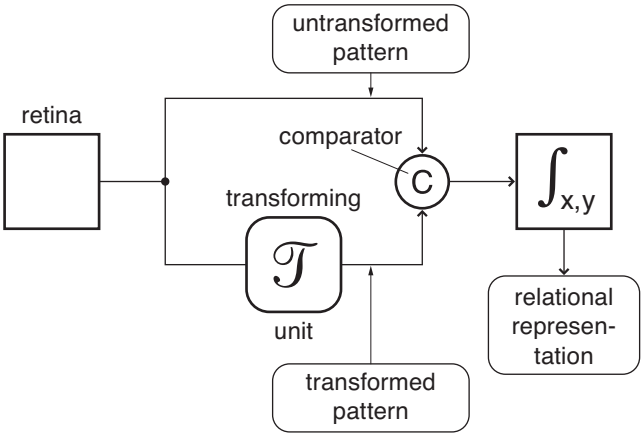

Fig. 8. Scheme of a relational pattern recognition concept

The simple pattern examples in Figure 9 provide an impression of the nature and meaning of these relational features. The patterns depicted in Figures 9a and 9b differ in their scale features: using the above mentioned compensation for the scale dependent intensity, the former pattern will result in $\hat{d}(m) = 0$, whereas the latter is zero only for $m = 1$ and increases for smaller and larger values. Thus, one aspect in which these patterns differ is revealed: the joining and the intersection of line elements. The same holds true for the patterns in Figures 9c and 9d. These, however, have also

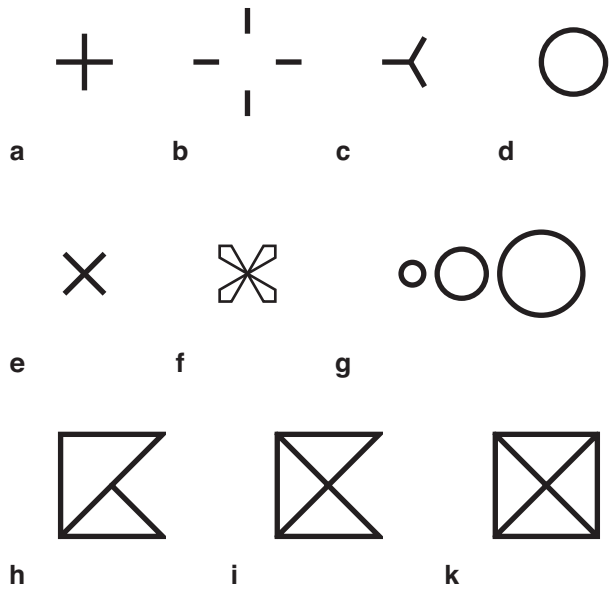

Fig. 9a–k. Examples of perceptual similarities and dissimilarities that are partly reflected by relational features

a different rotational symmetry, indicated by $\hat{d}(\varphi)$: while the circle leads to $\hat{d}(\varphi) = 0$, the pattern in Figure 9c is characterized by zeros for multiples of 120° with local minima for multiples of 60°. Obviously the patterns in Figures 9a, b have angle features with zeros for multiples of 90°. Hence, a difference from the pattern of Figure 9c is indicated, and another is given by the number of symmetry axes: three for the latter and four for the former ones. The pattern in Figure 9e delivers the same features as the cross of Figure 9a but differs from the one shown in Figure 9f in its rotational and scale feature.

Relational features can even reveal rules of composition for more complex patterns, e.g., of those shown in Figures 9g and 9h. The former is nicely described by the minima of $\hat{d}(m)$ that indicate the relative sizes of the three circles (their relative positions are contained in the fixed point feature). A dominant aspect of the latter is indicated by the combined feature $\hat{d}(\varphi, m)$ that becomes a minimum for $\varphi = 45°$ and $m = \sqrt{2}$. This feature function also permits the estimation of linelength-ratios, e.g., of the sides of a triangle. The axial symmetry feature provides the number of symmetry axes, e.g., one for the pattern of Figure 9i and four for the one in Figure 9k. It also delivers a measure indicating the highest degree of axial symmetry of a pattern that is not really axisymmetric, such as the one in Figure 9h.

The quantitative element of these features is quite useful, but what is surprising is their relevance with respect to Gestalt psychology: certain aspects of a pattern's "figural goodness" or "Prägnanz" are reflected by the symmetry and fixed point features. Furthermore, the examples just discussed show that the characterization of a pattern is, to a great extent, possible without using the shift feature. Apart from their utility for pattern descriptions, invariant relational features are also well suited for classification and association purposes. Since they are functions and not single values, higher relational features can be extracted, i.e., relations of relations.

