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 for pictorial data that allows one to derive features which are invariant under geometrical transformations. These features are based on intrinsic measures of the given signal. Beside these absolute inner geometrical relations it delivers information about the 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. the human visual system. Thus our approach is not well suited for implementations on v. 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, not some kind of code deciphering, i.e. the correct classification of a pattern independent from its presentation to the input device of the processing system. 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 the 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 will force us to get additional information about the position of the pattern relative to fixed coordinates as to distinguish for example between 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: this pattern looks like a "W" turned by 180 degrees, instead to "learn" a completely new set of characteristic features?

We should emphasize that invariances cannot be achieved merely by so-called invariant processing. Space invariant systems, e.g., deliver 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). Such approaches,however, imply the loss of positional information . We can overcome this disadvantage by using local operations[3,4].

We developed such "relational concepts" after 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 to ease the 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 should 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 neural nets shall be taken into account.
4 Multiplications as well as divisions, e.g. for normalizing 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 bandpassed or skeletted version of the input signal (for this, 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

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 . I.e., these versions are derived from the individual input representation by a fixed processing scheme. Figure 1 illustrates this approach.

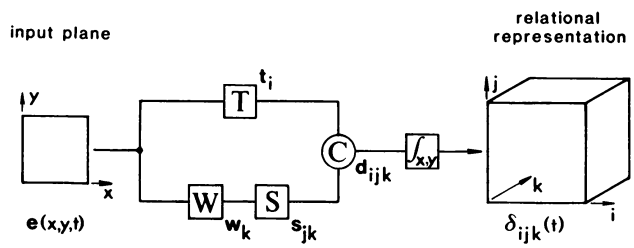

Fig. 1 Basic relational processing concept

The input signal e is fed to the transforming system 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 C. On the other hand the input pattern is windowed by a set W of fixed and highly overlapping window functions which may differ in size, shape and profile and are arbitrarily distributed over the input plane. This decomposition of the input into its parts w_k can again be achieved in parallel using selective connections and weighting. Since we want to achieve independence from a fixed centre of rotation and zoom and to introduce the set of translatory transformations we need a separate unit for shifts in the x and y directions which follows circuit T or, as assumed here, acts on each of the windowed details w_k.All shifted versions s_{jk} can be derived in parallel by appropriate "wiring" to the second input of stage C. Thus the whole 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 pointwise 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.) By this we obtain results which are restricted to values between 0 and the maximum signal amplitude. A perfect match is indicated by a zero output independent of the absolute signal strength in contrary to multiplicative comparators which need normalization in order to allow an interpretation of the results.

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

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

If we assume for better understanding a multiplying comparator 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 such spatially discrete processing we can give some estimates for an optimum bandwidth of the above mentioned differentiating preprocessor: the line width of its output should be about twice the shift increment of unit S.

At this point let us summarize the principal 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 Δ. Though it may be compressed or expanded its characteristic property expressed by the sequence of activity is not changed.

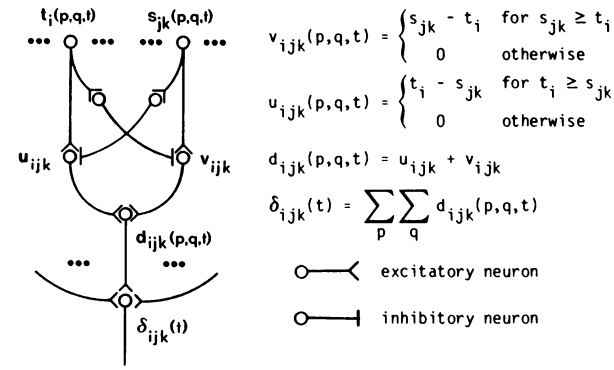

Fig. 2 Model for a neuronal comparator circuit

Extraction of Invariant Features

The relational representation Δ contains a tremendous amount of data and due to the fixed configuration of windows it is in no way transformation invariant.Therefore we have to look for data reduction and for methods to achieve independence from the windows. First we apply a threshold operation Θ. It serves for 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 will be computed as follows:

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

and thus leads to a more significant "activity configuration" $A = \{\alpha_{ijk}\}$.

