

ΣΠ-NETWORKS FOR MOTION AND INVARIANT FORM ANALYSES

Helmut GLÜNDER

Institut für Medizinische Psychologie
 der Ludwig-Maximilians-Universität
 Goethestraße 31, D-8000 München 2
 Federal Republic of Germany

Spatiotemporal relations that are represented by properly chosen multilinear forms of signal values are shown to be useful for form-independent motion analyses and for the extraction of form descriptors that are invariant under geometric transformations. Both tasks are performed by identical parallel computing ΣΠ-networks.

1. INTRODUCTION

The extraction of geometric relations between two or more points of a pictorial pattern plays a fundamental role in the evaluation of form-related pattern properties that are invariant under geometric transformations. Since any kind of statements about the presence or the degree of such relations involve decisions, nonlinear combinations of the signal values at two or more sites of a static image are inevitable. Multilinear terms, i.e. products of signal values that are taken from an image, represent convenient measures of spatial coincidences (Figure 1a). Motion analysis, on the other hand, additionally requires the evaluation of temporal relations, i.e., between those values of a continuously time-varying image that are found at the same location but are separated by one or more time intervals (Figure 1b). Together with a system for the analysis of spatial relations, spatiotemporal coincidence detection (Figure 2), i.e., motion analysis becomes feasible. It is remarkable that, from this point of view, the gradient and the correlative approaches to motion analysis do not differ.

For both applications, pooling of defined sets of multilinear terms, i.e., the computation of multilinear forms, is requisite if invariance of the form descriptors under geometric transformations, or independence of the velocity estimates from object form are demanded. These conditions are dual and, as will be shown, can be established by a common computing structure that is itself based on this dualism, in so far, as biological processing is concerned: Firstly, there is no need for invariant features if there is no variance, i.e., neither motion nor movement. Secondly, there is no self-organization of the highly specific motion analyzer networks if they are to function dependent on object form, because great numbers of equal stimuli are needed for the formation of suitable networks.

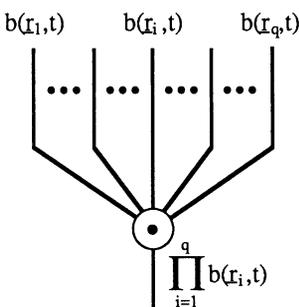


FIGURE 1a

Detection of spatial relations by simultaneous multiplication of q signal values

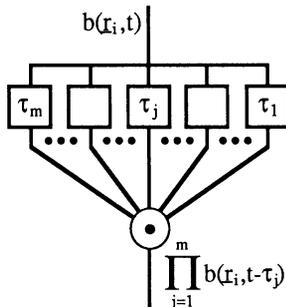


FIGURE 1b

Temporal coincidence detector multiplying m signal values sampled at different times at the same location

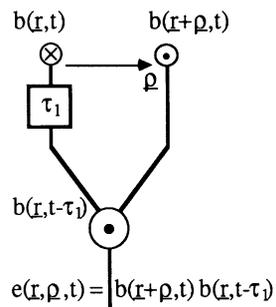


FIGURE 3

Simplest processing unit for the detection of spatiotemporal relations

2. BASIC COINCIDENCE DETECTORS

Figure 1a,b shows schematics of multiplying computing units for the extraction of spatial and temporal relations from a continuously time-varying image $b(\underline{r},t)$, with $\underline{r}=(x,y)^T$, which result in q - and m -linear terms respectively. In Figure 2 a general circuit for the analysis of spatiotemporal relations is depicted which leads to qm -linear terms. Although terms of orders $qm>2$ may be needed to solve special tasks in motion and invariant form analysis, the simplest spatiotemporal detector of order $qm=2$, with $\tau_{11}=\tau_1=\text{const}$ and $\tau_{21}=0$, is used in what follows. Such a unit represents a so-called ρ -tuned velocity detector with its tuning velocity defined by $v_{\text{tun}}=(\underline{r}_2-\underline{r}_1)/\tau_1=\rho/\tau_1$ (Figure 3). In other words, this bilocal detector delivers a coincidence signal $e(\underline{r},\rho,t)$, if a bright spot moves from point \underline{r} straight to point $\underline{r}+\rho$ with the constant speed $|v_{\text{tun}}|$.