It is quite feasible locally to extract relational features by applying many spatially restricted comparisons to a pattern. The resulting local relational features can again be compared on a higher processing level, thus forming a relational hierarchy (cf. Platzer 1976; Platzer and Glünder 1979). Related to this is an approach suggested by Palmer (1982, 1983). Although he gives no detailed mathematical description, it can be formulated as follows: cross correlations with a variety of template-like windows of different shapes are locally performed all over the input plane. The transformations are implicitly realized by spatially comparing the results of these local correlations. Contrary to the Perceptron, where the correlation coefficients of a set of *equally shaped templates* at different retinal locations are combined (integrating system), here all coefficients are compared with each other, independent of their belonging to such a set (differentiating system). Thus, by knowing the shape and position of each template, as well as all the results from these comparisons, it is possible to compute invariant features that are quite similar or even equal to those gained by explicit transformations. The choice of the templates, however, is crucial since they limit the performance of the following processing. For the purpose of shift invariance the locations of the templates must be eliminated. One way this can be done is by spatially integrating all those comparison results that have *equal relational properties*. Thus, this approach is also based on the principle defined by the question: which geometric transformations convert parts of a pattern into others? From a systems theoretical point of view, it is therefore essentially equivalent to the relational approach described before.

These results may lead to the following conjecture: in non-trivial recognition systems, i.e., systems in which the invariance of features is not traded for a significant loss of their information content, invariance can be achieved with respect to those transformations that are implemented explicitly or implicitly (Palmer) in the processing system.

In the final section comments are made concerning the biological plausibility of relational concepts. The geometrical transformations can be computed in parallel by using suitable interconnection schemes between "neural layers". With such point-to-point "wiring" the energy condition for the scale transformation is directly met.

A retinotopic pattern representation in a neural layer is geometrically transformed in parallel (with respect to all required qualities and quantities) by being transmitted to further layers in which they are compared with the untransformed representation. The number of long-reaching interconnections, is reduced if local operations are assumed. Palmer's approach requires that the transforming interconnection scheme be replaced by a "wiring" that compares the results of all template correlations with each other. In that case, the preceding structure resembles the first stage of a Perceptron which is assumed to be at least orientationally selective. This fact in particular corresponds well with the common neuroanatomical and neurophysiological findings (cortical columns). Relational structures afford large numbers of interconnections per computing unit and they provide a rapid computation due to the small number of layers that are involved in feature extraction.

As mentioned earlier, variance information is indispensable for biological recognition systems. Thus, the high invariance of such relational features must be confined according to a given context. For example, the rotation invariance must be limited to distinguish between the letters "d" and "p" and the numbers "6" and "9", when presented in the context of "writing". Information about the absolute position of a pattern, for example, is given by the optimal fixed point coordinates. In a hierarchically organized system the specific presentation of a pattern can be easily determined, since the localization increases as the processing level decreases.

For both relational approaches the amount of computation is high but still economical, since the processing structure is, contrary to the Perceptron, not specific for certain pattern classes: while, simply speaking, the number of templates for a Perceptron is given by the number of pattern classes times the average number of tolerable pattern variances per class, the latter alone defines the required number of transformation rules for the relational concept. As mentioned, a hierarchical evaluation of relational features can be achieved by the repeated application of equal or similar processing structures, e.g., comparators. Hence, features of a lower processing level can be recognized as "patterns" for the next level. This hierarchy is significantly different from that of Perceptrons: there, the meaning of a feature becomes increasingly specific according to the processing level – finally resulting in statements commonly associated with "grandmother neurons" – here, however, the degree of abstraction and invariance of the features increases.

The conjecture that the visual system uses information about inner relations of a pictorial representation in order to become invariant against certain changes of this representation is not a new one. The "retinex" theory of colour vision proposed by Land (1959; 1983) is a prominent example: independence of colour changes due to illumination effects is achieved by defining an object's colour through its relations to the colours of surrounding objects. Colour constancy is thereby explained, and one can suspect relational shape processing to be a key to other constancy phenomena of visual perception.

In this paper it was attempted to explain the important role that invariant, Gestalt-related features play in human vision. It was shown that, consequently, template-matching concepts are of little value for the explanation of visual perception. Furthermore, mathematically elegant approaches to invariant recognition were criticized because of their inferior biological plausibility. An alternative concept was introduced that avoids some of these shortcomings: the extracted features express neither abstract mathematical quantities, e.g., Fourier coefficients, nor pure numerical facts, e.g., numbers of pattern elements, but qualitative properties such as symmetries and similarities. Their evaluation requires spatial comparisons that can be explained by elementary neural processes. Contrary to related approaches, it was possible mathematically to formulate the feature extraction process. Thus, it can be tested and perhaps falsified.

