

Detection of First-order Elementary Components in Noisy Optic Flow Fields Through Context Sensitive Recurrent Filters

Silvio P. Sabatini, Fabio Solari, and Giacomo M. Bisio

Dept. of Biophysical and Electronic Engineering, University of Genoa
Via all'Opera Pia, 11a - 16145 Genova, Italy
{silvio, fabio, bisio}@dibe.unige.it
<http://www.pspc.dibe.unige.it>

Abstract. Measured optic flow fields are always somewhat erroneous and/or ambiguous. First, we cannot compute the actual spatial or temporal derivatives, but only their estimates, which are corrupted by image noise. Second, optic flow is intrinsically an image-based measurement of the relative motion between the observer and the environment, but we are interested in estimating the actual motion field. However, real-world motion field patterns contain intrinsic statistic properties that allow to define Gestalts as groups of pixels sharing the same motion property. By checking the presence of such Gestalts in optic flow fields we can make their interpretation more confident. We propose an optimal recurrent filter capable of evidencing motion Gestalts corresponding to 1st-order spatial derivatives or elementary flow components (EFCs). A Gestalt emerges from a noisy flow as a solution of an iterative process of spatially interacting nodes that correlates the statistics of the visual context with that of a structural model of the Gestalt.

1 Local motion Gestalts

Velocity gradients provide important cues about the 3-D layout of the visual scene. Formally, they can be described as *linear deformations* by a 2×2 velocity gradient tensor

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} \partial v_x / \partial x & \partial v_x / \partial y \\ \partial v_y / \partial x & \partial v_y / \partial y \end{bmatrix}. \quad (1)$$

Hence, if $\mathbf{x} = (x, y)$ is a point in a spatial image domain, the linear properties of a motion field $\mathbf{v}(x, y) = (v_x, v_y)$ around the point $\mathbf{x}_0 = (x_0, y_0)$ can be characterized by a Taylor expansion, truncated at the first order:

$$\mathbf{v} = \bar{\mathbf{v}} + \bar{\mathbf{T}}\mathbf{x} \quad (2)$$

where $\bar{\mathbf{v}} = \mathbf{v}(x_0, y_0) = (\bar{v}_x, \bar{v}_y)$ and $\bar{\mathbf{T}} = \mathbf{T}|_{\mathbf{x}_0}$. By breaking down the tensor in its dyadic components, the motion field can be locally described through 2-D maps representing *cardinal EFCs*:

$$\mathbf{v} = \alpha^x \bar{v}_x + \alpha^y \bar{v}_y + \mathbf{d}_x^x \left. \frac{\partial v_x}{\partial x} \right|_{\mathbf{x}_0} + \mathbf{d}_y^x \left. \frac{\partial v_x}{\partial y} \right|_{\mathbf{x}_0} + \mathbf{d}_x^y \left. \frac{\partial v_y}{\partial x} \right|_{\mathbf{x}_0} + \mathbf{d}_y^y \left. \frac{\partial v_y}{\partial y} \right|_{\mathbf{x}_0} \quad (3)$$

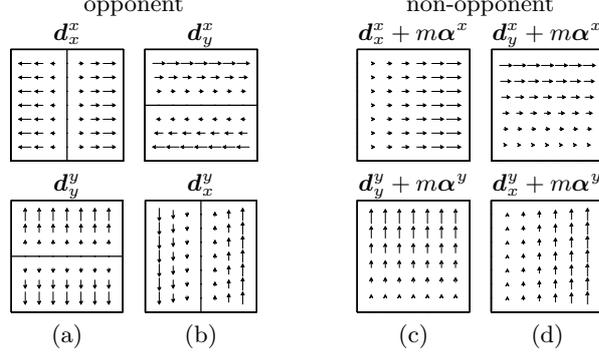


Fig. 1. Basic gradient type Gestalts considered. In stretching-type components (a,c) velocity varies *along* the direction of motion; in shearing-type components (b,d) velocity gradient is oriented *perpendicularly* to the direction of motion. Non-opponent patterns are obtained from the opponent ones by a linear combination of pure translations and cardinal deformations: $\mathbf{d}_j^i + m\alpha^i$, where m is a proper positive scalar constant.

