Carefully reworded reprint of a contribution to:
Artificial Neural Networks, 2
I. Aleksander and J. Taylor (Editors)

Relaxation in 4D state space – A competitive network approach to object-related velocity vector-field correction

H. Glünder[¶] and A. Lehmann[*]

Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität
Gœthestraße 31, D-8000 München 2, F.R. Germany

[*] Institut für Nachrichtentechnik, Technische Universität München
Arcisstraße 21, D-800 München 2, F.R. Germany

Abstract

A common principle of (energy-)minimization is applied to the problem of visual motion analysis. In contrast to well-known mathematical optimization procedures and universal optimizing networks, it is proposed to use a problem-adapted network architecture. Because of the here considered bilocal coincidence-type motion detector, object-related motion analysis appears as a geometric correspondence problem. Hence, the correct spatio-temporal correspondences between elements in consecutive images must be selected from all possible ones. This is performed by neighborhood operations that are repeatedly applied to the instantaneous signal representation in the space/velocity-domain until an acceptable estimate of the actual flow-field is reached.

1. INTRODUCTION

Per definition, any elementary measurement of velocity $\vec{v} = \mathrm{d}\vec{r}/\mathrm{d}t \approx \Delta\vec{r}/\Delta\mathrm{t}$ means a local operation. Consequently, motion in 2D images, for example from projections of 3D scenes, is often represented as dynamic velocity vector-fields $\vec{v}(\vec{r}, t)$. It is a basic fact that – without knowledge of (static) properties, such as shape (e.g. curvature), illumination and reflectance of the moving parts of an image – elementary velocity measurements do not suffice to characterize their motion [1, 2, 3, 4]. Although form analysis principally may precede the measurements of velocity, the proposed concept deals with *a posteriori* corrections of raw vector-field. Most problems associated with this concept are theoretically well understood and various so-called computational approaches – essentially algorithmic minimization procedures applied to suitably constrained integral expressions – have been proposed [for a survey see 5]. However, object-related network approaches – other than the resistive network proposed by Wang *et al.* [6] – are lacking.

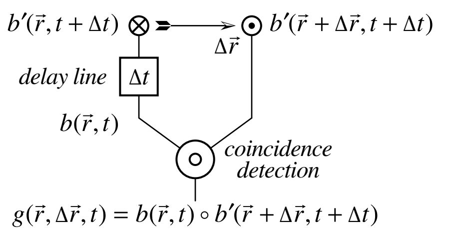
The here introduced concept is based on a constraint-defining kind of network that turned out being structurally related to the one Marr and Poggio [7] proposed for the computation of stereo disparity. It is object-related, thus there are no problems with the proper choice of integration areas as in [8] and it shows motion bias [9]. Despite local operations, that imply spatially confined network interconnections, it propagates velocity estimates into object regions where reliable measurements are not feasible. Unlike most approaches, the concept permits the computation of more than a single velocity vector at any location, a property that is requisite for the explanation of ‘motion transparency’ and associated after-effects [10, 11]. In contrast to cooperative resistive networks [6] it uses excitatory and inhibitory interactions and therefore it must be termed spatially competitive.

¶ The author is supported by the “Volkswagen-Stiftung” under the grant I/65 914.

2. SIGNAL REPRESENTATION IN 4D STATE SPACE

Instead of using elementary velocity *measurements*, for example according to the so-called gradient scheme, spatio-temporal correspondence *detection* is applied at the front-end: According to psychophysical evidence [10, 12] ensembles of spatio-temporal coincidence detectors, i.e. velocity-tuned versions [13] of Hassenstein/Reichardt-subunits [14] – of constant delay Δt but various spans Δr (see Figure 1) and direction angles – are assumed at every location of the pictorial representation. Though the latter may in principle contain form features – such as oriented edges – basic contrast measures, hence minimal information, are considered here. Such an arrangement of detector ensembles indicates all correspondences of gray-values (velocity spectrum) at every element (pixel) of the 2D representation, as depicted in the insert of Figure 2 for pixel 'D' of the rigidly displaced 2D object. Consequently, pictorial correspondences are binary (coincidence/no coincidence) signals $g(\vec{r}, \vec{v}, t)$ in the 4D space/velocity-domain. Any *retinotopic* projection of this 4D state space onto a 2D surface results in velocity-vector hyper-columns.