Acknowledgements. I thank first of all Helmut Platzer, to whom I am greatly indebted. Thanks also to Joe Hofer-Alfeis, who made the first step toward relational processing in our group many years ago; and to Alex Gerhard, who helped in corroborating the actual approach with sophisticated computer simulations. The many valuable suggestions of R. Bamler and Prof. T. Caelli during the preparation of the paper are also highly appreciated. The investigations were partially supported by the Deutsche Forschungsgemeinschaft under the grant SFB 50/A33.

References

Andrews HC (1972) Introduction to mathematical techniques in pattern recognition. Wiley-Interscience, New York

Barnea DI, Silverman HF (1972) A class of algorithms for fast digital image registration. IEEE Trans C-21:179–186

Barrow HG, Popplestone RJ (1971) Relational descriptions in picture processing. In: Meltzer B, Michie D (eds) Machine intelligence 6. Edinburgh University Press, Edinburgh, pp 377–396

Beck J (ed) (1982) Organization and representation in perception. Erlbaum, Hillsdale/NJ

Beck J, Hope B, Rosenfeld A (eds) (1983) Human and machine vision. Academic Press, New York

Braddick OJ, Sleigh AC (eds) (1983) Physical and biological processing of images. Springer, Berlin

Bremermann HJ (1971) What mathematics can and cannot do for pattern recognition. In: Grüsser OJ, Klinke R (eds) Pattern recognition in biological and technical systems. Springer, Berlin, pp 31–45

Caelli T, Moraglia G (1985) On the detection of Gabor signals and discrimination of Gabor textures. Vision Res 25:671–684

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol (Lond) 197:551–566

Casasent D, Psaltis D (1977) New optical transforms for pattern recognition. Proc IEEE 65:77–84

Daugman JG (1984) Spatial visual channels in the Fourier plane. Vision Res 24:891–910

Foster DH, Mason RJ (1979) Transformation and relational-structure schemes for visual pattern recognition. Biol Cybern 32:85–93

Fu KS (1968) Sequential methods in pattern recognition and machine learning. Academic Press, New York

Fukunaga K (1972) Introduction to statistical pattern recognition. Academic Press, New York

Fukushima K, Miyake S (1982) Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. Patt Recogn 15:455–469

Gaskill JD (1978) Linear systems, Fourier transforms, and optics. Wiley, New York

Gerrissen JF (1982) Theory and model of human global analysis of visual structure. IEEE Trans SMC-12:805–817

Gerrissen JF (1984) Theory and model of human global analysis of visual structure; part II: The space-time and visual value segment. IEEE Trans SMC-14:847–862

Gibson JJ (1966) The senses considered as perceptual systems. Houghton Mifflin, Boston/MA

Glezer VD, Cooperman AM (1977) Local spectral analysis in the visual cortex. Biol Cybern 28:101–108

Glünder H, Bamler R (1983) Coherent optically performed operations in a 4D compound space/spatial-frequency domain with applications in image analysis and quality control. In: Oosterlinck A, Tescher AG (eds) Applications of digital image processing. SPIE Press, Bellingham/WA (SPIE Proc 397), pp 207–215

Glünder H, Gerhard A, Platzer H, Hofer-Alfeis J (1984) A geometrical-transformation-invariant pattern recognition concept incorporating elementary properties of neural circuits. In: Wein M (ed) Proceedigs of the 7th International Conference on Pattern Recognition. IEEE Comp Soc Press, Silver Spring/MD, pp 1376–1379

Hartline HK (1940) The receptive fields of optic nerve fibers. Am J Physiol 130:690–699

Hebb DO (1949) The organization of behaviour. Wiley, New York

Hofer J, Platzer H (1978) Klassifikation in Modellsystemen. In: Hauske G, Butenandt E (eds) Cybernetics 1977. Oldenbourg, München, pp 426–430

Hoffman WC (1966) The Lie algebra of visual perception. J Math Psychol 3:65–98

Hsu YN, Arsenault HH, April G (1982) Rotation-invariant digital pattern recognition using circular harmonic expansion. Appl Opt 21:4012–4015