where $\alpha^x : (x, y) \mapsto (1, 0)$, $\alpha^y : (x, y) \mapsto (0, 1)$ are pure translations and $\mathbf{d}_x^x : (x, y) \mapsto (x, 0)$, $\mathbf{d}_y^x : (x, y) \mapsto (y, 0)$, $\mathbf{d}_x^y : (x, y) \mapsto (0, x)$, $\mathbf{d}_y^y : (x, y) \mapsto (0, y)$ represent cardinal deformations, basis of the linear deformation space.

It is worthy to note that the components of pure translations could be incorporated in the corresponding deformation components, thus obtaining generalized deformation components in which motion boundaries are shifted or totally absent. Although this does not affect the significance of the Taylor expansion in Eq. 3, the so-modified elementary components, present very different structural properties. Since a template-based approach cannot be used to extract single components, but only to perform pattern matching operations, the linear decomposition of the motion field has significance only for the definition of a proper representation space. Specific templates would be designed to optimally sample that representation space. In this work, we consider two different classes of deformation templates (opponent and non-opponent), each characterized by two gradient types (stretching and shearing), see Fig. 1. Due to their ability to detect the presence and the orientation of velocity gradients and kinetic boundaries, such cardinal EFCs and proper combinations of them resemble the characteristics of the cell in the Middle Temporal visual area (MT) [1] [2]. It is straightforward to derive that these MT-like components are well suited to provide the building blocks for the more complex receptive field properties encountered in the Medial Superior Temporal visual area (MST) [3] [4]:

$$\mathbf{v} = \alpha^x \bar{v}_x + \alpha^y \bar{v}_y + \frac{1}{2}(\mathbf{d}_x^x + \mathbf{d}_y^y)E + \frac{1}{2}(\mathbf{d}_x^x - \mathbf{d}_y^y)\omega + \frac{1}{2}(\mathbf{d}_x^x - \mathbf{d}_y^y)S_1 + \frac{1}{2}(\mathbf{d}_y^x + \mathbf{d}_x^y)S_2$$

where $E = (\bar{T}_{11} + \bar{T}_{22})/2$, $\omega = (\bar{T}_{12} - \bar{T}_{21})/2$, $S_1 = (\bar{T}_{11} - \bar{T}_{22})/2$, $S_2 = (\bar{T}_{12} + \bar{T}_{21})/2$ are the divergence, the curl and the two components of shear