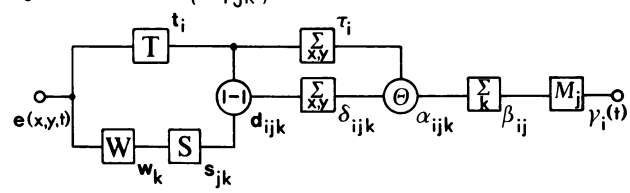

Fig. 3 Block-diagram of the proposed concept

A direct way to get 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 A: we project array A in the k-direction, i.e. we introduce the summation over all windows.Compared with a global concept we achieve two principal advantages:

1 The positional information is saved from the very beginning and thus we can use it together with the characteristic configurations, e.g. for motion analysis[10].
2 We obtain a higher significance for the resulting signal representation due to the threshold operation which is not applicable to global comparison techniques.

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

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

$$\gamma_i = \max_j (\beta_{ij})$$

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

Experimental Results

Figures 4b,c show sections through the representation $\overline{\Delta}$ 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). These results were computed coherent-optically applying correlation filtering[6]. Figure 4a shows the point response of the simulation system which was computer plotted. We implemented 200 transformations: 20 rotations in φ with an increment of 18 degrees and 10 scale factors m between 0.5 and 2. 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 transformation). 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 Γ corresponding to a binary chevron pattern presented in three positions. Apart from slight differences due to our rather coarse sampling the invariance is obvious. The feature vectors for a binary ring pattern can be seen in fig.6.

The following graphs are sections through array 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 over the shift coordinates j_x and j_y. For figs.7a,b,c a modified chevron having a vertical bar of double intensity was used as input. These plots let compare the different comparator mechanisms (part c was normalized). Figures 7d,e were calculated for the ring pattern of fig.6. We can state that the local difference technique leads to the sharpest peak.

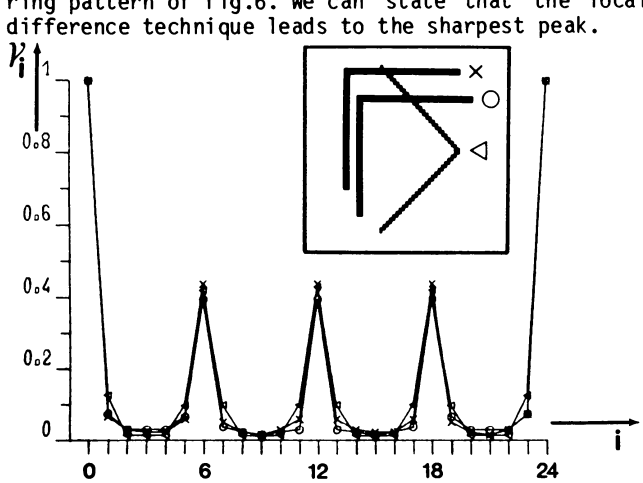

Fig. 5 Feature vectors for a chevron in 3 positions

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 serves for all patterns while in most of the others it is specialized to treat each input presentation separately. This economical aspect becomes particularly important when comparing highly invariant recognition systems. Since, e.g. systems of the cross correlation type will require an overproportional increase of effort caused by the combinatory explosion of necessary "masks" when demanding for more invariance.

It was shown that invariant features can be derived from our general signal representation. On the other hand it may also serve for or explain motion analysis of patterns as well as detection and analysis of temporal changes within a pattern.

Since all components of the array representation are time dependent we can "measure" for each, so-called receptive fields and tuning curves which are used in neurosciences to characterize types of neuronal processing. Hence we can call in question that the measurement of receptive fields implies hierarchical processing often described with perceptron structures. Furthermore we hope to give some hints for an explanation of the constancy phenomenon in visual perception which may also be based on parallel representations of geometrically transformed versions of the input.