Hubel DH, Wiesel TN (1959) Receptive fields of single neurons in the cat's striate cortex. J Physiol (Lond) 148:574–591
Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (Lond) 160:106–154
Jacobson L, Wechsler H (1982) A paradigm for invariant object recognition of brightness, optical flow and binocular disparity images. Patt Recogn Lett 1:61–68
Julesz B, Bergen JR (1983) Textons, the fundamental elements in preattentive vision and perception of textures. Bell Syst Techn J 62:1619–1645
Koffka K (1935) Principles of Gestalt psychology. Harcourt Brace & World, New York
Kröse BJA (1985) A structure description of visual information. Patt Recogn Lett 3:41–50
Kuffler SW (1953) Discharge patterns and functional organization of mammalian retina. J Neurophysiol 16:37–68
Kuhn TS ([2]1970, [1]1962) The structure of scientific revolutions. University of Chicago Press, Chicago/IL
Land EH (1959) Color vision and the natural image: Part I and II. Proc Natl Acad Sci USA 45:115–129, 636–644
Land EH (1983) Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. Proc Natl Acad Sci USA 80:5163–5169
Leibowitz HW, Post RB (1982) The two modes of processing concept and some implications. In: Beck J (ed) Organization and representation in perception. Erlbaum, Hillsdale/NJ, pp 343–363
Lendaris GG, Stanley GL (1970) Diffraction-pattern sampling for automatic pattern recognition. Proc IEEE 58:198–216
Livingstone MS, Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. J Neurosci 4:309–356
Lowenthal S, Belvaux Y (1967) Progrès récents en optique cohérente: filtrage des fréquence spatiales, holographie. Revue d'Optique 46:1–64
Lukes GE (1977) Rapid screening of aerial photography by OPS analysis. In: Leighty RD (ed) Data extraction and classification from film. SPIE Press, Bellingham/WA (SPIE Proc 117), pp 89-97
Marko H, Giebel H (1970) Recognition of handwritten characters with a system of homogeneous layers. Nachrichtentechnische Zeitschrift 23:455–459
Marko H (1973) Space distortion and decomposition theory: a new approach to pattern recognition in vision. Kybernetik 13:132-143
Marko H (1974) A biological approach to pattern recognition. IEEE Trans SMC-4:34–39
Marr D (1982) Vision – A computational investigation into the human representation and processing of visual information. Freeman, New York
Metzger W ([3]1975, [1]1936) Gesetze des Sehens. Kramer, Frankfurt/M
Minsky ML, Papert SA (1969) Perceptrons – An introduction to computational geometry. MIT Press, Cambridge/MA
Moore DJH, Parker DJ (1974) Analysis of global pattern features. Patt Recogn 6:149–164
Narasimhan R (1964) Labeling schemata and syntactic descriptions of pictures. Information and Control 7:151–179
Onoe M, Preston K, Rosenfeld A (eds) (1981) Real-time/parallel computing – Image analysis. Plenum, New York
Palmer SE (1982) Symmetry, transformation, and the structure of perceptual systems. In: Beck J (ed) Organization and representation in perception. Erlbaum, Hillsdale/NJ, pp 95–144
Palmer SE (1983) The psychology of perceptual organization: a transformational approach. In: Beck J, Hope B, Rosenfeld A (eds) Human and machine vision. Academic Press, New York, pp 269–339
Platzer H (1975) Holographic modelling of parallel information processing in the visual pathways of mammals. In: Greguss P (ed) Holography in medicine. IPC Science and Technology Press, Richmond/Surrey, pp 124–127
Platzer H (1976) Pattern processing and texture analysis by coherent optical computer. In: Rosenfeld A (ed) Proceedings of the 3rd International Joint Conference on Pattern Recognition. IEEE Comp Soc Press, Silver Spring/MD, pp 713–715
Platzer H, Glünder H (1979) Generation and use of local power spectra by coherent optics. In: Gardner WE (ed) Machine-aided image analysis, 1978. The Institute of Physics, Bristol (Conference Series no. 44), pp 114–119
Popper KR ([6]1972, [1]1959) The logic of scientific discovery. Hutchison, London
Pratt WK (ed) (1979) Image transmission techniques. Academic Press, New York
Rashevsky N ([3]1960, [1]1938) Mathematical biophysics – Physico-mathematical foundations of biology II. Dover Publications, New York.
Reitboeck HJ, Altmann J (1984) A model for size- and rotation-invariant pattern processing in the visual system. Biol Cybern 51:113–121
Rosenblatt F (1957) The Perceptron, a perceiving and recognizing automaton (Project PARA). Cornell Aeronautical Laboratory Report No 85-460-1, Ithaca/NY
Rosenblatt F (1962) Principles of neurodynamics – Perceptrons and the theory of brain mechanisms. Spartan Books, Washington/DC
Schade OH (1956) Optical and photoelectric analog of the eye. J Optical Soc Am 46:721–739
Schwartz EL (1984) Anatomical and physiological correlates of visual computation from striate to infero-temporal cortex. IEEE Trans SMC-14:257–271
Sebestyen GS (1962) Decision-making processes in pattern recognition. Macmillan, New York
Shapiro LG, Haralick RM (1982) Organization of relational models for scene analysis. IEEE Trans PAMI-4:595–602
Singh J (1966) Great ideas in information theory language and cybernetics. Dover Publications, New York
Taube M (1961) Computers and common sense – The myth of thinking machines. Columbia University Press, New York
Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt, II. Psychol Forsch 4:301–350