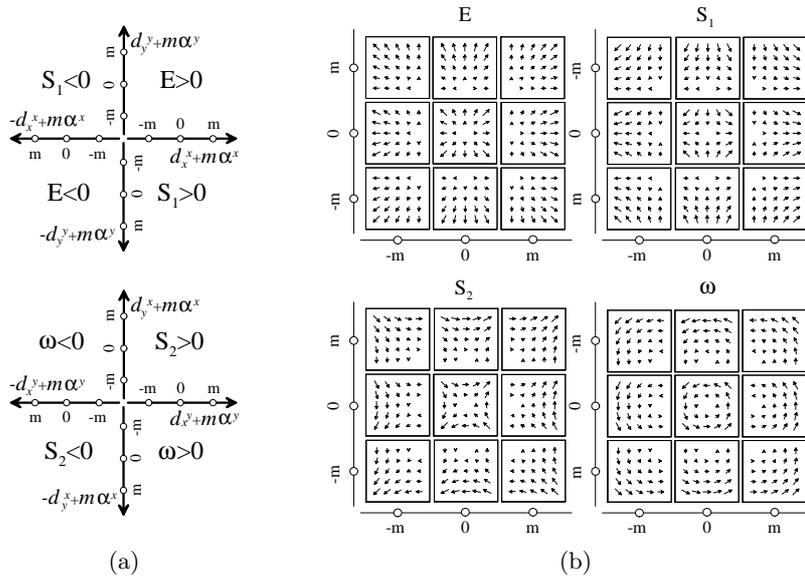


Fig. 2. (a) Two deformation subspaces obtained by the set of cardinal EFCs with different values of the parameter m . The quadrants of each subspace characterize an elementary deformation, as evidenced in (b) for expansion ($E > 0$), horizontal positive shear ($S_1 > 0$), oblique positive shear (S_2), and counterclockwise rotation ($\omega > 0$).

deformation, respectively (cf. [5]). These mixed EFCs constitute, together with the pure translations, an equivalent representation basis for the linear properties of the velocity field (see Fig. 2). Yet, they are rather complex since not only the speed, but also the direction of feature motion varies as a function of spatial position. Rigid body motion often generates simpler flow fields characterized by unidirectional patterns, as the cardinal EFCs considered in this study.

2 The context sensitive filter

The problem of evidencing the presence of a certain complex feature in the optic flow on the basis of both local and contextual information, can be posed as an adaptive filtering problem (estimation), where local information act as the input *measurements* and the context acts as the *reference signal*, e.g., representing a specific motion Gestalt. In the following, we propose a solution in the form of a generalized Kalman filter (KF) [6]. Due to its recurrent formulation, KF appears particularly promising to design *context-sensitive filters* (CSFs) based on recurrent cortical-like interconnection architectures.

Let us assume the optic flow $\tilde{v}(i, j)$ as the corrupted measure of the actual velocity field $v(i, j)$. The difference between these two variables can be represented

as a constant noise term $\varepsilon(i, j)$:

$$\tilde{\mathbf{v}} = \mathbf{v} + \varepsilon. \quad (4)$$

Due to the intrinsic noise of the nervous system, the neural representation of the optic flow $\mathbf{v}(i, j)[k]$ can be expressed by a *measurement equation*:

$$\mathbf{v}[k] = \tilde{\mathbf{v}} + \mathbf{n}_1[k] = \mathbf{v} + \varepsilon + \mathbf{n}_1[k] \quad (5)$$

where \mathbf{n}_1 represents the uncertainty associated with a neuron's response. The Gestalt is formalized through a *process equation*:

$$\mathbf{v}[k] = \Phi \mathbf{v}[k-1] + \mathbf{n}_2[k-1] + \mathbf{s} \quad (6)$$

with $\lim_{k \rightarrow \infty} \mathbf{v}[k] = \mathbf{v}$ if $\mathbf{n}_2 = 0$. The state transition matrix Φ is *de facto* a spatial interconnection matrix that implements a specific Gestalt rule (i.e., a specific EFC); \mathbf{s} is a constant driving input; \mathbf{n}_2 represents the process uncertainty. The space spanned by the observations $\mathbf{v}[1], \mathbf{v}[2], \dots, \mathbf{v}[k-1]$ is denoted by \mathcal{V}_{k-1} and represents the internal noisy representation of the optic flow. We assume that both \mathbf{n}_1 and \mathbf{n}_2 are independent, zero-mean and normally distributed: $\mathbf{n}_1[k] = N(0, \mathbf{\Lambda}_1)$ and $\mathbf{n}_2[k] = N(0, \mathbf{\Lambda}_2)$. The index k takes explicitly into account the time necessary for spatial recurrence. More precisely, Φ models space-invariant nearest-neighbor interactions within a finite region Ω in the (i, j) plane that is bounded by a piece-wise smooth contour. Interactions occur, separately for each component of the velocity vectors (v_x, v_y) , through anisotropic interconnection schemes:

$$\begin{aligned} v_{x/y}(i, j)[k] = & w_N^{x/y} v_{x/y}(i, j-1)[k-1] + w_S^{x/y} v_{x/y}(i, j+1)[k-1] + s_{x/y}(i, j) \\ & + w_W^{x/y} v_{x/y}(i-1, j)[k-1] + w_E^{x/y} v_{x/y}(i+1, j)[k-1] + n_1^{x/y}(i, j)[k-1] \end{aligned}$$

where (s_x, s_y) is a steady additional control input, which models the boundary conditions. The process equation has a *structuring effect* constrained by the boundary conditions that yields to structural equilibrium configurations, characterized by specific first-order EFCs. The resulting pattern depends on the anisotropy of the interaction scheme and on the boundary conditions. By example, considering, for the sake of simplicity, a rectangular domain $\Omega = [-L, L] \times [-L, L]$, the cardinal EFC \mathbf{d}_x^x can be obtained through:

$$\begin{aligned} w_N^x = w_S^x = 0 & & w_N^y = w_S^y = 0 & & s_x(i, j) = \begin{cases} -\lambda & \text{if } i = -L \\ \lambda & \text{if } i = L \\ 0 & \text{otherwise} \end{cases} & & s_y(i, j) = 0 \\ w_W^x = w_E^x = 0.5 & & w_W^y = w_E^y = 0 & & & & & & \end{aligned}$$

where the boundary value λ controls the gradient slope. In a similar way we can obtain the other components.

Given Eqs. (5) and (6), we may write the optimal filter for optic flow Gestalts. The filter allows to detect, in noisy flows, intrinsic correlations, as those related to EFCs, by checking, through spatial recurrent interactions, that the spatial context of the observed velocities conform to the Gestalt rules, embedded in Φ .

To understand how the CSF works, we define the *a priori* state estimate at step k given knowledge of the process at step $k - 1$, $\hat{\mathbf{v}}[k|\mathcal{V}_{k-1}]$, and the *a posteriori* state estimate at step k given the measurement at the step k , $\hat{\mathbf{v}}[k|\mathcal{V}_k]$. The aim of the CSF is to compute an *a posteriori* estimate by using an *a priori* estimate and a weighted difference between the current and the predicted measurement:

$$\hat{\mathbf{v}}[k|\mathcal{V}_k] = \hat{\mathbf{v}}[k|\mathcal{V}_{k-1}] + \mathbf{G}[k] (\mathbf{v}[\mathbf{k}] - \hat{\mathbf{v}}[\mathbf{k}|\mathcal{V}_{k-1}]) \quad (7)$$

The difference term in Eq. (7) is the *innovation* $\boldsymbol{\alpha}[k]$ that takes into account the discrepancy between the current measurement $\mathbf{v}[\mathbf{k}]$ and the predicted measurement $\hat{\mathbf{v}}[\mathbf{k}|\mathcal{V}_{k-1}]$. The matrix $\mathbf{G}[k]$ is the Kalman gain that minimizes the *a posteriori* error covariance:

$$\mathbf{K}[k] = E \{ (\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathcal{V}_k])(\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathcal{V}_k])^T \} . \quad (8)$$

Eqs. 7 and 8 represent the mean and covariance expressions of the CSF output.

The covariance matrix $\mathbf{K}[k]$ provides us only information about the properties of convergence of the KF and not whether it converges to the correct values. Hence, we have to check the consistency between the innovation and the model (i.e., between observed and predicted values) in statistical terms. A measure of the reliability of the KF output is the Normalized Innovation Squared (*NIS*):

$$NIS_k = \boldsymbol{\alpha}^T[k] \boldsymbol{\Sigma}^{-1}[k] \boldsymbol{\alpha}[k] \quad (9)$$

where $\boldsymbol{\Sigma}$ is the covariance of the innovation. It is possible to exploit Eq. (9) to detect if the current observations are an instance of the model embedded in the KF [7].