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.3rd IJCPR/Coronado'76,pp.713-715
|4| Platzer,H.,Glünder,H.;in:Machine-Aided Image Analysis,The Inst.Physics/London,1979,pp.114-119
|5| Hofer,J.,Platzer,H.;in:Kybernetik '77,Oldenbourg München/Wien,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 N.Y.,1981,pp.347-359
|9| Marr,D.;Vision,W.H. Freeman and Comp. S.F.,1982
|10|Lenz,R.;Proc.7th ICPR/Montréal 1984

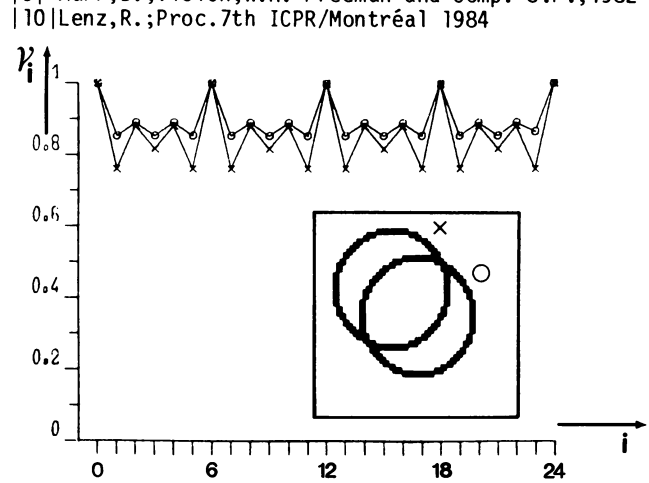

Fig. 6 Feature vectors for a ring in two positions

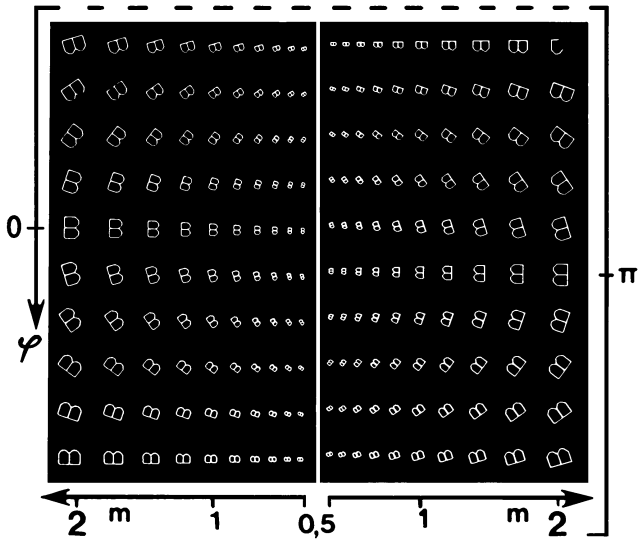

Fig. 4a Output of the transforming unit T for the input pattern "B" $T\{\text{"B"}\}$

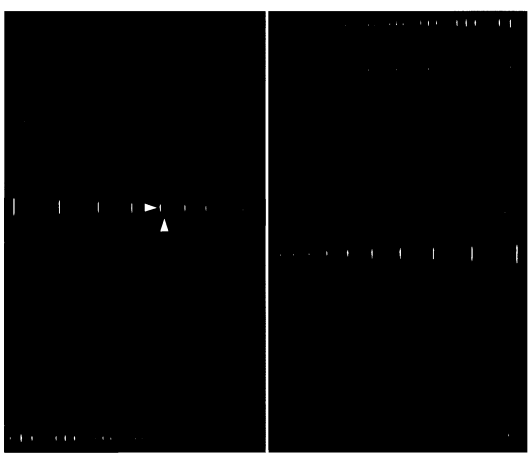

Fig. 4b Section through representation $\underline{\Delta}$ for the indicated window $T\{\text{"B"}\} \circledast \text{"B"}$

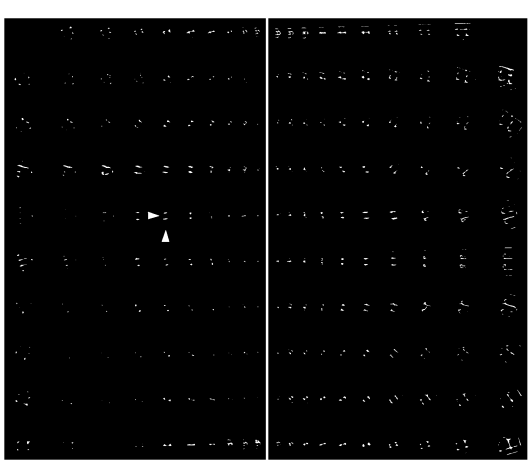

Fig. 4c Same as fig.4b but for a different window $T\{\text{"B"}\} \circledast \text{"D"}$