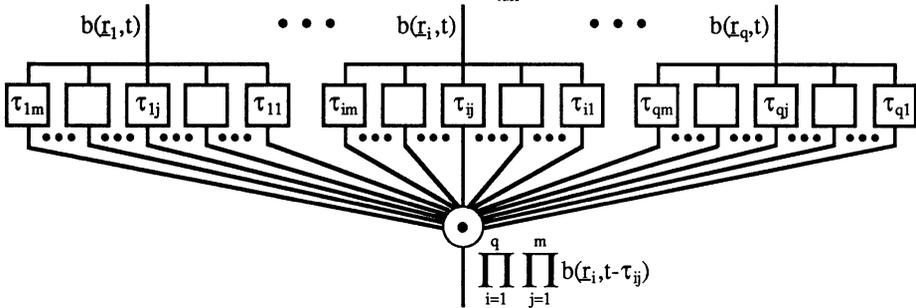


FIGURE 2

General analyzer unit for spatiotemporal relations of time-varying images producing qm -linear terms

3. MOTION ANALYSIS

If a single detector unit of the type shown in Figure 3 is confronted with motion of more complex but still rigid patterns $b(\underline{r})$, it will in general give rise to form-related deviations from the ideal velocity estimate. This is due to the well-known correspondence and window problems (see for instance [1]). They can be eased or even eliminated if extended arrays of elementary detectors are applied in combination with suitable schemes for the pooling of their coincidence signals. For the analysis of frontoparallel translations, for instance, a spectrum of differently tuned units must be isotropically arranged around every location of the image representation and the coincidence signals of all units with identical vectors ρ must be summed. Thereby, the modified autocorrelation function (MACF) of second order is computed, with each of its coefficients representing a bilinear form.

$${}^2k_{\rho d}(\rho, t) = \iint b(\underline{r} + \rho, t) b(\underline{r}, t - \tau_1) d\underline{r} \quad \text{for real-valued } b(\underline{r}, t) \quad (1)$$

Form-independent velocity estimates $v_d(t)=\rho_d(t)/\tau_1$ can be obtained from the MACF by evaluating the coordinates $\rho_d(t)$ of its absolute maximum which is always defined if the pattern is nonperiodic.

In order to analyze expansions or frontoparallel rotations, the input signals to each detector unit are taken from one of the straight lines of a line-bundle, or from one of the circles of a set of concentric circles respectively. In the first case, summation is applied to all bilinear terms that stem from units whose input positions are related by the same scale factor s with respect to the point of intersection of the line-bundle. In the second case, all terms are summed that result from units whose input positions are related by the same rotation angle ψ with respect to the centre of the circles. A universal formulation of such systems is given by the generalized MACF of order q which also allows for any kinds of impulse responses $h(t)$, and not only of the ideal delay $h_d(t)=\delta(t-\tau_1)$.

$${}^qk_{z_h}(z_1 \dots z_q, h_1 \dots h_q, t) = \iint \prod_{i=1}^q [b(\mathcal{J}(z_i)(\underline{r}), t) * h_i(t) \delta(\underline{r})] d\underline{r} \quad \text{for real-valued } b(\underline{r}) \quad (2)$$

Operator $\mathcal{J}(z_i)$ denotes mappings $\underline{r} \rightarrow \underline{r}'$ defined by parameter vectors z_i . It was shown recently [2] that form-independent velocity estimates, e.g. of frontoparallel translations $v_p(t)=\rho_p(t)/\tau_1$, can also be obtained from systems comprising bilocal detectors with lowpass filters of arbitrary characteristic, if the unequivocally defined centroid location $\rho_c(t)$ of the generalized MACF is used in conjunction with a characteristic time constant τ_h of the filter. Since the evaluation of centroids relies on in-

tegral operations, it is less susceptible to noise and less costly than maximum detection.

Owing to the specific network structure, i.e., to the geometric arrangement of the detector units in the plane of analysis, each analyzer system 'views' a specific aspect of object motion. Consequently, certain kinds of compound motion are decomposed according to the implemented systems. Cycloidal motion of points on a rolling wheel, for instance, can be decomposed into their translatory and rotatory components. This selectivity allows one to deal with motion of nonrigid objects as well [2]: Each system reacts to motion of moderately deforming patterns by adequate deviations of its analysis from the one that would have been caused by the corresponding rigid motion process. The fact that velocity estimates are independent of the object form is essentially due to the spatial integration of motion information. Globally acting systems, however, can only analyze motion of a single object correctly and thus, among other reasons, local autocorrelative networks are required. They are feasible, if the regions of analysis allow at least two applications of the considered transformation parameter and if the overlap between neighbouring regions measures its single application. In order to obtain regional or global MACFs, the locally computed results need simply be added. This grouping especially makes sense if it depends on size and location of the object [2,3,4].