Figure 1. Detector for spatio-temporal coincidences tuned to the velocity $\vec{v} \approx \Delta\vec{r}/\Delta t$



3. NETWORK ARCHITECTURE FOR VECTOR-FIELD CORRECTION

For the 4D representation defined above, the correction of velocity vector-fields consists in the selection of appropriate vectors from every velocity spectrum. To change this generally ill-posed problem into a better defined one, it is usually assumed that the true velocity vectors at neighboring pixels are identical (constraint 1) or at least similar (constraint 2). Hence, the idea is to discard all those vectors from a spectrum that are not common or sufficiently similar to the spectra at the surrounding pixels, and to continue this process until a qualitatively or temporally defined criterion is met. This relaxation process can be performed by networks in which summation/threshold-units (formal neurons) stand for points in the 4D state space – each representing a specific velocity vector at a certain pixel.

Constraint (1) applies perfectly to rigid objects that translate in isolation. The corresponding network is of the conjunctive type: A unit's threshold is surpassed, i.e. the attributed velocity vector survives, if and only if the same vector is present at *all* its activated neighboring pixels. This strict regime is of little value, if arbitrary object motion is to be described by local translations. Thus, a more flexible network implementation (constraint 2) is considered: Every active unit selectively excites its neighbors of the same or a similar vector specification and inhibits units of vectorially neighboring velocities at the proper pixel. With respect to a retinotopic representation, this scheme results in (velocity-)specific *inter*columnar excitation and in (patchy) *intra*columnar inhibition that are known from the cortical area 17 [15]. Figure 2 exemplifies this relaxation approach for the shifted object that is shown in the insert. The smallest isotropic excitatory $w_{\text{exc}}(\vec{r})$ and inhibitory $w_{\text{inh}}(\vec{v})$ correlation subkernels (receptive subfields) are applied. They are shown for a single unit (vector) at pixel 'G'. In this temporally discrete simulation, stable velocity estimates at all pixels are reached after $n = 4$ relaxation steps. The two remaining vectors at pixel 'F' nicely demonstrate the feasibility of (in this case incorrectly) coexisting correspondences – a property that results from the less strict constraint (2). In this example, the constant threshold θ of the units equals the coincidence value and all coupling strengths are the same $|w_{\text{exc}}| = |w_{\text{inh}}| = \varepsilon$. The choice of the latter is not critical as long as the condition $|\varepsilon| < 1/(m-1)$ is met, where m is the number of pixels of the excitatory subkernel.

