Carefully reworded reprint of a contribution to:
Proceedings of the 7th International Conference on Pattern Recognition. Wein M. (ed.) IEEE Computer Society Press, Washington/DC, 1984

A GEOMETRICAL-TRANSFORMATION-INVARIANT PATTERN RECOGNITION CONCEPT INCORPORATING ELEMENTARY PROPERTIES OF NEURONAL CIRCUITS

H. Glünder, A. Gerhard, H. Platzer, J. Hofer-Alfeis

Institut für Nachrichtentechnik, Technische Universität München
Arcisstraße 21, D-8000 München 2, FRG

Abstract

We introduce a rather general representation of pictorial data that allows one to derive features which are invariant under geometrical transformations. These features are based on intrinsic measures of given signals. Beside these absolute inner geometrical relations, it delivers information about a signal's situation with respect to external coordinates.

The processing concept was developed for parallel computing structures and we try to give some design hints for suitable parallel computing mechanisms. This allows us, to a certain degree, to compare its capabilities with those demonstrated by biological pattern processing systems, e.g. by the human visual system. Thus, our approach is less suited for implementations on "von Neumann"-type computers.

To illustrate our concept, we show results obtained from an early coherent-optical simulation and from a more extensive experiment performed on a digital general purpose computer.

Introduction

Considering principles of true pattern recognition, we mean invariant pattern recognition, i.e. the correct classification of a pattern independent from its presentation to the input device of the processing system, not some kind of code deciphering. In many cases, we are already satisfied if we can achieve a classification which is invariant under variations of luminance, contrast, shift, scale, and orientation of signals. What we need for that are features remaining unchanged under those transformations and allowing convenient learning and classifying. Concerning the geometrical transformations, we think it is advantageous to be invariant to their full extent, i.e. within the limits of the sensory plane and not only for small deviations from a prototype representation. For many tasks, however, this claim will force us to get additional information about the position of the pattern relative to fixed coordinates as to distinguish, for example the patterns "6" and "9", or "g" and "9" in a given context.

Such an approach can simplify complicated classification tasks and may also lead to better interpretations of human pattern recognition capabilities. Let us assume, e.g., a system which cannot yet identify the symbol "M" but already classifies the letter "W" correctly. Wouldn't it be better to state: pattern "M" looks like a "W" turned by 180 degrees, instead to "learn" a completely new set of characteristic features?

We should like to emphasize that invariances cannot be achieved merely by so-called invariant processing. Space invariant systems, e.g., show equal processing properties independent of the signal location at the input but the output location is that of the input. In many cases, we might still get the desired features by applying spatial integration to the output signal[1], however, at the expense of the pattern separability. Many partially invariant recognition systems, e.g. of the Perceptron-type, suffer from this problem[2]. Instead of template matching (cross correlation) or related operations, we suggest operations that compare the signal with itself (auto comparison). But these approaches imply the loss of positional information that we can overcome by using local operations[3,4].

We developed such "relational concepts" following investigations of Perceptron-like structures which turned out to be rather limited in their explanation of processing capabilities shown by the human visual system[5,6,7]. Our goal was to generate a signal representation which allows statements about the position of a pattern with respect to the sensory unit as well as the straightforward extraction of transformation invariant features, thereby applying techniques showing elementary neuronal relevance.

Boundary Conditions

Since our concept should show some biological plausibility, we use the following guidelines for our approach:

1. The processing shall be fitted to the pictorial aspect of the signals. Hence, we are to prefer "manipulations of pictorial content rather than pixel states" as pointed out by Sternberg[8].
2. The highest possible degree of parallel computation shall be incorporated.
3. The high amount of interconnections per elementary processing unit of about 10^4 which is found in cortical tissue shall be taken into account.
4. Multiplications as well as divisions, e.g. for normalization etc., shall be avoided (except for constant coefficients).

Today it can be regarded as a fact that cortical processing of patterns is done on some kind of bandpass filtered or skeletal version of the input signal (see the references in Marr's book[9]). There exist already useful theories about possible mechanisms for this purpose so that we can assume such preprocessed patterns as input for our system[5].