Higher order analyses (cf. Figure 2) become necessary if analyzer systems are to detect not only specific types of motion but velocity profiles, such as those caused by motion or movement along the line of sight. In this case, scale or expansion factors are a hyperbolic function of time which may be analyzed by a third order system (qm=3) built from elementary detectors incorporating two different delay times $\tau_{11} > \tau_{21} \neq 0$ and $\tau_{31} = 0$. However, this approach to the analysis of depth motion depends strongly on object size. One can cope with this problem by using a two-stage design of second order systems which computes the reciprocal of the 'time to contact' [5].

4. INVARIANT FORM ANALYSIS

If motion analyzers of the type described by equation (2) are confronted with stimuli that remain static for periods longer than the maximum duration of its impulse responses $h_1(t)$, their reaction is described by the generalized autocorrelation function (ACF) of order q

$${}^q k_z(z_1 \dots z_q) = \iint \prod_{i=1}^q b(\mathcal{T}(z_i)\{\underline{r}\}) d\underline{r} \quad \text{for real-valued } b(\underline{r}), \quad (3)$$

with each of its coefficients representing a q-linear form. Generalized ACFs that depend on a single type of transformation are inherently invariant under this very transformation. This holds true for every single autocorrelation coefficient, since all q-linear terms that stem from image points that are related by the same set of transformation parameters $z_1 \dots z_q$ are summed. The invariance of autocorrelative features is exclusively due to this spatial averaging. Hence, the extent of feature invariance for a given type of geometric transformation depends on the number of q-linear terms that contribute to a coefficient. For certain groups of geometric transformations it is even possible to derive form descriptors from the corresponding generalized ACF that are multiple invariant, e.g. under the whole group of similitudes [6]. Such features describe patterns by their general symmetries, self-congruences and self-similarities, i.e., by categories known from Gestalt-psychology.

Even if all coefficients, e.g. of the translatory second order ACF ${}^2 k_p(\rho)$, are available, they do not unequivocally determine the form of a pattern, i.e., its exact reconstruction is not possible. This situation is due to the loss of information about the relative positions between the two-point relations of a pattern (loss of coherence). The translatory third order ACF ${}^3 k_p(\rho_1, \rho_2)$, with $\rho_3 = 0$, however, perfectly defines patterns of finite extent, except for their shift position (invariance), and thus allows for the unequivocal reconstruction of their form [7,8,9]. Comparable considerations hold for the generalized triple ACF ${}^3 k_{s\psi}(z_1, z_2)$, with $z_i = (s_i, \psi_i)^T$ and $z_3 = 0$, which depends on scale factors s and rotation angles ψ . This becomes immediately clear if one imagines an image representation in polar coordinates with a logarithmically graded radial axis, for which both types of transformations are changed into pure translations. Pattern description by this ACF is rotation as well as scale invariant with respect to the considered fixed point. The reconstruction of the pattern form is unequivocally possible but will not include size and angular position. It remains to be mentioned that pattern reconstruction in both cases is essentially based on all those autocorrelation coefficients that depend on the parameter vector by which the two most distant points of a pattern are related. Therefore, local analyzers, even if the results of many of them are combined, cannot provide perfect invariant pattern descriptions (cf. the similar conclusions in [9]).

5. COMPUTING ELEMENTS FOR MULTILINEAR FORMS

The direct computation of a q -linear form is performed by a $\Sigma\Pi$ computing element which contains coincidence detectors and sums their q -linear terms. Hence, each coefficient of a generalized MACF can be calculated by a single $\Sigma\Pi$ element, provided the input signals are properly grouped. A polynomial approximation, using threshold logic elements (TLUs), results in cross correlation coefficients of the time-varying image and spatiotemporal masks that consist of q nonzero values. These masks are the counterparts of the q signal values which contribute to each q -linear term. The computed coefficients are nonlinearly weighted and summed according to the same schemes that serve for the production of the corresponding q -linear forms. A necessary condition for their approximation is a nonlinearity of polynomial degree $p \geq q$. The most severe problem associated with TLU-approaches is the fact that the desired q -linear terms are contained in sums with many other polynomial terms. The latter are either redundant, if, e.g. all q signals are nonzero, or they may distort the result, if, e.g. one of them is zero: $(1+1+1)^3 = (1.5+1.5+0)^3$. Furthermore, processing must cope with a large signal dynamic and provide sufficient accuracy in order to evaluate the contribution of q -linear terms to the polynomials. Obviously, the situation is aggravated after pooling.