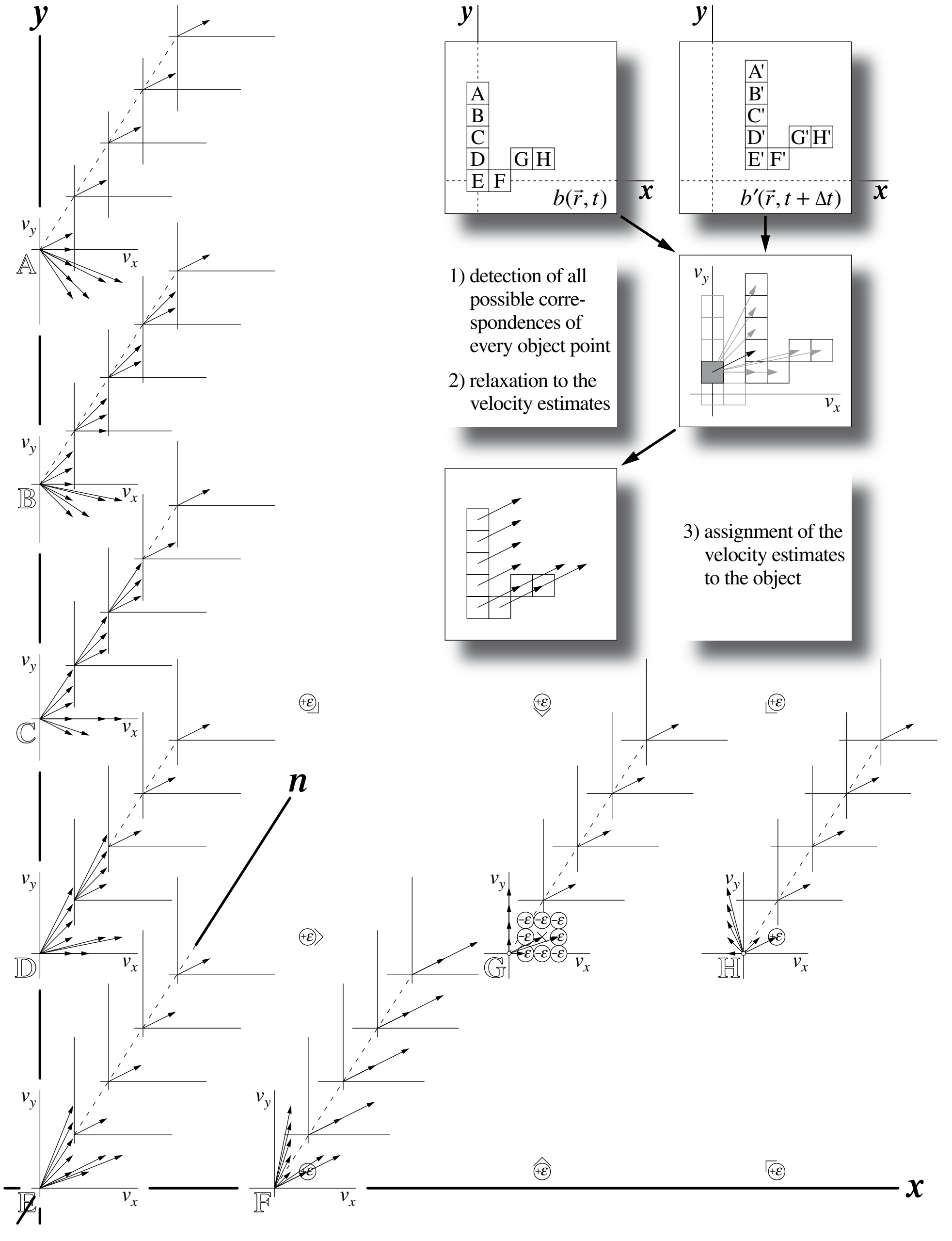


Figure 2. Relaxation in a planar 'hyper-columnar' representation of the 4D space/velocity-domain. The insert illustrates the idea of local velocity estimation. The 'receptive field' (4D correlation kernel) of the 'hyper-columnar' units is displayed for one unit at location 'G' only. After a fast increase in velocity vector-selectivity at points of high pattern curvature a stable result is reached after $n = 4$ relaxation steps

4. DISCUSSION

For rigid and planar objects that translate in isolation, it can be proven that the correct correspondences must survive if complementary subfields, i.e. $w_{\mathrm{exc}}(\vec{r}) = -w_{\mathrm{inh}}(\vec{v})$, are applied. However and in contrast to constraint (1), additional and wrong correspondences may remain – especially with objects showing extended unstructured regions. A higher degree of flexibility with respect to arbitrary motion can be obtained by interconnection schemes that either incorporate motion-specific excitatory interconnections, or that better comply with deviations from the exact translatory correspondences by means of less specific excitation. While the former method leads to a combinatorial explosion of the types of receptive-fields, the latter results in an even higher percentage of wrong correspondences. Owing to constraint (2), that relies on smoothly varying vector fields, any discontinuity that may be caused by immediately neighboring objects of different velocities, as in the case of structured backgrounds, disturbs the local velocity estimation, but – within limits – compliance can be traded in for a loss of correctness in cases of pure translations. For continuous motion, the excitation gives rise to motion bias [9], i.e. objects enter selectively sensitized regions thus accelerating and improving the relaxation.