CH2046-1/84/0000/1376$01.00©1984 IEEE

General Relational Signal Representation

To obtain intrinsic measures of a pattern, we propose to compare it, or parts of it, with itself. In order to become independent of the individual geometrical presentation of a pattern, we compare it with a set of geometrically transformed versions of itself that are derived from the input representation by a fixed processing scheme. Figure 1 illustrates this approach.

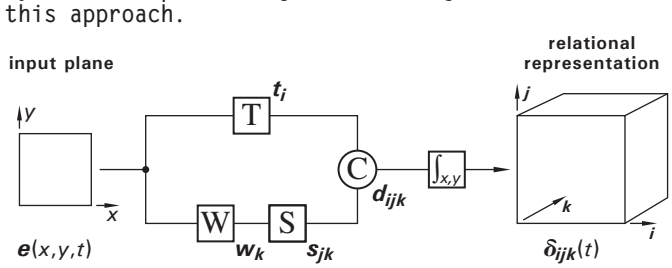

Fig. 1 Basic relational processing concept

On the one hand, an input signal $e(x,y,t)$ is fed to the transforming system $\mathbf{T}$ which produces rotated and zoomed versions t_i with respect to an arbitrarily chosen point. This can be done in parallel by fixed connections reaching the comparison stage $\mathbf{C}$. On the other hand, it is windowed by a set $\mathbf{W}$ of fixed and highly overlapping window functions that may differ in size, shape, and profile and that are arbitrarily distributed over the input plane. The decomposition of a signal into parts w_k can again be achieved in parallel using selective connections and weighting. Because we want to become independent of a fixed center of rotation and zoom, and wish to introduce translations as well, we need a separate unit for shifts in the x- and y-directions which follows circuit $\mathbf{T}$ or, as assumed here, acts on each of the windowed details w_k. The shifted versions s_{jk} can be derived in parallel by the appropriate "wiring" to the second input of stage $\mathbf{C}$. Consequently, the output of this comparator circuit has an extent of $(k \cdot j \cdot i)$-times the size of the input plane.

The comparison is performed point-wise in parallel between all pattern combinations t_i and s_{jk}. We propose the absolute value of the difference expressed by $d_{ijk} = |t_i - s_{jk}|$. (Another possible measure is the squared difference.) Therefore, we obtain results which are restricted to values between zero and the maximum signal amplitude. A perfect match is indicated by a zero output, independent of the absolute signal strength, contrary to multiplying comparators which need normalization in order to allow an interpretation of the results.

The desired relational representation $\mathbf{\Delta} = \{\delta_{ijk}\}$ is achieved by spatial integration of the difference images d_{ijk}:

$$\delta_{ijk}(t) = \iint d_{ijk}(x,y,t)\ \mathrm{d}x\ \mathrm{d}y$$

If we assume, for better understanding, a multiplying comparator $\mathbf{C}$, then each output δ_j is the correlation function of the signals t_i and w_k.

Figure 2 shows how a neuronal circuit of a difference comparator for spatially sampled patterns may look like in principle. Assuming spatially discrete processing, we can give some estimates for the optimum bandwidth of the above mentioned differentiating pre-processor: The line width of its output should be about twice the shift increment of unit $\mathbf{S}$.

Let us now summarize the principal properties of the properties of our relational representation:

1 An input pattern is described by its angular and positional congruences. These measures are represented as characteristic "activity configuration".
2 A different geometrical presentation of the same pattern yields this configuration at a different location in $\mathbf{\Delta}$. Though it may be compressed or expanded, its characteristic property, expressed by the sequence of activity, is not changed.

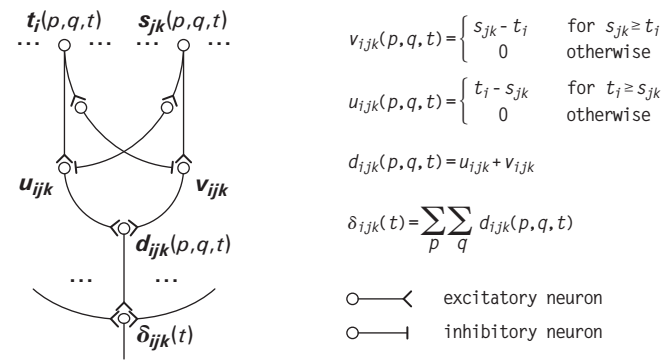

Fig. 2 Model of a neuronal comparator circuit

Extraction of Invariant Features

The relational representation $\mathbf{\Delta}$ contains a tremendous amount of data and, as a consequence of the fixed set of windows, it is no at all transformation invariant. Therefore, we must consider data reduction and methods for becoming independent of the windows. First, we apply a threshold operation Θ which serves the suppression of components δ_{ijk} that exceed the spatial integral value τ_i of the transformed signal as shown in Fig.3. The output of this stage computes as

$$\alpha_{ijk} = \begin{cases} \tau_i - \delta_{ijk} & \text{for } \delta_{ijk} < \tau_i \\ 0 & \text{otherwise} \end{cases}$$

thus leading to a more pronounced "activity configuration" $\mathbf{A} = \{\alpha_{ijk}\}$.

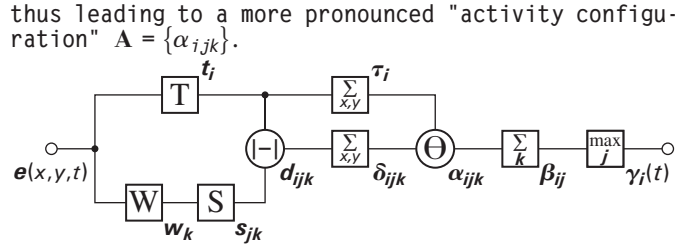

Fig. 3 Block-diagram of the proposed concept

A direct way of getting rid of the fixed window scheme is to omit this unit and to apply global comparisons. Here, however, we use another simple method based on our universal representation $\mathbf{A}$: we project array $\mathbf{A}$ in the k-direction, i.e. we introduce the summation over all windows. Compared to a global concept, we achieve two principal advantages:

1 The positional information is saved from the very beginning, thus we can use it in conjunction with the characteristic configurations, e.g. for motion analysis[10].
2 We obtain a higher significance of the resulting signal representation by thresholding which is not applicable in case of global comparison techniques.

The resulting array $\mathbf{B} = \{\beta_{ij}\}$ is now a shift invariant representation. It is, however, variant under

rotation and zoom since it still depends on the parameter j that determines the centers of these transformations. For the elimination of this variable, we propose maximum selection which becomes plausible when considering the correlative nature of the preceding computation.

$$\gamma_i = \max_j \left[\beta_{ij}\right]$$

The resulting invariant feature vector $\mathbf{\Gamma} = \{\gamma_i\}$ expresses inner angle relations and shape similarities of the input pattern.

Experimental Results

Figures 4b,c show sections through the representation $\overline{\mathbf{\Delta}}$ of the letter "B" in upright position corresponding to the window function that selects the vertical bar (Fig.4b) and the one selecting the upper half of this symbol (Fig.4c). The results were computed by coherent-optical correlation filtering[6]. Figure 4a shows the computer plotted point response of the simulation system. We implemented 200 geometrical transformations: 20 rotation angles φ with an increment of 18 degrees and 10 scale factors m from the interval 0.5 to 2.0. Shifts are continuous since correlation was applied. The untransformed input pattern in Fig.4a is defined by $\varphi = 0$ and $m = 1$. The corresponding outputs are indicated by arrows.

For the digital simulations, we used an input array of 64 x 64 pixels. We restricted ourselves to 24 rotations with an increment of 15 degrees (no scale transformations). The shift increment was one pixel. We applied 81 equally spaced and overlapping windows of the size 16 x 16 pixels.

Figure 5 shows the feature vectors $\mathbf{\Gamma}$ for a binary chevron pattern in three positions. Apart from slight differences due to our rather coarse sampling scheme, the invariance is obvious. The feature vectors for a binary ring pattern can be seen in Fig.6.

Finally, we consider sections through array $\mathbf{B}$ for $i = 0$, i.e. the "auto comparison" of the input, for both, global and local comparison techniques. Figure 7 shows the "activity" in such sections plotted as a function of the shift coordinates j_x and j_y. In Figs.7a,b,c the input was a modified chevron with the vertical bar having double intensity. The plots let compare the different comparator mechanisms (part c is normalized). Figures 7d,e were calculated for the ring pattern of Fig.6 and we state that the local technique leads to a more pronounced peak.

Concluding Remarks

We introduced a signal representation which is calculated by a fixed parallel computational scheme. Thus, the computational speed strongly depends on the applied amplitude coding, e.g. impulse rates in the nervous system.

The effort spent in our system equally serves all patterns while with most pattern recognition concepts every input presentation is separately treated. This economical aspect becomes particularly important when comparing highly invariant recognition systems, since, e.g. systems of the cross correlation type, require a more than proportional increase of effort caused by the combinatorial explosion of necessary "masks" when demanding increased invariance.

It was shown that invariant features can be derived from our general signal representation that may also serve for or explain the detection and analysis of moving as well as deforming patterns.

Since all components of the array representation are time dependent, we can determine for each so-called Receptive Fields and tuning curves which are used in neurosciences to characterize types of neuronal processing. Hence, we can call into question that the concept of Receptive Fields implies hierarchical processing that is often associated with Perceptron structures. Furthermore, we hope to give some hints for the explanation of constancy phenomena in visual perception which may also be explained by parallel representations of geometrically transformed versions of patterns.

References

|1| Marko,H.; IEEE-Trans. SMC-4, no.1, pp.34-39
|2| Minsky,M.,Papert,S.; Perceptrons. MIT-Press, 1969
|3| Platzer,H.; Proc. 3.IJCPR/Coronado'76, pp.713-715
|4| Platzer,H.,Glünder,H.; in: Machine-Aided Image Analysis. Inst. Physics/London, 1979, pp.114-119
|5| Hofer,J.,Platzer,H.; in: Kybernetik'77. Oldenbourg München, 1978, pp.426-430 (in German)
|6| Report of the SFB 50 for 1977-1979, Project A33. pp.81-90 (in German, available from the authors)
|7| Tilgner,R.D.; Nachrichtentechn. Berichte, vol.8 1982 (in German, available from the authors)
|8| Sternberg,S.R.; in: Realtime/Parallel Computing/ Image Analysis. Plenum-Press, 1981, pp.347-359
|9| Marr,D.; Vision. Freeman, 1982
|10| Lenz,R.; Proc. 7.ICPR/Montréal, 1984, pp.546-548

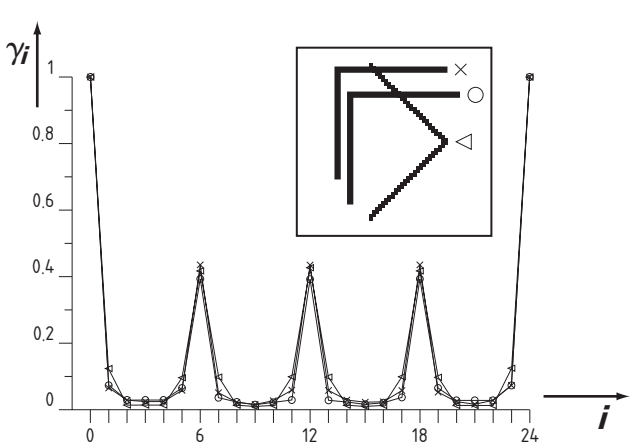

Fig. 5 Feature vectors for a chevron in 3 positions

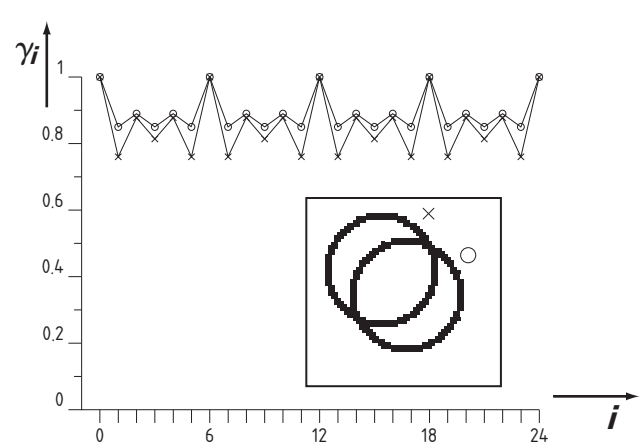

Fig. 6 Feature vectors for a ring in two positions

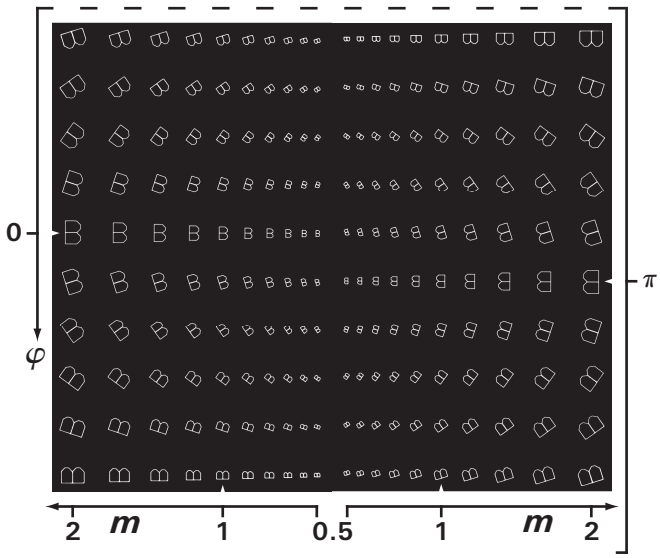

Fig. 4a Output of the transforming unit **T** for the input pattern "B" $\mathbf{T}\{\text{"B"}\}$

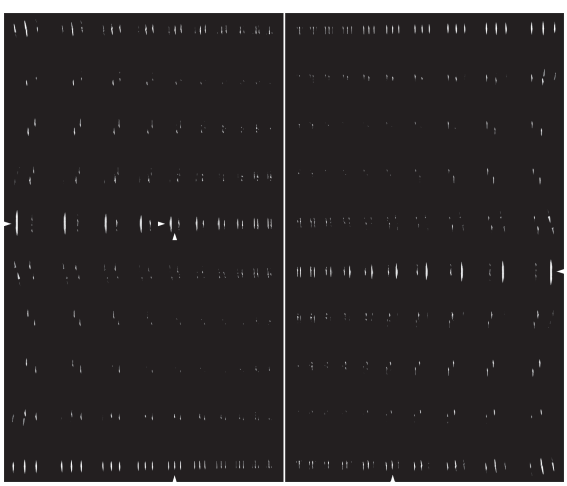

Fig. 4b Section through representation $\overline{\mathbf{\Delta}}$ for the indicated window $\mathbf{T}\{\text{"B"}\} \circledast \text{"B"}$

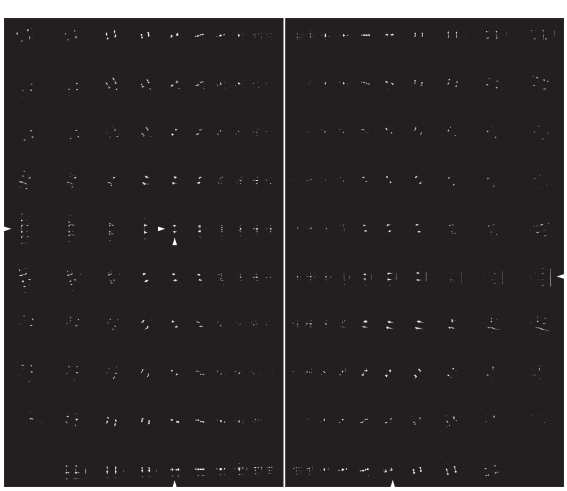

Fig. 4c Same as Fig.4b but for a different window $\mathbf{T}\{\text{"B"}\} \circledast \text{"B"}$

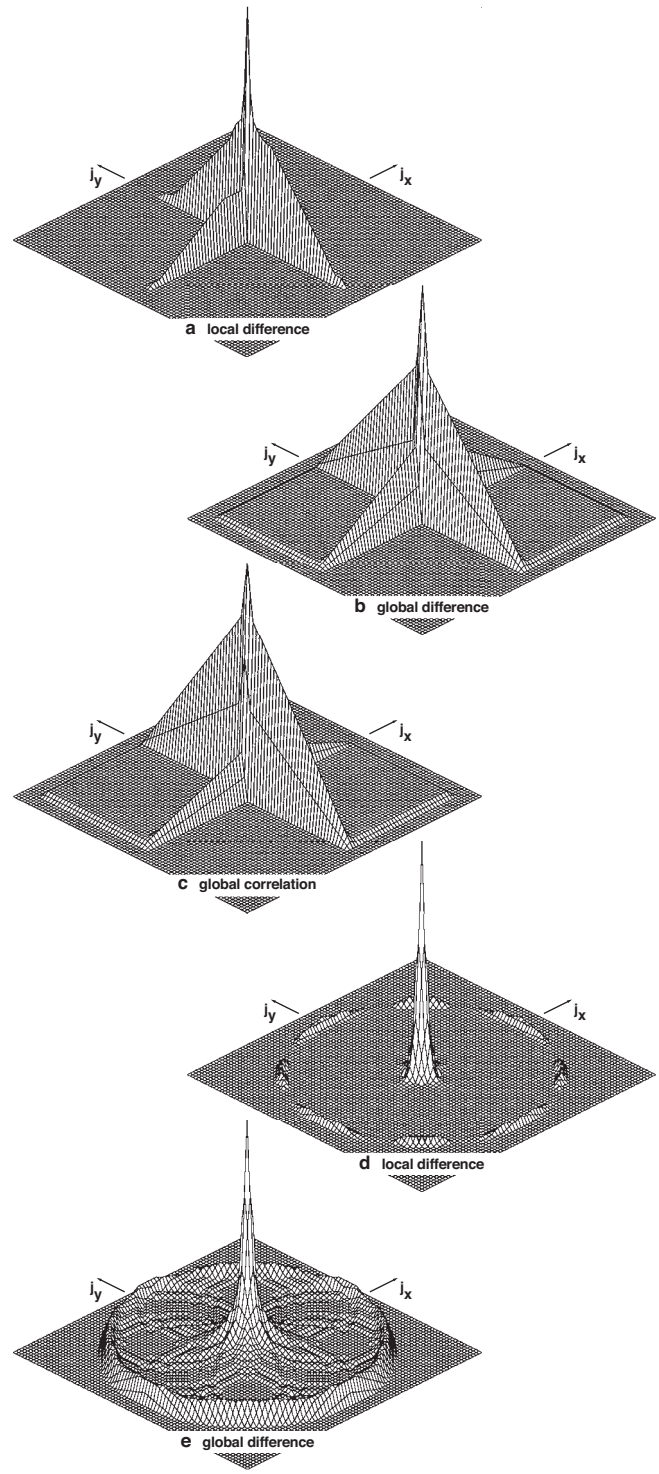

Fig. 7 "Auto comparison functions" (see text)

Acknowledgement

This paper essentially summarizes insights to and results from simulations of neuronal pattern processing obtained during several years of excellent research conditions made possible and supported by the Deutsche Forschungsgemeinschaft (DFG) under the grant SFB 50/A33 in the Sonderforschungsbereich 50 "Kybernetik".