3 Results

Fig. 3 shows the responses of the CSF in the deformation subspaces for two different input flows. Twentyfour EFC models have been used to span the deformation subspaces shown in Fig. 2a. The grey level in the CSF output maps represents the probability of a given Gestalt according to the *NIS* criterium: lightest grey indicates the most probable Gestalt. Besides Gestalt detection, context information reduces the uncertainty on the measured velocities, as evidenced, for the circled vectors, by the Gaussian densities, plotted over the space of image velocity.

4 Conclusions

Given motion information represented by an optic flow field, we specified a CSF to recognize if a group of velocity vectors belong to a specific pattern, on the basis of their relationships in a spatial neighborhood. Casting the problem as a KF, the detection occurs through a spatial recurrent filter that checks the consistency between the spatial structural properties of the input flow field pattern

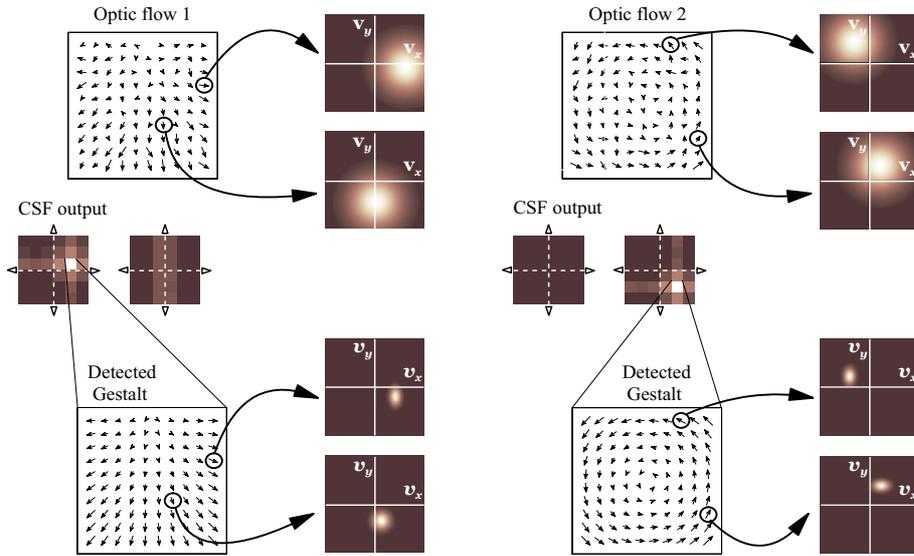


Fig. 3. Example of Gestalt detection in noisy flows.

and a structural rule expressed by the process equation of the KF. The CSF behaves as a template model. Yet, its specificity lies in the fact that the template character is not built by highly specific feed-forward connections, but emerges by stereotyped recurrent interactions (cf. the process equation). Furthermore, the approach can be straightforwardly extended to consider adaptive cross-modal templates (e.g. motion and stereo). By proper specification of the matrix Φ , the process equation can, indeed, potentially model any type of multimodal spatio-temporal relationships (i.e., multimodal spatio-temporal context).

References

1. V.L. Marcar, D.K. Xiao, S.E. Raiguel, H. Maes, and G.A. Orban. Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *J. Neurophysiol.*, 74(3):1258–1270, 1995.
2. S. Treue and Andersen R.A. Neural responses to velocity gradients in macaque cortical area MT. *Visual Neuroscience*, 13:797–804, 1996.
3. C.J. Duffy and R.H. Wurtz. Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J. Neuroscience*, 15:5192–5208, 1995.
4. M. Lappe, F. Bremmer, M. Pekel, A. Thiele, and K.P. Hoffmann. Optic flow processing in monkey STS: A theoretical and experimental approach. *J. Neuroscience*, 16:6265–6285, 1996.
5. J.J. Koenderink. Optic flow. *Vision Res.*, 26(1):161–179, 1986.
6. S. Haykin. *Adaptive Filter Theory*. Prentice-Hall International Editions, 1991.
7. Y. Bar-Shalom and X.R. Li. *Estimation and Tracking, Principles, Techniques, and Software*. Artech House, 1993.