More generally, the relaxation process can be formulated as a 4D correlation with the speed-variant kernel function $w\left(\vec{r}/cv, \vec{v}/cv\right)$ – for example a (not necessarily isotropic) Gaussian function – that is followed by a (sigmoidal) nonlinear transfer characteristic $\mathcal{S}$

$$g_{i+1}(\vec{r}, \vec{v}) = \mathcal{S}\left\{g_i(\vec{r}, \vec{v})\ \underline{\mathrm{corr}}\ w\left(\vec{r}/cv, \vec{v}/cv\right)\right\}, \qquad \text{where } c \text{ denotes a constant.}$$

It turns out that simulations suffer less from the restrictions due to constraint (2), if the relaxation process is restarted when the first units get near the saturation level of function $\mathcal{S}$. Such *temporally* competitive processing better copes with various kinds of motion discontinuities.

5. REFERENCES

[1] Ullman S. (1979) *The Interpretation of Visual Motion*. MIT-Press, Cambridge/MA.

[2] Hildreth E.C. (1984) *The Measurement of Visual Motion*. MIT-Press, Cambridge/MA.

[3] Poggio T., Torre V. and Koch C. (1985) Computational vision and regularization theory. *Nature* **317**: 314-319.

[4] Uras S., Girosi F., Verri A. and Torre V. (1988) A computational approach to motion perception. *Biol. Cybern.* **60**: 79-87.

[5] Nagel H.-H. (1986) Image sequences – ten (octal) years – from phenomenology towards a theoretical foundation. In: Guiho G. (ed.) *Proceedings of the 8th International Conference on Pattern Recognition, vol. 2*. IEEE Computer Society Press, Washington/DC, pp. 1174-1185.

[6] Wang H.T., Mathur B. and Koch C. (1989) I thought I saw it move: computing optical flow in the primate visual system. In: Gluck M.A. and Rumelhart D.E. (eds.) *Neuroscience and Connectionist Theory*. Erlbaum, Hillsdale/NJ.

[7] Marr D. and Poggio T. (1976) Cooperative computation of stereo disparity. *Science* **194**: 283-287

[8] Bülthoff H., Little J. and Poggio T. (1989) A parallel algorithm for real-time computation of optical flow. *Nature* **337**: 549-553.

[9] Snowden R.J. (1989) Motions in orthogonal directions are mutually suppressive. *J. Opt. Soc. Am. A* **6**: 1096-1101.

[10] Doorn A.J. van and Koenderink J.J. (1982) Spatial properties of the visual detectability of moving spatial white noise. *Exp. Brain Res.* **45**: 189-195.

[11] Doorn A.J. van, Koenderink J.J. and Grind W.A. van de (1985) Perception of movement and correlations in stroboscopically presented noise patterns. *Perception* **14**: 209-224.

[12] Grind W.A. van de, Koenderink J.J. and Doorn A.J. van (1986) The distribution of human motion detector properties in the monocular visual field. *Vision Res.* **26**: 797- 810.

[13] Glünder H. (1990) Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors. *J. Opt. Soc. Am. A* **7**: 255-263.

[14] Hassenstein B. and Reichardt W. (1956) Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*. *Z. Naturforschg.* **11b**: 513-524.

[15] Braitenberg V. and Schüz A. (1991) *Anatomy of the Cortex*. Springer, Berlin.