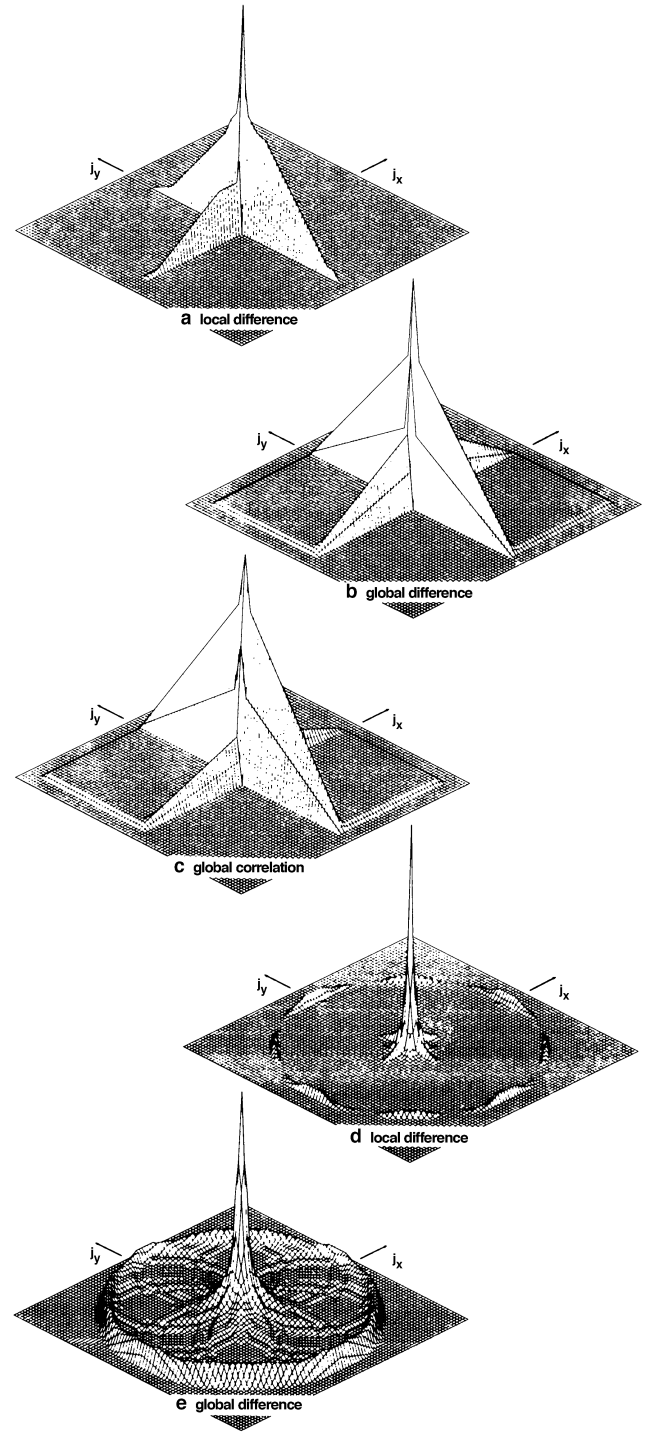

Fig. 7 "Auto comparison functions" (see text)

Acknowledgement

This paper somewhat summarizes the 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 within the Sonderforschungsbereich 50 "Kybernetik".

Seventh International Conference on Pattern Recognition

Montreal, Canada
July 30 – August 2, 1984

Proceedings
Volume 2

Organizers:

CIPPRS:
Canadian Image Processing and Pattern Recognition Society

IAPR:
International Association for Pattern Recognition

ISBN 0-8186-0545-6
IEEE Catalog No. 84CH2046-1
Library of Congress No. 84-80909
Computer Society Order No. 545

Printed in the United States of America

IEEE COMPUTER SOCIETY PRESS