Since networks of orders $q > 3$ do not considerably improve the performance but – owing to the exponential decrease of the probability for coincidences –, have little chance to self-organize, the optimum order is estimated to be $2 \leq q \leq 3$. 'TLU-neurons' that perform cross correlations by linearly weighted synaptic transmission of signal values should therefore receive only about three input signals which is inconsistent with neurobiology. A direct evaluation of q -linear forms by TLUs, i.e., through masks of large q in conjunction with polynomial degrees $2 \leq p \leq 3$, is in general not feasible for the pooling schemes considered here. ' $\Sigma\Pi$ -neurons' perform coincidence detection by nonlinear dendritic interaction of neighbouring synaptic inputs and produce outputs that essentially are sums of coincidence signals. It is conjectured that coincidence detection in real neurons is based on dendritic conductance changes that depend in a highly nonlinear way on the membrane potential [10]. Synaptic modification, i.e., self-structuring, is assumed to depend on the success of the coincidence [11]. Thus, learning and processing are based on the same local mechanism. However, a global rule is additionally needed for the self-organization of the highly specific pooling structure.

ACKNOWLEDGEMENTS

This paper was prepared and written while on leave at the 'Ecole Nationale Supérieure des Télécom. de Bretagne'. The support of the 'Dépt. Math. et Syst. de Com.' is acknowledged.

REFERENCES

- [1] Poggio, T.; Woodward, Y. and Torre, V., Optical flow: computational properties and networks, biological and analog, in: Durbin, R.; Miall, C. and Mitchison, G. (eds.), *The Computing Neuron* (Addison-Wesley, Wokingham, 1989) pp.355-370.
- [2] Glünder, H., Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors, *J. Opt. Soc. Am. A* **7** (1990) 317.
- [3] Bülthoff, H.; Little, J. and Poggio, T., A parallel algorithm for real-time computation of optical flow, *Nature* **337** (1989) 549.
- [4] Yuille, A.L. and Grzywacz, N.M., A computational theory for the perception of coherent visual motion, *Nature* **333** (1988) 71.
- [5] Glünder, H., A dualistic view of motion and invariant shape analysis, in: Simon, J.C. (ed.), *From Pixels to Features* (Elsevier, Amsterdam, 1989) pp.323-332.
- [6] Glünder, H., Invariant description of pictorial patterns via generalized autocorrelation functions, in: Meyer- Ebrecht, D. (ed.), *ASST '87* (Springer, Berlin, 1987) pp.84-87.
- [7] Lohmann, A.W. and Wirtzner, B., Triple correlations, *Proc. IEEE* **72** (1984) 889.
- [8] McLaughlin, J.A. and Raviv, J., Nth-order autocorrelations in pattern recognition, *Information and Control* **12** (1968) 121.
- [9] Minsky, M. and Papert, S., *Perceptrons* (The MIT Press, Cambridge/MA, ¹1969, ²1988)
- [10] Dingleline, R., N-methyl-aspartate activates voltage-dependent calcium conductance in rat hippocampal pyramidal cells, *J. Physiol.* **343** (1983) 385.
- [11] Kelso, S.R.; Ganong, A.H. and Brown, T.H., Hebbian synapses in hippocampus, *Proc. Natl. Acad. Sci. USA* **83** (1986) 5326.

PARALLEL PROCESSING IN NEURAL SYSTEMS AND COMPUTERS

Edited by

Rolf ECKMILLER

Division of Biocybernetics
Department of Biophysics
Heinrich Heine University Düsseldorf
F.R.G.

Georg HARTMANN

Electrical Engineering Department
University of Paderborn
F.R.G.

Gert HAUSKE

Communication Science Department
Technical University München
F.R.G.

PARTICIPANTS EDITION



1990

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO