

## Cortical Spatiotemporal Dimensionality Reduction for Visual Grouping

**Giacomo Cocci**

*giacomo.cocci2@unibo.it*

*Department of Electrical, Electronic, and Information Engineering  
"Guglielmo Marconi," University of Bologna, 40136 Bologna, Italy*

**Davide Barbieri**

*davide.barbieri@uam.es*

*Department of Mathematics, Autonomous University of Madrid,  
Facultad de Ciencias, 28049 Madrid, Spain*

**Giovanna Citti**

*giovanna.citti@unibo.it*

*Department of Mathematics, University of Bologna, 40126 Bologna, Italy*

**Alessandro Sarti**

*alessandro.sarti@ehess.fr*

*Centre d'Analyse et de Mathématique Sociales, EHESS, 75244 Paris, France*

The visual systems of many mammals, including humans, are able to integrate the geometric information of visual stimuli and perform cognitive tasks at the first stages of the cortical processing. This is thought to be the result of a combination of mechanisms, which include feature extraction at the single cell level and geometric processing by means of cell connectivity. We present a geometric model of such connectivities in the space of detected features associated with spatiotemporal visual stimuli and show how they can be used to obtain low-level object segmentation. The main idea is to define a spectral clustering procedure with anisotropic affinities over data sets consisting of embeddings of the visual stimuli into higher-dimensional spaces. Neural plausibility of the proposed arguments will be discussed.

### 1 Introduction ---

It is well understood from the psychological theory of the Berliner Gestalt that local properties of the visual stimulus, such as neighboring, good continuation, and common fate, have a central role in the execution of global visual tasks like image segmentation and grouping (Wertheimer, 1938; Wagemans et al., 2012).

A key concept for understanding visual perceptual tasks is that of association fields, introduced by Field, Hayes, and Hess (1993) to describe the structure of the field of good continuation, underlying the recognition of perceptual units in the visual space. These results have been obtained by psychophysical experiments, presenting stimuli made of an ensemble of oriented patches, a subset of which was consistently aligned along a continuous path. The study of these phenomena allowed the identification of properties that a stimulus should have near a given patch in order to recognize such samples as belonging to a curvilinear path, namely, colinearity and cocircularity. The detection of these properties is indeed compatible with the functional behavior of simple cells in the primary visual cortex V1 as linear feature detectors for local orientations (Hubel & Wiesel, 1977).

Several physiological experiments showed how the principles of association fields seem to be implemented in the V1 of mammals, where long-range horizontal connections preferentially link columns of neurons having similar preferred orientation (Bosking, Zhang, Schofield, & Fitzpatrick, 1997). By interpreting cortical columns as directional differential operators, Citti and Sarti (2006) showed how this specialized functional organization of V1 naturally leads to a geometric model of the association fields. The field lines are modeled with a family of integral curves on a contact structure based on the Lie algebra of the group of rigid motions of the Euclidean plane  $SE(2)$ . This geometric approach lies within a well-established research line founded by seminal work (Koenderink & van Doorn, 1987; Hoffman, 1989; Mumford, 1994; Petitot & Tondut, 1999), whose current state of the art can be found in Citti and Sarti (2014).

Further phenomenological experiments demonstrated the central role for the perception of global shapes of the features of movement direction and velocity (Rainville & Wilson, 2005) and that, similar to what happens for the integration of spatial visual information, the brain is capable of predicting complex stimulus trajectories and grouping together elements having similar motion or apparent motion paths (Verghese, Watamaniuk, McKee, & Grzywacz, 1999; Verghese, McKee, & Grzywacz, 2000; Watamaniuk, 2005). Also in this case, the detection of such features is performed at the level of V1, where specialized cells show spatiotemporal behaviors optimized for the detection of local velocities (DeAngelis, Ohzawa, & Freeman, 1995; Cocci, Barbieri, & Sarti, 2012).

The analysis of spatiotemporal properties and organization of cortical visual neurons, together with the indications given by experimental results on visual spatial and motion integration, has recently led to extensions of the  $SE(2)$  model that include local stimulus velocity. Barbieri, Citti, Cocci, and Sarti (2014) introduced new classes of spatiotemporal connectivities, providing a geometric model of association fields in the five-dimensional contact structure of cells' positions and activation times, together with locally detected features of orientations and velocities. Such a structure embeds purely spatial geometry in a layered fashion and integrates the association

mechanisms as extensions to a higher-dimensional space. Their capabilities in the elaboration of trajectories and in the tasks of spatiotemporal image completion showed good accordance with previous psychophysical and physiological results.

The aim of this letter is to study the capabilities of such geometric spatiotemporal connectivities with respect to the tasks of image segmentation and grouping. Such tasks are addressed with spectral analysis; in particular, we discuss and refine the proposed connectivity structures in order to describe them as operators on high-dimensional feature spaces. Our approach represents a spatiotemporal stimulus as a data set in a feature space and considers it as a weighted graph whose affinity matrix is determined by geometric connectivity. The grouping mechanisms that we describe arise from a spectral clustering of the associated graph Laplacian, making use of the probabilistic framework introduced in Meila and Shi (2001) followed by an adaptation of the simple and robust clustering technique proposed in Kannan, Vempala, and Vetta (2004). When dealing with nonsymmetric affinities, we followed the ideas introduced in Pentney and Meila (2005). We have tried to stick with minimal hypotheses at the algorithmic level in order to keep our focus on the role of the kernels. Our main results show that the introduced spatiotemporal geometry provides connectivities suitable for robustly grouping spatiotemporal stimuli, but also that a connectivity pattern based on local stimulus velocity can enhance the spatial grouping capabilities of a visual system.

With respect to the neural implementations of such principles, apart from the study in Barbieri, Citti, Cocci et al. (2014), we can see that psychophysical experiments such as those of Geisler, Perry, Super, and Gallogly (2001); Hess and Ledgeway (2003); and Ledgeway, Hess, and Geisler (2005) addressed the problems of the existence of association fields for local directions of motion and their role in visual grouping, and of comparison of human grouping with cocircular correlations in natural image statistics. Moreover, the recent results of Sarti and Citti (2014) show how spectral analysis of the spatial connectivity introduced in Citti and Sarti (2006), which is the more basic one studied in this letter, is actually implemented by the neural population dynamics of primary visual cortex. This suggests a fundamental role for spectral mechanisms in the phenomenology of perception, indicating that they may be concretely performed by the visual system and, hence, provide a stronger motivation for our detailed spectral analysis of connectivities and their segmentation properties of visual stimuli.

Finally, we note that while cocircularity is naturally implemented in kernels (see Sanguinetti, Citti, & Sarti, 2010), its use for the definition of affinity matrices whose spectra could be suited for line grouping was suggested by Perona and Freeman (1998).

The plan of this letter is as follows. In section 2, we describe the geometry arising from the spatiotemporal functional architecture of the visual cortex as introduced in Barbieri, Citti, and Sarti (2014); Barbieri, Citti, Cocci

et al. (2014), providing more detailed arguments on how to construct and implement the connectivity kernels we will use. In section 3 we recall the notions of spectral analysis on graphs and introduce the basic clustering principles that we adopt. Then we show how to construct affinity matrices based on spatiotemporal cortical geometry, providing discussion on how to deal with the different kinds of asymmetries present in the connectivities. Our main results are in section 4, where we use the previously introduced spectral clustering algorithm to automatically extract perceptual information from artificial stimuli living in the cortical feature spaces. In particular, we propose two connectivity mechanisms to perform spatiotemporal segmentation of motion contours and shapes, also providing parametric evaluations of the kernel performances, and we discuss the relationship with neural processes studied in several psychophysiological experiments.

## 2 The Geometry of V1

---

In this section, we present a model of the functional architecture of the visual cortex as a contact structure, where cortical connectivity is implemented as a diffusion process along its admissible directions. This same approach was taken in Barbieri, Citti, Cocci et al. (2014), where it was compared to psychophysical and physiological behaviors of the visual system.

**2.1 The Cortical Feature Space.** It is well known since the fundamental studies of Hubel and Wiesel (1977; Hubel, 1988) that the primary visual cortex (V1) is one of the first physiological layers along the visual pathway in carrying out geometrical measurements on the visual stimulus, decomposing it in a series of local feature components. The development of suitable electrophysiological techniques (Ringach & Shapley, 2004) has made it possible to reconstruct the linear filtering behavior of V1 simple and complex cells, that is, their spatiotemporal receptive profiles (RPs).

The RPs of orientation-selective cells in V1 have classically been modeled with two-dimensional Gabor functions (Jones & Palmer, 1987), which compute a local approximation of the directional derivative of the visual stimulus, minimizing the uncertainty between localization in position and spatial frequency (Daugman, 1985). Spatiotemporal RPs of V1 simple cells can be modeled by three-dimensional Gabor functions of the form (Cocci et al., 2012)

$$g_{q,p}^{\sigma}(x, y, t) = e^{2\pi i(p_x(x-q_x)+p_y(y-q_y)+p_t(t-q_t))} e^{-\left(\frac{(x-q_x)^2}{2\sigma_x^2} + \frac{(y-q_y)^2}{2\sigma_y^2} + \frac{(t-q_t)^2}{2\sigma_t^2}\right)}, \quad (2.1)$$

where  $q = (q_x, q_y, q_t)$  is the spatiotemporal center of the Gabor filter,  $p = (p_x, p_y, p_t)$  is the spatiotemporal frequency, and  $\sigma = (\sigma_x, \sigma_y, \sigma_t)$  is the

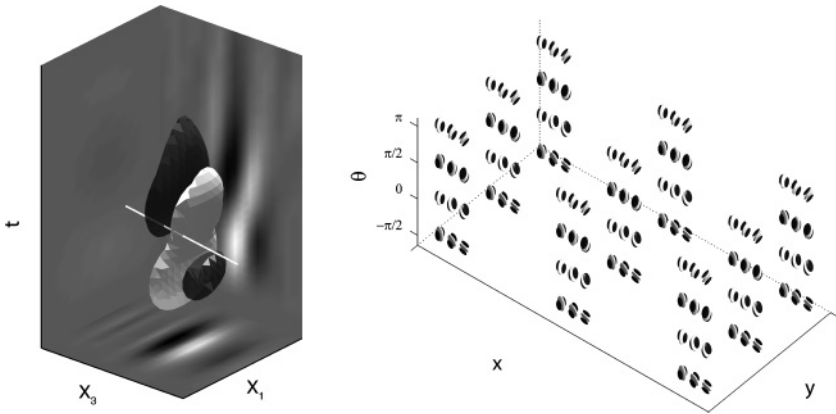


Figure 1: (Left) Isolevel surfaces of a V1 inseparable receptive profile depicted in space-time. Light and dark surfaces enclose, respectively, excitatory and inhibitory regions. The transversal line indicates the direction of the local vector  $\vec{X}_{\theta,v}$  (see the text for details). (Right) Schematization of the feature-wise organization of the primary visual cortex. For each spatiotemporal point  $(x, y, t)$  of the image hyperplane, there is a two-dimensional fiber of representing local orientation  $\theta$  and local velocity  $v$ .

spatiotemporal width. One of the crucial features of equation 2.1 is minimization of the uncertainty of simultaneous measurements in space-time and frequency. It is worth noting that this model strictly captures the features of so-called inseparable RPs, tuned for position, orientation, and motion detection, depicted in Figure 1 (left). Separable RPs can be obtained as linear superpositions.

Further analyses have also shown that the Gabor parameter distribution found in cortical cells covers only a subset of the Gabor family. Such subsets are optimized for the detection of the local features of orientation  $\theta$  and speed  $v$  (Cocci et al., 2012; Barbieri, Citti, & Sarti, 2014),

$$\theta = \arctan\left(\frac{p_y}{p_x}\right),$$

$$v = \frac{p_t}{\sqrt{p_x^2 + p_y^2}},$$

which can be interpreted as fundamental features of the visual stimulus. For this reason, we will not deal with the dependence on the spatial frequency

$\kappa = \sqrt{p_x^2 + p_y^2}$  or on the scales  $\sigma$ , but consider RPs of the form

$$g_{q,\theta}^{\sigma,\kappa}(x, y, t) = e^{-2\pi i \kappa ((x-q_x) \cos \theta + (y-q_y) \sin \theta + v(t-q_t))} e^{-\left(\frac{(x-q_x)^2}{2\sigma_x^2} + \frac{(y-q_y)^2}{2\sigma_y^2} + \frac{(t-q_t)^2}{2\sigma_t^2}\right)}, \quad (2.2)$$

where  $\kappa$  and  $\sigma$  are considered fixed. This corresponds to a neural processing stage where the visual stimulus is lifted from the spatiotemporal image space  $\mathbb{R}^2 \times \mathbb{R}^+$  to the extended five-dimensional feature space,

$$\mathcal{M}_T = \mathbb{R}^2 \times \mathbb{R}^+ \times S^1 \times \mathbb{R}^+ = \{\eta = (q_x, q_y, q_t, \theta, v)\},$$

where every point  $\eta \in \mathcal{M}_T$  corresponds to a filter, equation 2.2, as in Figure 1(right). The activity of V1 simple cells is then considered as modeled by the map

$$f \mapsto F^{\sigma,\kappa}(q, \theta, v) \doteq \langle g_{q,\theta}^{\sigma,\kappa}, f \rangle_{L^2(\mathbb{R}^3)}. \quad (2.3)$$

**2.2 Connectivity as a Differential Constraint.** The functional behavior of V1 simple cells modeled by equation 2.3 can be interpreted as a finite-scale spatiotemporal directional derivative of the stimulus around position  $q$ , performed along the direction

$$\vec{X}_{\theta,v} = (\cos \theta, \sin \theta, -v)$$

expressed in the coordinates  $\{\hat{e}_x, \hat{e}_y, \hat{e}_t\}$ . Accordingly, this derivation is maximal along the direction of the gradient of  $f$ . This implies that the lifting to  $\mathcal{M}_T$  of any smooth level set of  $f$  is always orthogonal to the vector field,

$$\vec{X} = (\cos \theta, \sin \theta, -v, 0, 0) \in \mathcal{T}_\eta \mathcal{M}_T,$$

expressed in the coordinates  $\{\hat{e}_{q_x}, \hat{e}_{q_y}, \hat{e}_{q_t}, \hat{e}_\theta, \hat{e}_v\}$ , where  $\mathcal{T}_\eta \mathcal{M}_T$  stands for the tangent space of  $\mathcal{M}_T$  at  $\eta = (q_x, q_y, q_t, \theta, v)$ .

These constraints induce the consideration as admissible surfaces on  $\mathcal{M}_T$  those whose tangent space at any point is spanned by the vector fields

$$\begin{aligned} \vec{X}_1 &= (-\sin \theta, \cos \theta, 0, 0, 0) & \vec{X}_2 &= (0, 0, 0, 1, 0), \\ \vec{X}_4 &= (0, 0, 0, 0, 1) & \vec{X}_5 &= (v \cos \theta, v \sin \theta, 1, 0, 0), \end{aligned} \quad (2.4)$$

defining the orthogonal complement to  $\vec{X}$  in  $\mathcal{T} \mathcal{M}_T$ .

The four-dimensional hyperplanes generated by  $\{X_1, X_2, X_4, X_5\}$  are called contact planes, the whole structure is named the contact structure. Contact structures have been used for modeling the functional architecture of the visual cortex in other work (Petitot & Tondut, 1999; Citti & Sarti, 2006).

Due to this contact structure, the connectivity among V1 cells on  $\mathcal{M}_T$  can be modeled geometrically as in Barbieri, Citti, Cocci et al. (2014) in terms of advection-diffusion processes along the directions of the vector fields, equation 2.4. Two corresponding stochastic processes were introduced in order to provide concrete realizations of the mechanisms of propagation of information along connections.

A first mechanism, aimed at modeling connectivity along lifted contours of a spatial image at a fixed time, consists of propagation along the direction  $\vec{X}_1$  forced by diffusion over  $\vec{X}_2$  and  $\vec{X}_4$ . This will be used for a single-frame segmentation out of a spatiotemporal streaming. It lives on a codimension 1 submanifold of  $\mathcal{M}_T$  at fixed time  $t_0$  that we will call  $\mathcal{M}_0 = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+ = \{\xi = (q_x, q_y, q_t, \theta, v) : q_t = t_0\}$ . It can be formally described by the following system of stochastic differential equations,

$$\left\{ \begin{array}{l} dq_x(s) = -\sin \theta(s) ds \\ dq_y(s) = \cos \theta(s) ds \\ d\theta(s) = \kappa dW_1(s) \\ dv(s) = \alpha dW_2(s) \\ \xi(s=0) = \xi_0, \end{array} \right. \quad (2.5)$$

where  $W = (W_1, W_2)$  is a two-dimensional Brownian motion and  $(\kappa, \alpha)$  are the corresponding diffusion constants. The Fokker-Plank equation associated with this process, which provides a transition probability density  $\rho_0$ , is

$$(\partial_s - L_0)\rho_0(\xi, s|\xi_0, 0) = \delta(\xi - \xi_0)\delta(s), \quad (2.6)$$

where the evolution operator is given by  $L_0 = X_1 - \kappa^2 X_2^2 - \alpha^2 X_4^2$ ; as customary, we have denoted with  $X_i$  the directional derivative along the vector field  $\vec{X}_i$ .

A second mechanism, aimed at modeling connectivity among moving contours of a spatiotemporal stimulus, will be used for spatiotemporal segmentation of apparent point trajectories. It consists of a propagation along  $\vec{X}_5$ , again forced by a diffusion over  $\vec{X}_2$  and  $\vec{X}_4$ , and is described by the stochastic process on  $\mathcal{M}_T$ :

$$\left\{ \begin{array}{l} dq_x(s) = v \cos \theta(s) ds \\ dq_y(s) = v \sin \theta(s) ds \\ dq_t(s) = ds \\ d\theta(s) = \kappa dW_1(s) \\ dv(s) = \alpha dW_2(s) \\ \eta(s=0) = \eta_0 \end{array} \right. . \quad (2.7)$$

The Fokker-Plank equation associated with this process, which provides transition probability density  $\rho_T$ , is

$$(\partial_s - L)\rho_T(\eta, s|\eta_0, 0) = \delta(\eta - \eta_0)\delta(s), \quad (2.8)$$

where the evolution operator is given by  $L = X_5 - \kappa^2 X_2^2 - \alpha^2 X_4^2$ .

The structure of these processes explicitly assigns different roles to the spatiotemporal variables  $q$ , where the stimulus is defined, and to the engrafted variables  $(\theta, v)$ . More precisely, we have advection in the  $q$  variables, while diffusion occurs in the  $(\theta, v)$  variables. It is worth noting that this construction naturally extends the process proposed by Mumford (1994) for the case of static images, which consists of propagation along the direction  $\vec{X}_1$  forced by diffusion over  $\vec{X}_2$ , on the three-dimensional submanifold  $\mathcal{M}_3 = \mathbb{R}^2 \times S^1 = \{\zeta = (q_x, q_y, \theta)\}$ :

$$\left\{ \begin{array}{l} dq_x(s) = -\sin \theta(s) ds \\ dq_y(s) = \cos \theta(s) ds \\ d\theta(s) = \kappa dW_1(s) \\ \zeta(s=0) = \xi_0. \end{array} \right. . \quad (2.9)$$

The Fokker-Plank equation associated with this process, which provides a transition probability density  $\rho_3$ , is

$$(\partial_s - L_3)\rho_3(\zeta, s|\zeta_0, 0) = \delta(\zeta - \zeta_0)\delta(s), \quad (2.10)$$

where the evolution operator is given by  $L_3 = X_1 - \kappa^2 X_2^2$ .

Each of equations 2.6, 2.8, and 2.10 is defined by a Markov generator  $\mathcal{L}$  of a stochastic process over a manifold  $X$ , with transition probability  $\varrho$  satisfying

$$(\partial_s - \mathcal{L})\varrho(x, s|x_0, 0) = \delta(x - x_0)\delta(s). \quad (2.11)$$

Our aim is to use the density  $\varrho$  to define a connectivity kernel over the manifold  $X$ , disregarding the dynamics over the evolution parameter  $s$ .



This connectivity can be obtained by integrating  $\varrho$  over  $s$ , against some appropriately chosen weight  $p$ , hence defining a connectivity kernel as

$$\Gamma(x, x_0) = \int_0^\infty \varrho(x, s|x_0, 0) p(s) ds.$$

When  $p$  is a positive weight normalized to 1, this kernel has the well-known probabilistic interpretation of being the transition probability for a stochastic process subordinated to a time process with independent increments distributed with  $p$  (see Sato, 1999). One common choice for the weight  $p$  is that of Mumford (1994) and Williams and Jacobs (1995): an exponential decay representing a decay of the signal during the propagation (for a quantitative analysis, see Bosking et al., 1997), which amounts to replacing  $\varrho$  with its Laplace transform at a fixed timescale. Another possible choice is the one used in Sanguinetti et al. (2010) and Barbieri, Citti, Cocci et al. (2014), where  $p$  was identically set to 1, which amounts to considering the density of points reached at any value of the evolution parameter. In this work, we deal with an intermediate choice between these two: we will choose a weight depending on an evolution scale parameter  $H$  as

$$p(s) = \frac{1}{H} \chi_{[0,H]}(s) = \begin{cases} \frac{1}{H} & s \in [0, H] \\ 0 & \text{otherwise} \end{cases}.$$

This choice amounts to assigning a uniform weight to all values of the evolution parameter, but allows as to keep track of the evolution length over which the stochastic paths are evaluated. Our notion of a connectivity kernel will then be

$$\Gamma^H(x, x_0) = \frac{1}{H} \int_0^H \varrho(x, s|x_0, 0) ds. \quad (2.12)$$

It is worth noting that both the diffusion parameter  $\kappa$  and the evolution scale parameter  $H$  could be treated in principle as additional fiber variables of the model, linking their different associated connectivities to some prominent feature of the image detected by the visual cortex (e.g., curvature and scale). For model simplicity, though, in this letter, we treat these variables as parameters, addressing the extension of the model in future work.

**2.3 Discrete Connectivity Kernels.** In this section, we outline the numerical method we have used to compute the connectivity kernels  $\Gamma^H$ , which are needed due to the lack of analytic solutions. A notable exception is the  $SE(2)$  case, equation 2.10, for which an analytic solution was obtained in Duits and Van Almsick (2008) and compared in Zhang, Duits, and

Romeny (2014) to several numerical approaches, all of them different from ours. Our choice of numerics has privileged the robustness and flexibility of the implementation, which has been applied to all three cases.

The differential equations of type 2.11 that we need to solve originate from a stochastic process over  $X$ ,  $\mathcal{M}_0$ ,  $\mathcal{M}_T$ , or  $\mathcal{M}_3$ , that we can write in terms of its sample paths  $\gamma : \mathbb{R}^+ \rightarrow X$  as

$$d\gamma(s) = A(\gamma)ds + B(\gamma)dW,$$

where  $A$  is a vector field and  $B$  is a matrix field over  $X$  representing, respectively, equations 2.5, 2.7, and 2.9. A flexible and efficient numerical technique is then that of Markov chain Monte Carlo methods (MCMC) (see Graham & Talay, 2013), which was implemented as follows. We first fix a discretization over the parameter  $s$ , using without loss of generality a step  $\Delta s = 1$  so that the discrete evolution will be performed over  $\mathbb{N}$ , and a discrete covering grid  $\{\Omega_j\}_{j \in \mathbb{N}}$  of  $X$  (i.e., a collection of subsets of  $X$  satisfying  $\Omega_i \cap \Omega_j = \emptyset$  if  $i \neq j$  and  $\bigcup \Omega_j = X$ ). For a given  $\gamma_0 \in X$ , we then simulate  $N$  several discrete-time random paths over  $X$  with the recursive equation

$$\gamma_{h+1} = \gamma_h + A(\gamma_h) + B(\gamma_h)\delta_h, \quad h \in \mathbb{N},$$

where  $\{\delta_h\}_{h \in \mathbb{N}}$  are (vector valued) independent and identically (i.i.d.) gaussian random variables, and assign to each region  $\Omega_j$  a value between 0 and 1 corresponding to the number of paths that passed through it divided by  $N$ . This provides a distribution over the cells  $\Omega_j$  that, up to a multiplicative constant, for large values of  $N$ , gives a discrete approximation of the solution to equation 2.11 that we will denote  $\varrho(\Omega_j, h|\gamma_0, 0)$ . The resulting discrete approximation of the connectivity kernel, equation 2.12, will then be computed as

$$\hat{\Gamma}^H(\Omega_j|\gamma_0) = \frac{1}{H} \sum_{h=0}^H \varrho(\Omega_j, h|\gamma_0, 0). \quad (2.13)$$

A deeper discussion of this kind of numerical approximation for stochastic differential equations can be found in Platen (1999).

During the simulations of section 4, we vary the parameters that characterize the kernels. In order to keep track of their dependency, we then use the notation  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  for the discrete approximations of the connectivity kernels in the cases, respectively, of equations 2.5 and 2.7, where  $\kappa$  and  $\alpha$  stand for the diffusion coefficients over  $\theta$  and  $v$ , and the notation  $\hat{\Gamma}_3^{H,\kappa}$  for the discrete approximation of the connectivity kernel in equation 2.9, where  $\kappa$  is the diffusion coefficient over  $\theta$ .

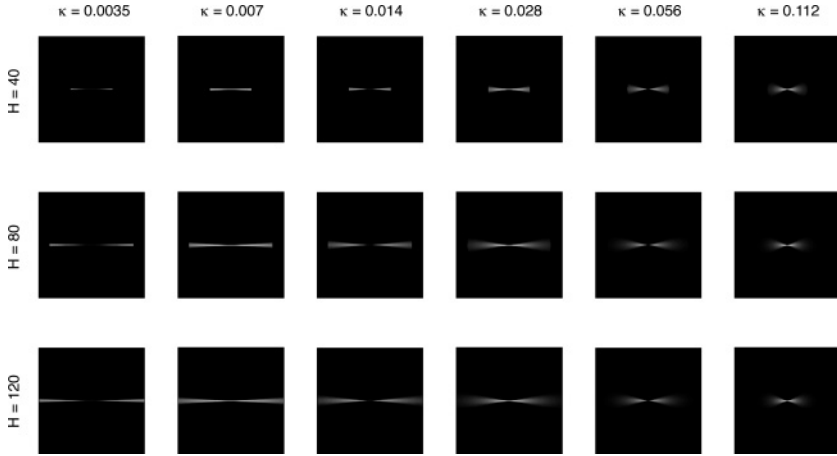


Figure 2: Marginal distributions over the  $(q_x, q_y)$  plane of the kernel  $\hat{\Gamma}_3^{H,\kappa}$  computed for different values of the parameters  $\kappa$  and  $H$ .

The effect of the variation of the parameter  $\kappa$  on the shape of  $\hat{\Gamma}_3^{H,\kappa}$  is shown in Figure 2; in Figure 3, we show the isosurfaces of two projections of the connectivity kernels  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  together with the related horizontal curves, that is, the curves obtained from systems 2.5 and 2.7 by substituting noise with a constant and varying the parameters  $\kappa$  and  $\alpha$ . These curves were introduced in Barbieri, Citti, Cocci et al. (2014) as the geometric counterpart, in the space-time contact structure, to the horizontal curves used in Citti and Sarti (2006) to model the orientation association fields. They provide a natural geometric extension of the notion of association fields to spatiotemporal stimuli with orientation and velocity features. As can be seen, the related kernels reach their maximum values in the proximity of the fan of such curves originating from the same starting point, which motivates their role as a model for a spatiotemporal neural connectivity.

### 3 Spectral Analysis of Connectivities

In this section, we recall the main notions to be used about clustering with spectral analysis on graphs and how to use the previously introduced geometric setting for these purposes. The task we address falls into the well-known class of problems broadly referred to as dimensionality reduction, which deals with the general problems of data partitioning and locality-preserving embeddings of high-dimensional data sets. The literature on these topics is huge, and we have chosen to refer only to work directly related to the one we present here, referring readers to the references in these other works.

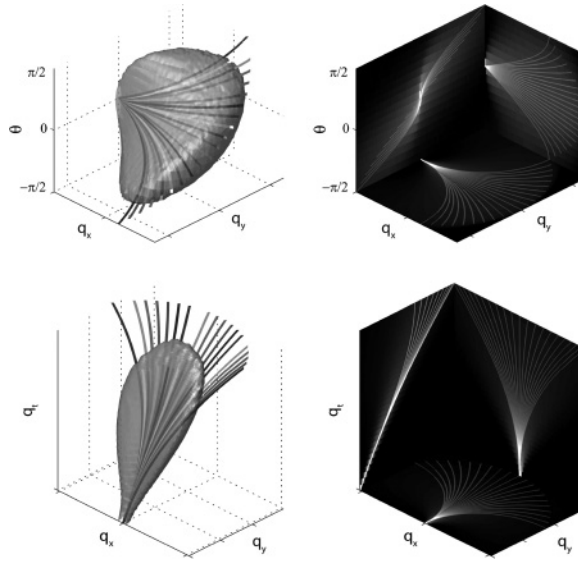


Figure 3: Stochastic kernels  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  (top) and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  (bottom) for  $N = 10^6$  stochastic paths, compared to the horizontal curves calculated as in Barbieri, Citti, Coci et al. (2014). (Left) Isosurface plot of the kernels. (Right) Kernel projections relative to the variables  $(q_x, q_y, \theta)$  for  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  and  $(q_x, q_y, q_t)$  for  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  (gray), under the projections of the horizontal curves (white).

The cognitive task of spatial or spatiotemporal visual grouping can be interpreted as a form of clustering. Two examples of spatial visual stimuli providing standard clustering problems are portrayed in Figure 4. On the left, three dense gaussian distributions of two-dimensional points are embedded within a sparser set of random points uniformly scattered throughout the domain. The human visual system normally segregates the points of the gaussian clouds into three separate groups of points (objects) lying on a noisy environment. On the right, two dashed continuous lines are embedded in a field of segments having random position and orientation. In this case, stimulus collinearity gives rise to a pop-out effect that makes the two lines easily distinguishable from the background, the phenomenon quantified by the psychophysical experiments of Field et al. (1993).

Although these examples show two quite different grouping effects, a common underlying mechanism can be formalized as follows. Given a data set of  $n$  points  $S = \{x_i\}_{i=1}^n \subset X$  living in an arbitrary metric space  $(X, d)$ , the task of grouping together the points that—according to the distance  $d$ —are closer, or more similar to each other (so that their ensemble forms an object), amounts to identifying  $K$  disjoint subsets  $S_i \subset S$ , with  $i \in \{1, \dots, K\}$ , where  $K - 1$  of them contain points relatively close to each other and relatively far

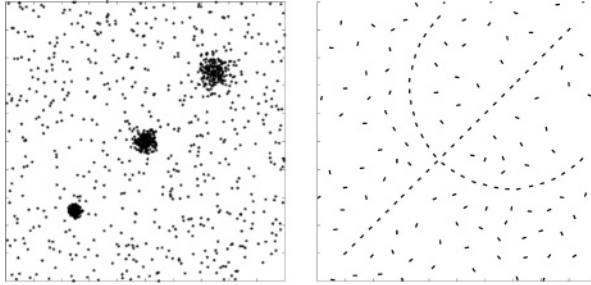


Figure 4: Two sample spatial visual stimuli containing perceptual units embedded in random point fields.

from the rest and one of them contains points that cannot be classified in such a way and will be considered as noise.

It is widely known that this problem is not easily resolvable with purely clustering algorithms like K-means (see Ng, Jordan, & Weiss, 2002). A major branch of research has been recently devoted to the development of spectral techniques that allow addressing these issues in terms of the spectral properties of symmetric positive semidefinite affinity matrices constructed from the input data set. The ensemble of these techniques can generally be subdivided into two classes (Lafon & Lee, 2006): methods for locality-preserving embeddings of large data sets that project the data points onto the eigenspaces of the affinity matrices (Roweis & Saul, 2000; Belkin & Niyogi, 2003; Coifman & Lafon, 2006) and methods for data segregation and partitioning that basically perform an additional clustering step, taking as input the projected data set (Perona & Freeman, 1998; Weiss, 1999; Shi & Malik, 2000; Meila & Shi, 2001).

In section 3.1, we describe the spectral clustering algorithm we use to perform visual grouping, and in section 3.2, we show how to use the geometric feature spaces and the cortical connectivities developed in section 2 for constructing affinity matrices associated with spatial and spatiotemporal visual stimuli. We apply this method in section 4 to several stimuli.

**3.1 Spectral Clustering.** Let us consider the data set  $S = \{x_i\}_{i=1}^n$  in the metric space  $(X, d)$  as the vertices of a weighted graph, where the edge weights  $\{a_{ij}\}_{i,j=1}^n$  define an affinity matrix  $A$ . Perona and Freeman (1998) showed that when  $A$  is a real symmetric matrix, its first eigenvector can serve as an indicator vector for basic grouping purposes. They also proposed a partitioning algorithm that recursively separates the foreground information from the data set. While this algorithm's implementation is straightforward and efficient, it can easily lead to clustering errors due to noise, nonlinear distributions, or outliers (Weiss, 1999). Perona and Freeman's

argument was later improved, also in view of its relation to several other problems such as that of minimal graph cuts (Shi & Malik, 2000), by performing the spectral analysis of a suitably normalized affinity matrix (see Von Luxburg, 2007). We use the normalization proposed in Meila and Shi (2001), which turns a real symmetric affinity matrix  $A$  into the transition matrix  $P$  of a reversible Markov chain via row-wise normalization. More precisely, if  $D$  is the diagonal degree matrix, having elements

$$d_i = \sum_{j=1}^n a_{ij},$$

the normalized affinity matrix  $P$  is given by

$$P = D^{-1}A. \quad (3.1)$$

This matrix in general will not be symmetric, but it can be shown that its eigenvalues  $\{\lambda_j\}_{j=1}^n$  are real and satisfy  $0 \leq \lambda_j \leq 1$ , and its eigenvectors  $\{u_j\}_{j=1}^n$  accordingly can be chosen with real components. The clustering properties of the eigenvectors of  $P$  can be clearly understood in the following ideal case. Suppose that the graph  $G = (S, A)$ , with nodes given by  $S$  and edge weights given by  $A$ , has  $K$  connected components  $\{G_i\}_{i=1}^K$ , and that all the elements of each component have the same edge weight connecting them. The resulting normalized affinity matrix  $P$  would then be a block diagonal matrix, with only  $K$  nonnull eigenvalues  $\{\lambda_i\}_{i=1}^K$ , each of them equal to 1, whose corresponding eigenvectors  $\{u_i\}_{i=1}^K$  are piece-wise constant indicator functions of the partitions.

In real applications, the affinity matrices are perturbed versions of the block diagonal ones and do not possess an ideal binary spectrum with purely indicator eigenvalues, thus making the partitioning problem generally illposed. The best situation one can hope for is that of a good approximation of the ideal case, where the affinity on the data set is such that there are clusters of points that are strongly connected mainly to their cluster neighbors and only weakly connected to the rest. In general, several authors demonstrated that the eigenvectors  $\{u_i\}_{i=1}^K$  of the normalized affinity matrix  $P$ , corresponding to the  $K$  largest eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$ , solve the relaxed optimization problem of normalized graph cuts and give a nice probabilistic interpretation of the clustering problem.

A crucial step in real applications is that of choosing the value of  $K$ , that is, deciding how many eigenvalues are worth taking into consideration and consequently how many eigenvectors possess relevant clustering information. Many authors have proposed different solutions, for example, looking for the maximum eigengap or trying to minimize a particular cost function (see Zelnik-Manor & Perona, 2004). We decided to adopt a

## Algorithm 1: Spectral Clustering Algorithm.

- 
1. Build the affinity matrix  $A$  on an appropriate connectivity measure
  2. Compute the normalized affinity matrix  $P = D^{-1}A$
  3. Solve the eigenvalue problem  $Pu_i = \lambda_i u_i$ , where the order of indices  $i = 1, \dots, n$  is such that  $\{\lambda_i\}_{i=1}^n$  is decreasing
  4. Fix a threshold  $\epsilon$  and a diffusion parameter  $\tau$
  5. Define  $q = q(P; \epsilon, \tau)$  as the largest integer such that  $\lambda_i^\tau > 1 - \epsilon$  for all  $i = 1 \dots q$
  6. Define the preclusters  $\{\hat{C}_k\}_{k=1}^q \subset \mathbb{R}^d$  so that for any  $i = 1, \dots, n$ , the point  $x_i$  belongs to  $\hat{C}_k$  whenever
 
$$k = \arg \max_{j \in \{1, \dots, q\}} \{u_j(i)\}$$
  7. Fix a minimum cluster size  $M$ , join together the preclusters with less than  $M$  elements into the cluster  $C_0$ , and order the remaining preclusters so to obtain a partition of the data set into  $\{C_k\}_{k=1}^K \subset \mathbb{R}^d$ , where  $K = K(P; \epsilon, \tau, M)$ .
- 

semisupervised solution consisting of fixing an a priori significance threshold  $\epsilon$ , and consider as clustering eigenvectors all those  $u_i$  whose  $\lambda_i > 1 - \epsilon$ .

Since the more  $P$  is far from being similar to a block diagonal matrix, the more its spectrum will be far from being dichotomous, with the ordered  $\lambda$ 's decreasing more smoothly (see Figure 5), the sensitivity to small changes on the values of  $\epsilon$  may become very high in the more ill-posed cases. In order to facilitate this delicate passage, we have used a technique suggested by the well-known diffusion map approach (Coifman, Maggioni, Zucker, & Kevrekidis, 2005; Coifman & Lafon, 2006; Lafon & Lee, 2006): we have introduced an auxiliary thresholding integer parameter  $\tau$  and have evaluated the exponentiated spectrum  $\{\lambda_i^\tau\}_{i=1}^n$ , which for sufficiently large values of  $\tau$  is closer to dichotomy, against the threshold  $\epsilon$ . This exponentiated spectrum has an easy probabilistic interpretation, being the spectrum of the matrix  $P^\tau$ . Indeed, due to the normalization, equation 3.1,  $P$  can be seen as the Markov transition matrix of a random walk over the graph, so that  $P^\tau$  represents the transition probability of the same random walk in  $\tau$  steps.

Once the number  $K$  of eigenvectors to use has been selected, we have used a variation of a simple clustering technique proposed in Kannan et al. (2004) in order to extract the clustering information. It corresponds, in the notation of these authors, to the clustering of the rows of  $P$  based on its reduced matrix of eigenvectors  $\hat{U} = [u_1, \dots, u_q]$ . The main differences are that once the thresholding parameters are fixed, this algorithm dynamically assigns the number of (pre)clusters, and that a notion of background is explicitly introduced in the last step. Algorithm 1 is our spectral clustering algorithm.

As a first application of algorithm 1, consider the two data sets presented in Figure 4 as sets of points in  $\mathbb{R}^2$  endowed with the isotropic Euclidean distance  $d(x, y) = \|x - y\|$ , and define the affinity matrix

$$a_{ij} = e^{-\frac{d(x_i, x_j)^2}{2\sigma^2}}, \quad (3.2)$$

where  $\sigma$  is a scale parameter that has to be chosen based on the characteristics of the data set to be clustered. The result is two connected graphs where the similarity between vertices decays as a gaussian with their Euclidean distance. This is intuitively suitable to describe the visual clustering of the first data set, but does not take into consideration the kind of similarity that characterizes the visual grouping of the second one. The results of the application of the spectral clustering algorithm with affinity given by equation 3.2 are shown in Figure 5. In the first case, the algorithm performs the clustering process correctly, automatically finding the number of the main perceptual units and assigning the remaining elements to the background cluster. It is worth noting that the spectrum of the normalized affinity matrix  $P^r$  counts many eigenvalues that are close to 1, each representing only a single perceptual unit, that are mostly composed of a few elements. The segment data set was not clustered correctly; this was predictable, as these kind of stimuli are characterized by a local feature of orientation and hence should be better considered on the position-orientation domain  $\mathbb{R}^2 \times S^1$ , together with an anisotropic affinity.

*3.1.1 On Possible Neural Implementations of the Algorithm.* Besides the purely computational presentation that we have given of our spectral clustering algorithm, it may be relevant to consider whether plausible neural computations exist that could be responsible for its implementation in the visual cortex.

The discussion in section 2 and the work of Barbieri, Citti, Cocci et al. (2014) on the psychophysical and physiological nature of the connectivity kernels constitute our main motivation to consider the first step of the algorithm as a cortical implementation of affinities in the feature spaces. Regarding the anatomical implementation, these kernels may be implemented via long-range connections in a single area, as in Ben-Shahar and Zucker (2004). However, they may also be compatible with an effective connectivity due to the integration of different areas (see Angelucci et al., 2002).

For what concerns the normalization step 2, we observe that this is formally analogous to the one introduced in Tononi, Sporns, and Edelman (1994), which is generally motivated by evidence that biological neural systems tend to adjust the weight of afferent connections so that a neuron with few incoming connections will weight those inputs more heavily than a neuron with many incoming connections (see also the discussion in (Barnett, Buckley, & Bullock, 2009)). However, these effects mainly refer to long-term



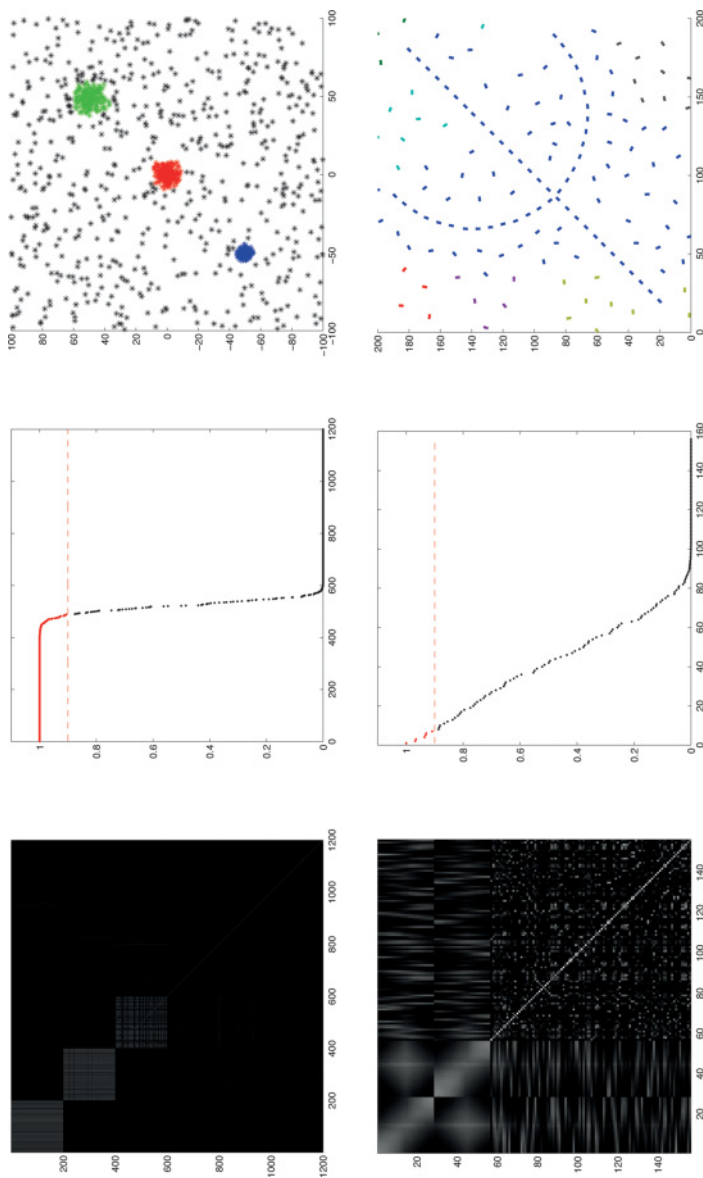


Figure 5: Result of the proposed algorithm for the two example data sets. The picture shows how a Euclidean isotropic kernel is an optimal choice to cluster groups of points living in the image plane  $\mathbb{R}^2$ , but the results are inadequate when trying to separate boundaries or contours, whose additional feature of local orientation suggests considering them as points on the contact manifold  $\mathcal{M}_3 = \mathbb{R}^2 \times S^1$ .

learning processes and involve all afferent connections, while our step 2 is related only to the ones that are activated by stimuli. Short-term adaptive effects providing the remaining normalization could be due to dendritic interactions (see Mel & Schiller, 2004, or London & Häusser, 2005).

With respect to step 3, previous work such as Bressloff, Cowan, Golubitsky, Thomas, and Wiener (2002) and Faugeras, Veltz, and Grimbert (2009) describes a possible implementation of the spectral analysis as a mean field neural computation, obtaining the emergence of eigenvectors of the connectivity as a symmetry breaking in the evolution equation associated with sufficiently high eigenvalues. An extension of this work can be found in Sarti and Citti (2014), who show that in the presence of a visual input, the emerging eigenvectors correspond to visual perceptual units, which are obtained from a spectral clustering on excited connectivity kernels. According to these models, when the magnitude of an eigenvalue trespasses a given threshold, its corresponding eigenvector becomes a locally unstable solution, hence generating a distinguishable activity pattern. This threshold is what we aim to reproduce with steps 4 and 5.

Coming to step 6, we first note that its combinatorial formulation represents a concrete implementation of the following principle: eigenvectors with high eigenvalues represent perceptual units, and a point of the visual stimulus is assigned the unit whose eigenvector component over such a point has the higher magnitude. As we will discuss in detail, such a clustering will perform better the more the affinity matrix is close to a block diagonal matrix, since in such cases, the eigenvectors relative to high eigenvalues will be close to indicator functions of weakly overlapping regions of the stimulus. In these situations, the concrete neural implementation of steps 3, 4, and 5 already includes our step 6. Indeed, the computation of an almost-indicator eigenvector that is an unstable solution to the neural field equation means that the underlying cells have been selected and participate in the associated activity pattern. We also observe at this point that in this normalized model, all sufficiently high eigenvalues have very close magnitudes that in the ideal case is 1 for all of them. If a neural field model with two populations is considered, the magnitude of the eigenvalue is associated with the frequency of oscillation of the neural population, which in this situation is almost the same for all perceptual units. This aspect marks a substantial difference with other nonnormalized models of spectral neural computations. Indeed the binding process associated with perception (Gray & Singer, 1989) shows coherent activity in a single frequency range (the gamma range), a behavior that can be reproduced by our model, where all the eigenvalues associated with perceptual units are almost the same.

The aim of step 7 is to measure the salience of the perceptual units and introduce a thresholding on it. The notion of salience that we consider is proportional to the total neural activity involved in the detection of a perceptual

unit that, in the special artificial cases under consideration, corresponds to the number of related elements of the data set.

**3.2 Cortical Affinity Matrices.** As we described in section 2, a spatial or a spatiotemporal visual stimulus is represented by different classes of V1 specialized neurons on different kinds of higher-dimensional feature spaces, and we have focused on three of them that we have called  $\mathcal{M}_0$ ,  $\mathcal{M}_T$ , and  $\mathcal{M}_3$ . The neural response, equation 2.3, is in general non-null on the whole feature space, but according to the stronger or weaker presence of the locally preferred feature, the magnitude of the response will be higher or lower. Here we will adopt a great simplification and consider the response magnitude to be allowed to take only value 0 or 1, so that the data sets we will deal with will be purely point data sets embedded in a feature space. The main advantage of this simplification is the possibility of testing the grouping capabilities of the geometric connectivities in the feature spaces without dealing with the many delicate questions related to the stimulus representation itself. Only a few classes of stimuli can be represented in such a way. For this reason, we work only with synthetic stimuli and generate the corresponding data sets directly in the feature space. This seems to be the safest way to separate the problems related to the geometric properties of the proposed connectivity from the ones related to the good representation of stimuli onto the feature space.

Let  $X$  be a feature space, such as  $\mathcal{M}_0$ ,  $\mathcal{M}_T$ , or  $\mathcal{M}_3$ ;  $\{\Omega_j\}_{j \in \mathbb{N}}$  be a discrete covering such as the one introduced in section 2.3; and  $S = \{x_i\}_{i=1}^n \subset X$  be the data set representing the visual stimulus on this higher-dimensional space, being of type  $S_0 = \{\xi_i\}_{i=1}^n \subset \mathcal{M}_0$ ,  $S_T = \{\xi_i\}_{i=1}^n \subset \mathcal{M}_T$  or  $S_3 = \{\zeta_i\}_{i=1}^n \subset \mathcal{M}_3$ . Our aim is to define an affinity that makes two points  $x_i$  and  $x_j$  more similar the higher is their geometric connectivity, computed as in equation 2.12 with  $\hat{\Gamma}^H(x_i, x_j)$ . A necessary compatibility condition between the data set and the discretization of the space is that each  $\Omega_j$  contains at most one point  $x_i$  of the data set. This indeed makes the kernels well defined on the data set and allows us to write  $\hat{\Gamma}^H(x_i, x_j)$ . However, since the associated Markov generator  $\mathcal{L}$  is not self-adjoint, this kernel is not symmetric. We then need to construct symmetric affinities from these connectivities or, rather, adapt the previously described theoretical setting to couple with this asymmetry. We will choose one or the other way depending on the kernel and the geometry of the information carried.

**3.2.1 Reciprocal Connectivities for Spatially Distributed Features.** A reasonable neural assumption when modeling long-range horizontal connection among V1 cells on  $\mathcal{M}_3$  or on  $\mathcal{M}_0$ , which refer to the spatially distributed features of local orientations and local velocities, is that these connections are reciprocal (Kisvarday & Eysel, 1992). This means that two cells that are

selective for the features  $(\theta_1, v_1)$  and  $(\theta_2, v_2)$  are assumed to be symmetrically connected. We then take as affinities  $A^H = (a_{ij}^H)$  a symmetrization of the corresponding connectivity kernels. The symmetrization we have chosen consists of taking the Hermitian component of the kernels:

$$a_{ij}^H = \frac{\hat{\Gamma}^H(x_i, x_j) + \hat{\Gamma}^H(x_j, x_i)}{2}. \quad (3.3)$$

This symmetrization has a specific geometric meaning, which can be easily understood for the kernel  $\hat{\Gamma}_3^{H,\kappa}$ : it is equivalent to sum the fundamental solution of equation 2.10 with the fundamental solution of the same operator under an angular shift of 180 degrees in the variable  $\theta$ . Such rotation turns the drift term  $X_1$  into  $-X_1$ , hence transforming the forward Kolmogorov equation into the corresponding backward equation, so the sum of the two solutions is clearly symmetric. That sum allows us to identify angles up to 180 degrees, hence turning a process that was a priori defined over the group  $SE(2)$  of positions and angles into a process properly defined on positions and orientations. For this reason, such symmetrization is customary when dealing with such nuclei with the purpose of describing pure orientation (see Sanguinetti et al., 2010). We observe that this symmetrization does not modify the single cell response of modeled simple cells, which still detect angles, but rather introduces a symmetry in their geometric connectivities so that a cell having a preferred angle is considered to be long range connected to other cells along the two angles corresponding to the associated orientation. The same argument applies to the kernel  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ , whose Fokker-Planck operator differs only for an additional diffusion term on the velocities.

Depending on the application to clustering with the kernels  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  or  $\hat{\Gamma}_3^{H,\kappa}$  we will then obtain from equation 3.3 different affinity matrices over data sets  $\{\xi_j\} \subset \mathcal{M}_0$  or  $\{\zeta_i\} \subset \mathcal{M}_3$ . The associated spectral clustering will then depend on the parameters  $(\epsilon, \tau)$  and  $M$  from the algorithm and on the parameters  $H, \kappa$ , and  $\alpha$  from the kernels.

The grouping capabilities of such anisotropic affinities, and the role of the parameters, may be first evaluated on the simplest connectivity  $\hat{\Gamma}_3^{H,\kappa}$ . In order to do so, we have applied the spectral clustering algorithm with the cortical affinity to the second stimulus of Figure 4. Such stimulus was indeed considered as a data set in  $\mathcal{M}_3 = \mathbb{R}^2 \times S^1$ , each segment having a position in  $\mathbb{R}^2$  determined by its center and a position in  $S^1$  corresponding to its orientation. The results obtained are displayed in Figure 6 for different sets of kernel parameters  $H$  and  $\kappa$ . The clustering parameters used here and in the examples of the next sections were  $\epsilon = 0.05$ ,  $\tau = 150$ , and  $M = 3$ . We have kept those parameters fixed for all the tests in order to focus on the differences in the grouping properties due to the connectivity kernels.

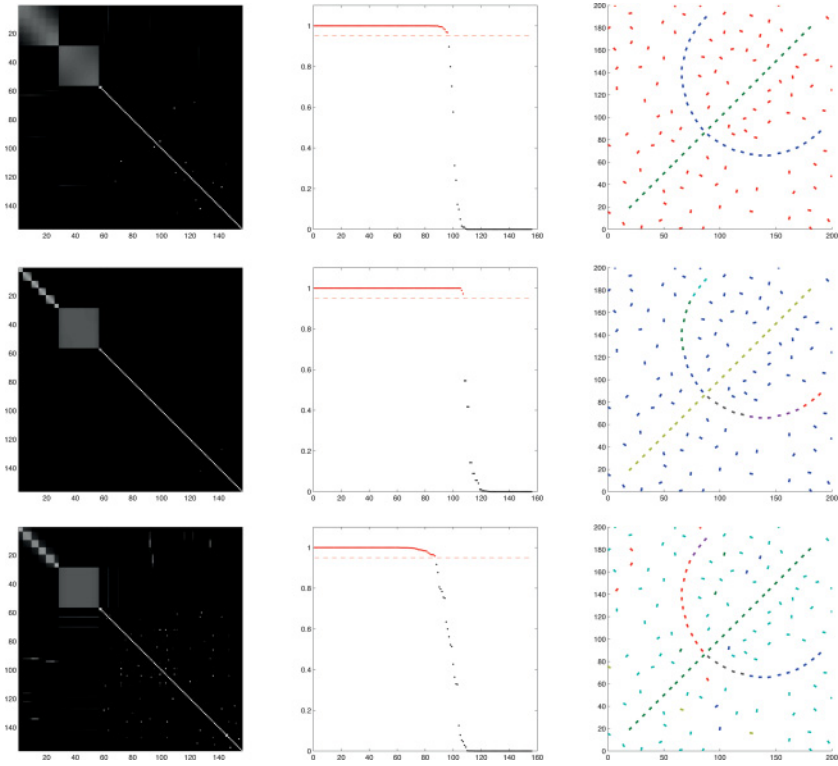


Figure 6: Result of the proposed algorithm for the second example data set and different parameters for the kernel  $\hat{\Gamma}_3^{H,\kappa}$ . Columns and shadings organized as in Figure 5. Different kernel parameters modify the look of the affinity matrix, its spectrum, and the resulting data set partitioning.

As we can see, the presence of an affinity that is more semantically and geometrically adapted to the data set with respect to equation 3.2 positively influences the grouping capabilities of the method. However, this first simulation already shows the relevance of the kernel parameters on the quality of the grouping.

For the top plots of Figure 6, we have used an evolution integration step of  $H = 40$  and an orientation diffusion coefficient  $\kappa = 0.014$ , which coincides with the curvature of the semicircular object. The algorithm clearly succeeds in distinguishing the two perceptual units from each other and correctly assigns the remaining elements to the same background and noise partition. For the middle plots, we have reduced the value of the diffusion constant to  $\kappa = 0.0035$ : while the algorithm correctly retrieves the straight contour and distinguishes the units from the background, the semicircle

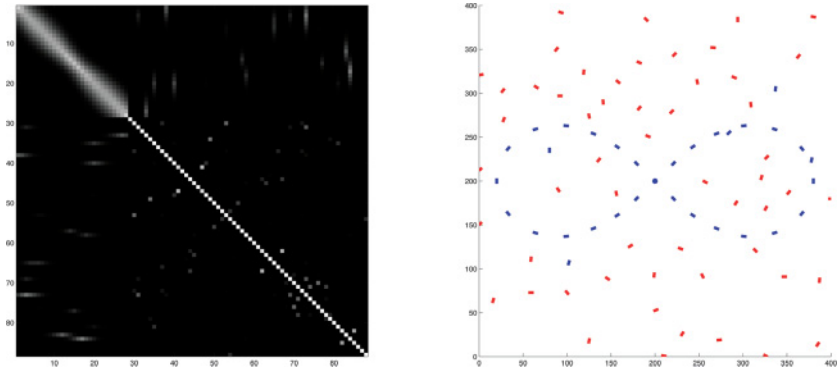


Figure 7: Affinity matrix (left) and grouping results (right) of a lemniscate with random noise. The segments pertaining to the lemniscate are assigned to the same group (shown in blue), while the others are assigned to the background partition (shown in red).

gets overpartitioned. As in the previous case, the affinity matrix is close to being a block diagonal matrix, but in this case, the partitioned connected components of the subgraph containing the two objects are more than two. Again setting the diffusion coefficient to the original value, for the bottom plot, we have increased the evolution integration step value to  $H = 100$ . As the stimulus domain is  $200 \times 200$  pixels wide, this means that every segment in the example could potentially have a non-null connectivity value with almost half of the other segments if the cocircularity conditions are satisfied. Indeed, by observing the resulting affinity matrix  $A_3^{H=100, \kappa=0.014}$ , we can see high affinity values between the objects and the random noise and in between the background elements. Again, even if the straight line is correctly retrieved as a single object, two random elements, approximately cocircular with the beginning of the line, are incorrectly interpreted as being part of the object. Furthermore, the semicircular contour again gets over-segmented, and many of the randomly collinear points, very far from each other, are interpreted as being a perceptual unit.

We stress that these kernels do not prevent contours having multiple orientations in the same spatial position to be recognized as one object. For example, Figure 7 show, a lemniscate embedded in a field of randomly oriented segments. On the lifted space where the kernel lives, this eight-figure contour is indeed continuous and nonoverlapping, so that the grouping algorithm will assign all of its elements to the same perceptual unit, as human vision would tend to do.

**3.2.2 Time Asymmetry: A Directed Graphs Approach.** In order to set an affinity over a space-time data set  $S_T = \{\eta_i\}_{i=1}^n \subset \mathcal{M}_T$ , one must consider

the intrinsic notion of causality associated with the time component of  $\eta$ , which, from the neural point of view, represents the mean activation time of a V1 cell. This causality is reflected in the kernel  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  so that any symmetrization would compromise the information content carried by such kernels and negatively affect the efficiency of the grouping algorithm. We then need to work with a nonsymmetric affinity matrix  $A_T^{H,\kappa,\alpha}$ , hence defining a directed graph, given by

$$(a_T^{H,\kappa,\alpha})_{ij} = \hat{\Gamma}_T^{H,\kappa,\alpha}(\eta_i, \eta_j). \quad (3.4)$$

In order to cope with this asymmetry, we modify some of the spectral clustering criteria. We still first normalize the affinity matrix as in equation 3.1. This produces a transition probability matrix  $P_T$ , which, however, is associated with a random walk that does not satisfy the detailed balance condition because the Markov chain is no longer reversible. Its eigenvalues  $\{\lambda_i\}_{i=1}^n$  and eigenvectors  $\{u_i\}_{i=1}^n$  will then in general be complex valued. In order to perform spectral clustering in this situation, we follow the approach introduced in Pentney and Meila (2005), which consists concretely in replacing the eigenvectors  $\{u_i\}$  with vectors obtained by the sum of their real and imaginary parts  $\{u_i^+ = \Re u_i + \Im u_i\}$  and defining their clustering strength in terms of the square modulus of their eigenvalues, hence performing the thresholding step 5 of algorithm 1 with respect to the set of real numbers  $\{|\lambda_i|^2\}_{i=1}^n$ .

This is by far not the only possible choice. In particular, this choice will not preserve the interpretation of minimal graph cuts; rather, it will produce clusters with the property of having elements with approximately the same outward transition probability. Indeed, as Meila and Pentney (2007) discussed, the minimal cuts clustering and the probabilistic clustering in general have different solutions for directed graphs, which sets a marked difference with the undirected case associated with symmetric affinities.

The choice of working with the couples  $\{(u_i^+, |\lambda_i|^2)\}$ , which are not in general eigenvalue-eigenvector pairs of a Hermitian matrix, can be better understood as follows. We can see that the eigenvalue problem associated with the real  $n \times n$  matrix  $P = D^{-1}A$  for a nonsymmetric affinity  $A$  can be restated equivalently as an eigenvalue problem with real eigenvectors by doubling the dimension of the space and replacing  $P$  by  $\tilde{P} = \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix}$ . In a more formal way, let

$$Pu_\lambda = \lambda u_\lambda, \quad P\bar{u}_\lambda = \bar{\lambda} \bar{u}_\lambda, \quad (3.5)$$

and call  $\lambda = x + iy$ . We introduce the following notation: set  $\vec{u}_\lambda = \begin{pmatrix} u_\lambda \\ \bar{u}_\lambda \end{pmatrix} \in \mathbb{C}^{2n}$ , denote with  $\tilde{P} = \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix} \in \mathbb{I}_{2 \times 2} \otimes P$ , and also let  $\Lambda_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix} \otimes \mathbb{I}_{n \times n}$ ,

where  $\otimes$  is the usual Kronecker product and  $\mathbb{I}_{n \times n}$  stands for the  $n \times n$  identity matrix. Then solving problem 3.5 for  $u_\lambda \in \mathbb{C}^n$  is equivalent to solving

$$\tilde{P}\vec{u}_\lambda = \Lambda_\lambda \vec{u}_\lambda \quad (3.6)$$

for  $\vec{u}_\lambda \in \{(z_1, z_2) \in \mathbb{C}^{2n} : z_2 = \overline{z_1}\}$ .

Consider now the auxiliary real vectors  $\vec{v}_\lambda = \begin{pmatrix} u_\lambda^+ \\ u_\lambda^- \end{pmatrix} = \begin{pmatrix} \Re u_\lambda + \Im u_\lambda \\ \Re u_\lambda - \Im u_\lambda \end{pmatrix}$ . If we introduce the matrix  $B = \begin{pmatrix} \frac{1-i}{2} & \frac{1+i}{2} \\ \frac{1+i}{2} & \frac{1-i}{2} \end{pmatrix} \otimes \mathbb{I}_{n \times n}$ , we have that  $\vec{v}_\lambda = B\vec{u}_\lambda$ , and we can rewrite equation 3.6 as the real system for  $\vec{v}_\lambda \in \mathbb{R}^{2n}$ ,

$$\tilde{P}\vec{v}_\lambda = \Sigma_\lambda \vec{v}_\lambda, \quad (3.7)$$

where  $\Sigma_\lambda = \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \otimes \mathbb{I}_{n \times n} = B\Lambda_\lambda B^{-1} \in \mathbb{R}^{2n \times 2n}$ . The magnitude of the action of  $\tilde{P}$  over  $\vec{v}_\lambda$  is given by  $\det \Sigma_\lambda = |\lambda|^2$ , which provides a natural clustering parameter to threshold. Moreover, since the matrix  $\tilde{P}$  is a double copy of the matrix  $P$ , in order to cluster its rows, it is sufficient to work on a half-dimensional space and choose, without loss of information, only the  $u_\lambda^+$  component of  $\vec{v}_\lambda$ .

#### 4 Visual Grouping with Cortical Affinities

---

Several phenomenological findings indicate that the grouping properties obtained by spatial collinearity can easily be broken if one associates a speed and an orthogonal direction of movement to each oriented segment. Limits have been found on the maximum rate of change of local speed along a contour that makes possible the perception of boundaries and shapes. Indeed, a random speed distribution over a dashed line could completely destroy the perception of a single unit as a whole while enhancing the impression of different segments pertaining to the random background field. These observations have led to the notion of motion contour in Rainville and Wilson (2005), where it is shown by psychophysical experiments how the brain groups features together while also relying on the local speed perpendicular to their orientation axis, with coherent velocities being represented by velocity fields that vary smoothly over space. These authors expanded the already known notions that local stimulus velocity is discernible (thus determinant for grouping purposes) only when it is orthogonal to the perceived contour or is not part of a trajectory (Hess & Ledgeway, 2003; Ledgeway et al., 2005; Vergheze & McKee, 2006). Coherently, the analysis carried out in Cocci et al. (2012) over a data set of cortical neurons in the primary visual cortex showed how the spatiotemporal shape of their RPs is biased to



optimally measure the local stimulus orthogonal velocity. Thus, it may be inferred that stimulus local direction of movement and speed are additional features driving the spatial integration involved in the perception of shapes and contours.

The simulations performed in this section deal with the geometric connectivities constructed in section 2, which aim to model the neural connections in the cortical area V1 of the visual cortex (Barbieri, Citti, Cocci et al., 2014). These connectivities are defined in feature spaces that include the local orthogonal velocities, which are detected by V1 specialized cells.

**4.1 Grouping with Spatial Features.** The perceptual bias toward collinear stimuli has classically been associated with the long-range horizontal connections linking cells in V1 having similar preferences in stimulus orientation. This specialized form of intrastriate connectivity pattern is found across many species, including cats (Kisvarday, Toth, Rausch, & Eysel, 1997), tree shrews (Chisum, Mooser, & Fitzpatrick, 2003), and primates (Angelucci et al., 2002), the main difference being the specificity and the spatial extent of the connections. Furthermore, axons seem to follow the retinotopic cortical map anisotropically, with the axis of anisotropy being related to the orientation tuning of the originating cell (Bosking et al., 1997). The clustering algorithm with the affinity matrix  $A_3^{H,\kappa}$ , constructed as in equation 3.3 using the kernel  $\hat{\Gamma}_3^{H,\kappa}$ , may then be neurally motivated by the assumption that contour completion and visual grouping are strongly influenced by the mutual position and local orientation of the elements in the visual space. Other prominent features of the visual stimulus, namely, its velocity and direction of motion, seem to play an important role in the spatial integration of oriented elements (Rainville & Wilson, 2005). In addition to stimulus orientation, cells in the striate areas are selective for the direction of motion orthogonal to the cell's preferred orientation (DeAngelis et al., 1995; Barbieri, Citti, & Sarti, 2014). These selectivities are also structurally mapped in the cortical surface, with nearby neurons being tuned for similar motion direction (Weliky, Bosking, & Fitzpatrick, 1996), and it has been shown that excitatory horizontal connections in the V1 of the ferret are strictly iso-direction-tuned (Roerig & Kao, 1999). In order to model the grouping effect of these connectivities, we make use of the affinity matrix  $A_0^{H,\kappa,\alpha}$  constructed with the extended kernel  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ .

**4.1.1 The Stimuli.** We have considered a data set made of  $n$  points  $\{\zeta_i\}_{i=1}^n$  in the feature space of positions and orientations  $\mathcal{M}_3 = \mathbb{R}^2 \times S^1$  having the following structure. Two perceptual units consisting of segments aligned along circle arcs with curvature  $k$  are embedded in a background environment of  $r$  segment having random positions and orientations. The data set will be denoted as  $S_r^k$  and is depicted in Figure 8 (top).

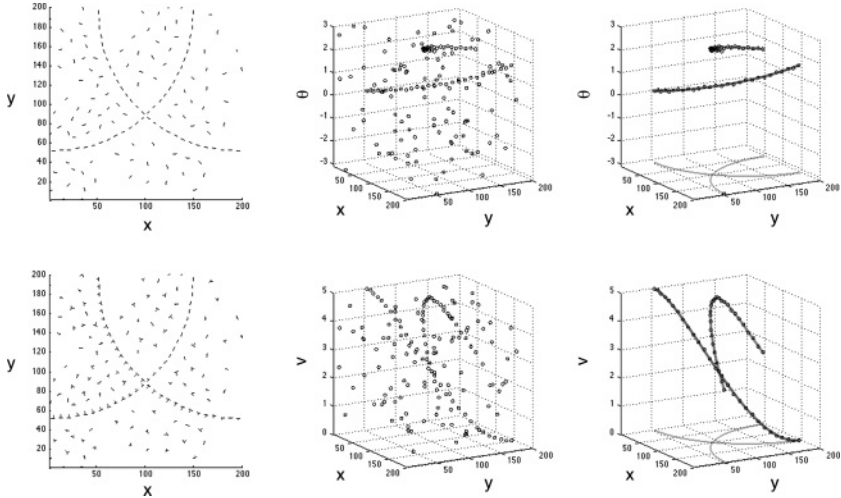


Figure 8: An instance of the data set  $S_r^k$ , where the left column shows the corresponding visual stimulus, the middle column shows the data set in the feature space, and the right column shows the perceptual units, which constitute the target segmentation objects, which in this instance have curvature  $k = 0.023$ . The top row shows the data set in the feature space  $\mathcal{M}_3 = \mathbb{R}^2 \times S^1$ , and the bottom row shows it in the feature space  $\mathcal{M}_0 = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$  with a velocity assignment producing a shape deformation. The bottom left image contains the spatial stimulus together with the magnitude of its instantaneous orthogonal velocity (drawn as arrows).

By assigning to each point  $\zeta_i$  an instantaneous orthogonal velocity  $v_i$ , the data set can be considered as  $n$  points  $\{\xi_i = (\zeta_i, v_i)\}_{i=1}^n$  embedded in the larger feature space  $\mathcal{M}_0 = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ . A velocity field that is constant on each perceptual unit would describe a rigid motion, and the analysis of such a case with affinity  $A_0^{H,\kappa,\alpha}$  can be easily understood as a direct extension of the static case, clearly providing grouping improvement. A more interesting case is that of a velocity field that changes along each perceptual unit, which represents shape deformation. We then generated a distribution of the velocity feature in the following way: given a fixed maximal velocity  $V$ , we assigned to points belonging to the perceptual units a velocity feature  $v_i$  that varies sinusoidally along the arc circle, passing from zero to  $V$ , and assigned a random velocity between 0 and  $V$  to the background points. The data set  $\{\xi_i\}_{i=1}^n \subset \mathcal{M}_0$  is depicted in Figure 8 (bottom).

**4.1.2 The Numerical Experiments.** We have applied the method described in section 3 using the same spectral clustering parameters as for the examples of Figure 6— $\epsilon = 0.05$ ,  $\tau = 150$ , and  $M = 3$ —on a set of stimuli  $S_r^k$  for

different values of  $k$  and  $r$ . We used both the affinity  $A_3^{H,\kappa}$  on  $\mathcal{M}_0 = \mathbb{R}^2 \times S^1$  and  $A_0^{H,\kappa,\alpha}$  on  $\mathcal{M}_0 = \mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ , in this case setting the maximum velocity assignment to  $V = 5$ . We then evaluated the grouping performances for various kernel parameter sets  $(H, \kappa)$  and  $(H, \kappa, \alpha)$  by computing, for each iteration, a percentage error measure,

$$E = \frac{E_1 + E_2 + E_3}{n},$$

where  $n$  is the total number of points in the stimulus,  $E_1$  is the number of points incorrectly assigned to the noise or background set,  $E_2$  is the number of random points incorrectly recognized as part of a perceptual unit, and  $E_3$  is the points pertaining to an over- or underpartitioned contour. In order to correctly compare results obtained by partitioning the data set with a different number of random points, we averaged  $E$  over 100 repetitions, where at each repetition, we changed the random part of the stimulus and calculated new kernels  $\hat{\Gamma}_3^{H,\kappa}$  and  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ . The resulting mean percentage error measure  $\hat{E}$  has been taken as a measure of the quality of the grouping.

**4.1.3 The Results.** The results of the experiments are shown in Figures 9 and 10.

In Figure 9a, we represent with a grayscale intensity the error  $\hat{E}$  for the spectral clustering of  $S_{120}^{0.056}$  with the affinity  $A_3^{H,\kappa}$  in  $\mathbb{R}^2 \times S^1$  by varying the parameters  $(H, \kappa)$  of the kernel. The parameters  $(k, r) = (0.056, 120)$  correspond to the stimulus in the feature space of positions and orientations composed of the perceptual units with the highest curvature and having the highest number of random elements among all those tested in our analysis. A first significant feature that emerged is that the set of kernel parameters that gives the lowest grouping error value is  $\kappa = 0.056$ ,  $H = 20$ , that is, the same curvature value of the contours in the stimulus and short stochastic path length. Moreover, we observe that by maintaining the kernel parameter  $H$  set to its minimum value and decreasing the kernel diffusion coefficient  $\kappa$ , we have an approximately constant higher error value  $\hat{E}$ , whose dominating components are  $E_1$  and  $E_3$ . This indicates that a reduction in the width of the fan of stochastic curves generating  $\Gamma_3$  impairs the connectivity between high curvature contour elements, so that the algorithm perceives them as separate units and assigns part or all of them to the background set. It is also worth noting that regardless of the kernel diffusion coefficient, increasing the parameter  $H$  has a negative impact on the quality of the grouping, this time mainly because of the error component  $E_2$ . This was predictable because the longest stochastic path lengths can generate affinities between elements very distant from each other. A high number of random elements can in this case induce the algorithm to recognize them as part of a distant contour. This gives worse effects when both  $\kappa$  and  $H$  have high values,

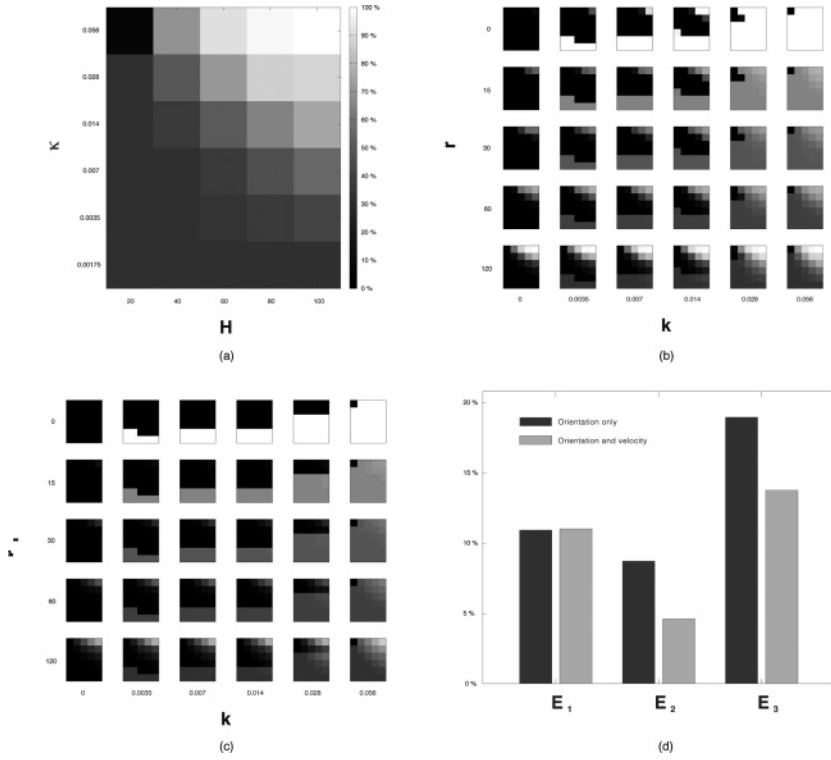


Figure 9: Parametric analysis for visual grouping in  $\mathbb{R}^2 \times S^1$  and  $\mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ . The intensity of the shading each region is proportional to the percentage of misinterpreted points averaged over 100 repetitions. (a) Grouping results for the stimulus  $S_{120}^{0.056}$  in  $\mathbb{R}^2 \times S^1$  by varying the kernel parameters  $(\kappa, H)$ . (b) Grouping results as in panel a for the stimuli  $S_r^k$  in  $\mathbb{R}^2 \times S^1$  by varying the stimulus parameters  $(k, r)$ . (c) Grouping results for the stimuli  $S_r^k$  in  $\mathbb{R}^2 \times S^1 \times \mathbb{R}^+$  with nonuniform velocity assignments. (d) Comparison of the grouping performances in position-orientation versus position-orientation-velocity with separated analysis of the three error sources.

because with the associated kernel, a point in  $\mathcal{M}_3$  can potentially be connected to distant elements having a very different orientation. In such cases, the error component  $E_3$  makes a contribution, as the two perceptual units can be underpartitioned and interpreted as one unique object, each one having reciprocally affine contour elements.

Figure 9b resumes the same kind of grouping analysis carried out with 30 different stimuli  $S_r^k \in \mathcal{M}_3$  by varying the parameters  $k$  and  $r$ . We notice a positive correlation between the stimulus contour curvature  $k$  and the

smallest value of the kernel diffusion coefficient  $\kappa$  that provides optimal grouping. Smaller values of  $\kappa$  mostly lead to overpartitioned or unrecognized contours, while higher  $\kappa$ 's generally do not impair the contour grouping capability of the algorithm, even if we have observed that the error component  $E_2$  tends to grow together with the number of random elements in the space  $r$  and the length of the stochastic paths  $H$ . In general, thus, in order to correctly identify contour-like objects, the algorithm should be based on a connectivity whose diffusion coefficient is above the contour's curvature.

These results suggest that for a successful grouping, the parameters  $\kappa$  and  $H$  could be associated with some property of the image detected by the visual cortex. In the discrete case, for example, the model could be extended by adding two fiber variables, local stimulus curvature and scale, governing the numerical value taken by the two connectivity parameters (August & Zucker, 2003). One could also argue—and our results are consistent with this view—that curvature and scale are very close concepts, strongly influencing each other, so that  $\kappa$  and  $H$  should not be independent. A deeper study of these aspects will be addressed in a future paper.

In Figure 9c, we show the grouping results obtained by analyzing the stimuli  $S_r^k$  embedded in  $\mathcal{M}_0$ , with the described nonuniform velocity assignment, using the connectivity  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ . From a first visual inspection, we can see that the correlation constraint between the contour curvature  $k$  and the orientation diffusion coefficient  $\kappa$  is still present. However, the error at the highest values of  $\kappa$  and  $H$  is significantly reduced, if not almost completely eliminated, for the stimuli having fewer random elements due to the influence on the algorithm of correlations on the direction of motion. We have not presented the detailed analysis on the parameter  $\alpha$ , which is very similar to the one related to  $\kappa$ . We limit ourselves to the observation that the best results were obtained for values of  $\alpha$  close to  $\bar{\alpha} = \frac{\pi}{L}V$ , where  $L$  is the length of the perceptual unit (circle arc) and  $V$  is the maximum velocity assigned, which was set to  $V = 5$ . The reason can be explained in terms of the velocity assignment, which we chose to be sinusoidally varying along the contour, namely,

$$v(\phi) = V \sin\left(\frac{\pi\phi}{L}\right) + \text{const.},$$

where  $\phi \in [0, L]$  is the arc parameter of the circular curves. Then  $\bar{\alpha} = \max |\dot{v}|$ , which is a diffusion constant that ensures that grouping capabilities will be preserved even at the maximum local speed change rate along the contour. In particular, the results shown refer to clustering with the parameter  $\alpha$  set to  $\bar{\alpha}$ .

A comparison of the grouping quality obtained with the orientation connectivity  $\hat{\Gamma}_3^{H,\kappa}$  and the orientation and velocity connectivity  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ , with

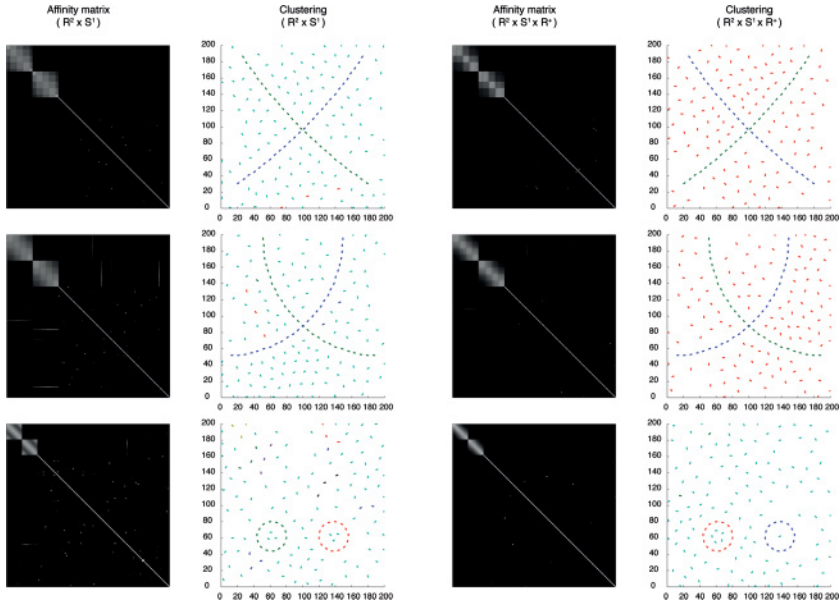


Figure 10: Results of the proposed algorithm for a data set with perceptual units with different curvatures. (Left column) Results obtained by using the position-orientation connectivity given by  $\Gamma_3$ . (Right column) Results obtained by using the position-orientation-velocity connectivity given by  $\Gamma_0$ . Affinity matrices built with  $\Gamma_0$  are cleaner than those built with  $\Gamma_3$ ; they generally avoid spurious affinities between perceptual units and noise while maintaining the approximate object block diagonal structure.

separate presentation of the three error sources considered, is shown in Figure 9d. The histograms correspond to the average of the three error components  $E_1$ ,  $E_2$ , and  $E_3$  over all stimuli, the kernel parameters, and the repetitions. The analysis of  $E_1$  shows that both kernels tend to confuse the perceptual units with noise in approximately the same way. However, from the analysis of  $E_2$ , we can see that information on local velocity significantly enhances the performance of the algorithm in the presence of noise, improving the assignment of the random elements to the background. Also, the analysis of  $E_3$  shows that the connectivity  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  reduces the underpartitioning effect of the stimulus, which typically occurs when the parameters  $(\kappa, H)$  make  $\hat{\Gamma}_3^{H,\kappa}$  too long range and widespread.

Finally, some particular cases of the discussed grouping simulations are shown in Figure 10, where the use of  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  instead of  $\hat{\Gamma}_3^{H,\kappa}$  concurs in reducing different kinds of grouping errors of the algorithm.

**4.2 Spatiotemporal Grouping.** In this section, we consider grouping performed in space and time. Similar to what happens for the integration of spatial visual information, our visual system is capable of easily predicting stimulus trajectories and grouping together elements having similar motion or apparent motion paths. At the psychophysical level, the facilitation in detecting moving stimuli, given a previous cue with a coherent trajectory, is found to be significantly higher than the one expected from the temporal response summation given by the onset and offset dynamics of classical RPs (Verghese et al., 1999). One possible explanation could be the existence of a specialized facilitatory network linking cells anisotropically and coherently with their direction of motion.

Ledgeway and Hess (2002) studied the perception of spatial contours defined by nonoriented stimuli moving coherently and tangentially along a path. They suggested rules similar to the ones driving facilitation in position and orientation and hinted a possible role played by a trajectory-specialized network. More possible evidence of a trajectory-driven connectivity comes from a recent study of the dynamics of neural population response to a sudden change of motion direction, where it is shown that for low angular changes, a nonlinear part of the response provides a sort of spatiotemporal interpolation (Wu, Tiesinga, Tucker, Mitroff, & Fitzpatrick, 2011).

We recall also that at the physiological level, although the basic mechanisms of velocity detection are already present in V1, the estimation of stimulus motion has been classically associated with neurons in visual area MT/V5. There, cells with high selectivity in direction of movement and extended sensitivities to a wide range of stimulus velocities are indeed present (Maunsell & Van Essen, 1983; Rodman & Albright, 1987). Extrastriate areas are retinotopically organized, with anisotropic and asymmetric connectivity bundles reaching columns of cells tuned for similar orientation and direction preference (Malach, Schirman, Harel, Tootell, & Malonek, 1997).

On the other hand, the important role played by V1 collinear horizontal connectivity in motion perception has been suggested by Series, Georges, Lorenceau, and Frégnac (2002), after showing that perception of speed is biased by the direction of motion. Also, recent results in Pavan et al. (2011) showed how trajectories of oriented segments are significantly more detectable for orientations orthogonal to the path of motion, thus supporting the hypothesis of two different facilitatory mechanisms (Hess & Ledgeway, 2003).

**4.2.1 The Stimulus.** The data set we used in this simulation is depicted in Figure 11. We created a set of points  $\eta_i = (x_i, y_i, t_i, \theta_i, v_i) \in \mathcal{M}_T = \mathbb{R}^2 \times \mathbb{R}^+ \times S^1 \times \mathbb{R}^+$ , represented in the figures by segments moving in time for  $n_t = 32$  frames. Three perceptual units, a circular shape of curvature  $k = 0.02$  and two bars, move as rigid bodies, with the circle translating with speed  $V_{\text{circ}} = 7.5$  spatial units/frame in the opposite direction to the bars, which both move with speed  $V_{\text{bars}} = 3.75$  spatial units/frame. We recall that



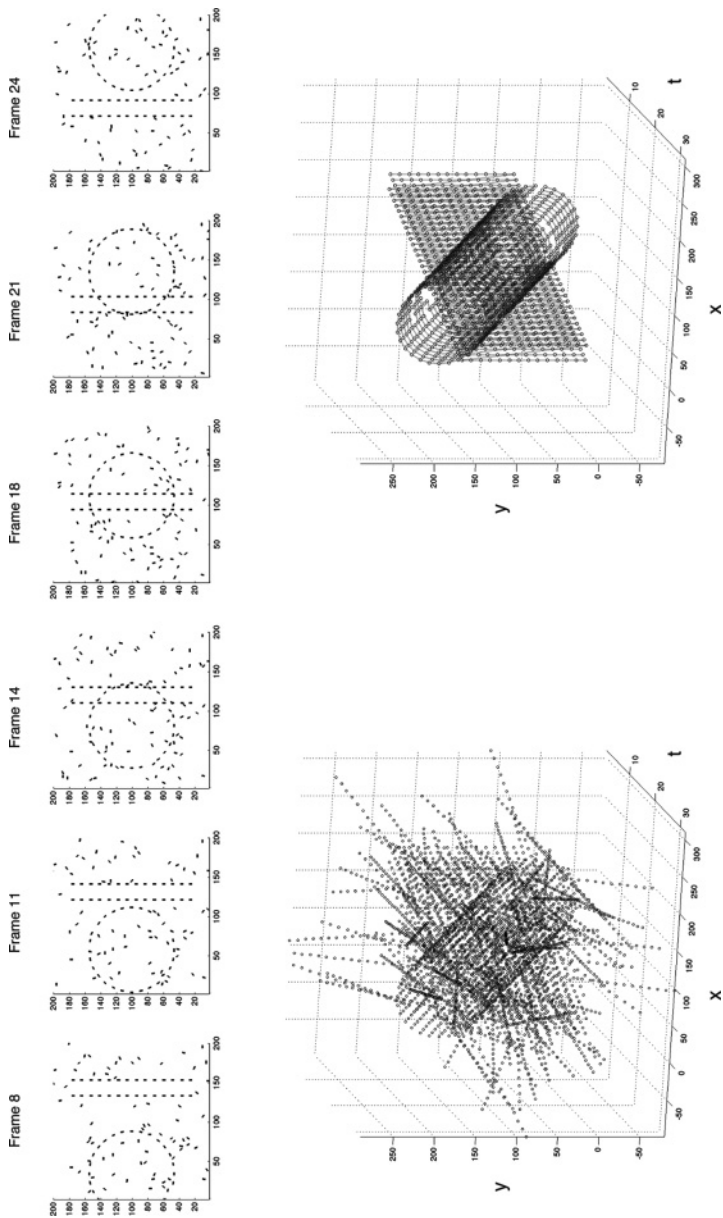


Figure 11: An instantiation of the stimulus living in  $\mathbb{R}^2 \times \mathbb{R}^+ \times S^1 \times \mathbb{R}^+$  used in the tests for spatiotemporal grouping of moving shapes. The circle has curvature  $k = 0.02$ , and its velocity is 7.5 units per frame, that is, twice that of the bars (3.75 units per frame).



the fiber coordinate of local velocities  $v_i$  of each point represents just the projection onto  $\bar{X}_3$ , that is, orthogonal to the segment orientation, of the real velocity vector. Similar to what we did in the previous example, these perceptual units were embedded into a background consisting of a variable number  $r$  of random elements, each having a uniform motion path along its  $\bar{X}_3$  direction during all the stimulus frames.

**4.2.2 The Numerical Experiment.** The aim of the grouping algorithm that we use for this experiment is to carry out a segmentation of the full spatiotemporal surfaces representing the moving objects.

From the point of view of the structure of the stimulus that gives rise to the data set, this represents a composite task. It consists of a visual grouping at the spatial level, which identifies the objects, and a visual grouping at the spatiotemporal level, where each previously identified cluster is recognized as constituting the same object during its movement. It is then reasonable to assume that the two connectivity mechanisms  $\hat{\Gamma}_0^{H,\kappa,\alpha}$ , modeling the interactions between points of a motion contour, and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$ , which models motion integration of point trajectories, combine in the clustering of spatiotemporal perceptual units.

The presence of more than one grouping law governing the detection of contours, with different underlying implementing structures, was experimentally confirmed at the psychophysical level in Ledgeway et al. (2005). Moreover, the composition of the different mechanisms that resulted is compatible with a probabilistic summation.

Guided by such arguments, we then performed the spectral clustering on this data set with a matrix  $P$  obtained as the sum of the transition probability (normalized affinity) matrices obtained from the cortical connectivities  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$ . More precisely, we constructed the symmetric matrix  $A_0$ , based on  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  as in equation 3.3, with the additional condition of setting zero affinity between points having different temporal coordinates, and we have constructed the nonsymmetric matrix  $A_T$ , based on  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  as in equation 3.4. We then normalized both of them as in equation 3.1, obtaining the transition probabilities  $P_0$  and  $P_T$ , and we defined the combined spatiotemporal normalized affinity as

$$P = \frac{P_0 + P_T}{2}. \quad (4.1)$$

From the neural point of view, we observe that by relying on the normalized affinity, equation 4.1, we are implicitly assuming a much faster propagation along the connectivity defined by  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  with respect to the temporal dynamics. More precisely, by assigning zero  $A_0$  affinity to temporally separated points of the data set, we are considering spatial connections that fully

operate at each single frame, which corresponds to an almost instantaneous velocity detection and a high horizontal transmission speed.

Finally, in order to perform the spectral clustering over  $P$ , we use the approach described in section 3.2.2. Indeed, since  $A_T$  is not symmetric,  $P$  in general will have complex eigenvalues and eigenvectors, but it was constructed in such a way to keep a probabilistic structure.

**4.2.3 The Results.** The results obtained for various instances of the circle and bars stimulus by varying the number of background elements  $r$  are shown in Figure 12.

The parameters chosen to run the algorithm were  $\epsilon = 0.01$ ,  $\tau = 150$ , and  $M = 3$ , where a smaller threshold  $\epsilon$  with respect to the previous experiments is suggested by the presence of a semantically and geometrically sharper affinity. The integration parameter was set to  $H = 40$  for both  $\Gamma_0$  and  $\Gamma_T$ , according to the quality indications of the previous experiment. Similarly, we have set the angular diffusion parameters to the same value  $\kappa = 0.014$  for both kernels.

With respect to the diffusion coefficient over velocities, we are showing the results for values that for both kernels stay close to the optimal value for the circle discussed in section 4.2.2, that is,  $\bar{\alpha} = \frac{kV_{\text{circ}}}{2} = 0.75$ . Namely, we have set  $\alpha = 0.5$  for  $\Gamma_0$  and  $\alpha = 1$  for  $\Gamma_T$ . In contrast to all other parameters, which were chosen to be equal for both kernels, this one is indeed observed to perform better when it is larger on the temporal connectivity. Such behavior is coherent with the considerations made in Hess and Ledgeway (2003) and Verghese and McKee (2006), where the effect of local changes in the velocity of a motion contour over the perception of visual units is studied. The results of this work show that such changes tend to weaken the visual grouping, but this effect is much stronger when changes are orthogonal to contours with respect to changes that are tangential to trajectories. We are performing visual grouping at both the level of motion contours, with  $\Gamma_0$ , and of trajectories, with  $\Gamma_T$ , describing two copresent connectivities, and the parameter  $\alpha$  describes the sensitivity of the corresponding kernel to local changes in the velocity, so the different levels of  $\alpha$  reflect a higher sensitivity (lower diffusion) to velocity for the  $\Gamma_0$  connectivity and a lower sensitivity (higher diffusion) for the  $\Gamma_T$  connectivity.

From Figure 12 it is possible to see how spectral clustering with the composite affinity, equation 4.1, performs in recognizing the spatiotemporal surfaces relative to the moving circle and the bars with different noise levels. It is successful in clustering the perceptual units, separating clearly their boundaries from the background and between themselves, up to relatively high noise levels. When the number  $r$  of random elements was lower than about 50% of the total, we obtain correct clustering, but for higher noise values, the algorithm began to give poor grouping results, as in the case of  $r = 100$  random segments, which correspond to about 64% of the total.

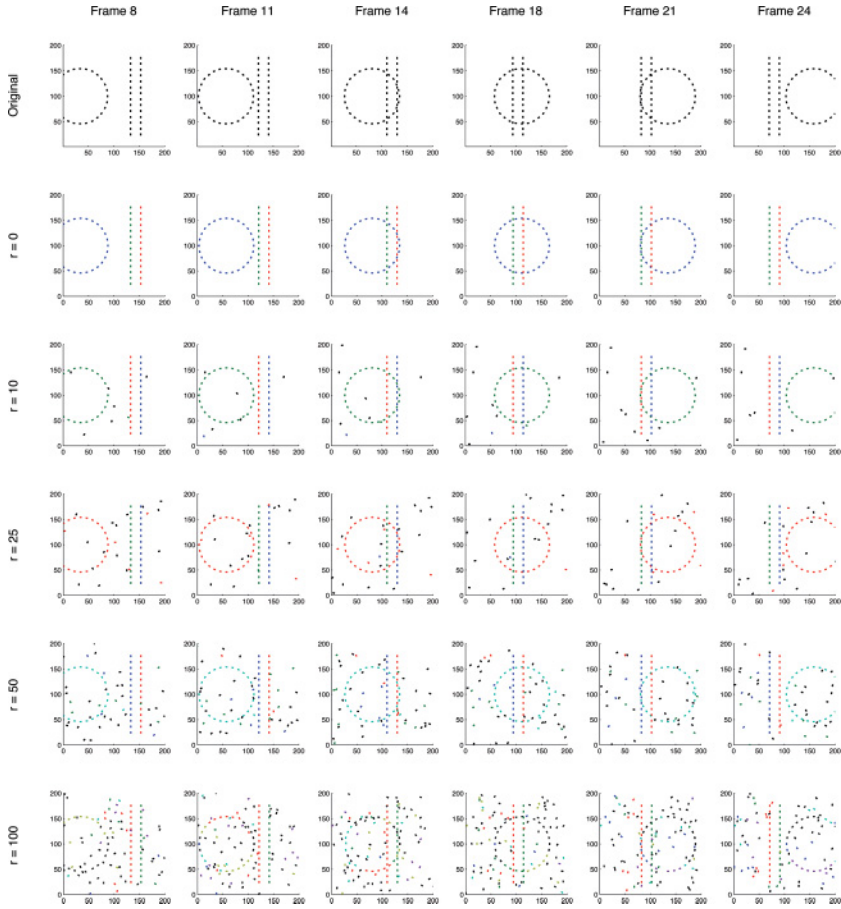


Figure 12: Results obtained by using both connectivities  $\hat{\Gamma}_0^{H,\kappa,\alpha}$  and  $\hat{\Gamma}_T^{H,\kappa,\alpha}$  simultaneously. Different groups are displayed in different colors; the segments assigned to the background partition are drawn in black. The grouping is successful for stimuli with noise levels up to 50%. At a higher noise level, the algorithm fails by overpartitioning the contours of the moving circle. The third row shows grouping results with  $r = 50$  background elements, corresponding to about 47% of the total. The fourth row shows grouping results with  $r = 100$  background elements, corresponding to about 64% of the total.

It is worth noting that even at the higher noise values, the bars are always correctly retrieved. The algorithm tends to fail in detecting contours with high curvature, confusing them with the background segments and thus leading to overpartitioning.

In general, though, we show that the connectivity kernels defined by the proposed cortical-inspired geometrical model applied to a simple spectral clustering algorithm is able to carry out a nontrivial grouping task. To better understand the powerful mechanics involved in the calculations, consider that the only segments of the circle that present a positive affinity value with their corresponding points at future temporal positions are the ones having an orientation value near  $\pm \frac{\pi}{2}$ , as only for them the vector field  $\vec{X}_5$  of connectivity propagation has the same direction of the global movement of the shape.

In fact, while the ability and the reliability of visual neurons in areas V1 and MT/V5 in measuring local stimulus orientation and speed have been studied extensively, the majority of cells in those areas respond solely to the local characteristics they are tuned for. In doing so, the measurements available in the first stages of the visual cortex are subject to the well-known aperture problem: with no information other than the local direction of movement, not much can be said about the direction and speed of the object to which that local measurement refers. For a continuously moving contour, for example, classical orientation- and direction-selective cortical cells measure, for each position along the contour, only the velocity component that is orthogonal to the contour tangent direction at that point. In the framework presented throughout this letter, this is modeled so that the fiber variable of local velocity refers to movements in the  $\vec{X}_3$  direction.

## 5 Conclusion

---

In this work we have constructed a clustering algorithm for visual grouping of spatiotemporal stimuli in terms of geometric connectivity kernels associated with the functional architecture of the visual cortex. The main purpose was to test the segmentation capabilities of such a geometric model of low-level vision areas with respect to spectral analysis mechanisms. Previous experimental investigations such as those of Ledgeway et al. (2005) have already indicated that spatiotemporal perceptual associations play an important role in visual recognition tasks. The recent results in Sarti and Citti (2014) also suggest the presence of concrete cortical implementations of a spectral analysis associated with lateral connection.

We used recent dimensionality reduction methods (Kannan et al., 2004), chosen for their robustness and their relatively simple structure, which allowed us to focus primarily on kernel properties. We then performed several spectral clusterings on the introduced spatiotemporal cortical feature space of position, time, orientation, and velocity by using anisotropic affinities obtained by the models of cortical connections. Such affinities present a structure that is geometrically adapted to the stimuli considered and proved to be able to better extract the relevant information compared to more classical kernels such as the isotropic gaussian one. In particular, the proposed algorithm is capable of grouping together elements belonging to a single

contour or shape moving in time, forming a spatiotemporal surface, and distinguishing them from a noisy background.

The first analysis that we carried out considered visual grouping when the affinity between points of a visual stimulus is assigned in the cortical feature space of positions and local orientations  $\mathbb{R}^2 \times S^1$  and in the feature space  $\mathbb{R}^2 \times S^1 \times \mathbb{R}^+$  taking locally detected velocities into account. The algorithm showed generally higher segmentation capabilities with respect to isotropic affinities defined only on the visual stimulus. Moreover, performance analyses showed that when the additional velocity feature is used, the results are less affected by the influence of random elements and had a significantly lower percentage of grouping errors due to over- or underpartitioning, thus allowing the horizontal connectivity to be spatially extended without suffering noise, as happens in the visual cortex (Bosking et al., 1997).

A second analysis considered grouping in space and time and produced an algorithm that is able to identify spatial perceptual units and follow them during their motion. This is done by extending the previous approach, combining the affinity over  $\mathbb{R}^2 \times S^1 \times \mathbb{R}^+$ , here treated as an instantaneous connectivity, with an affinity in the cortical feature space of positions, activation times, local orientations, and locally detected velocities  $\mathbb{R}^2 \times \mathbb{R}^+ \times S^1 \times \mathbb{R}^+$ . In order to cope with the intrinsic causality of time evolution, we have worked directly on asymmetric affinities and used probabilistic clustering arguments (Pentney & Meila, 2005). The copresence of more than one connectivity mechanism in the segmentation of spatiotemporal stimuli is a realistic assumption of visual cortex behavior (Hess & Ledgeway, 2003).

Modeling the neural dynamical aspects of visual perception is a very delicate task. Indeed, many open questions remain to be addressed regarding, in particular how mean field equations in space-time can be compatible with the fast timescales of visual processing, or how the delays introduced by the neural dynamics influence the functionalities of cortical architectures. This first model of functional architecture in space-time does not take into account the integration time constants for neurons, as well as other biophysical phenomena, which certainly matter and deserve future study. These aspects are also strictly related to the general problem of understanding the timescale of mean field dynamics at the physical level. We observe, however, that this model of connectivity is able to implement a preactivation mechanism induced by spatiotemporal stimuli, whose biological plausibility was discussed and compared with neurophysiological measurement in Barbieri, Citti, Cocci et al. (2014), showing good accordance.

Our investigation has thus provided a geometric framework to perform clustering in feature spaces, following principles of extensions that can be generalized to higher-dimensional detected features. Possible extensions could be achieved by the inclusion of features such as color, three-dimensional stereo, or scale as additionally detected features. Depending

on the geometry of the stimuli and the psychophysiological indications, such extensions can lead to the definition of connectivities of higher complexity or the addition of copresent association mechanisms, following the two modalities that we discussed. This approach provides a way to design artificial perceptual algorithms adapted to the geometry of the information and a tool to propose new models of connections in the visual cortex, also suggesting further psychophysical or physiological experiments to compare and tune the model's parameters in order to fit visual perception and cognition behaviors. A natural future step will be the inclusion of realistic feature detection mechanisms, allowing the application of this methodology to real stimuli.

### Acknowledgments

---

D. B. was supported by a Marie Curie Intra European Fellowship (Prop. N. 626055) within the 7th European Community Framework Programme.

### References

---

- Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, 22(19), 8633–8646.
- August, J., & Zucker, S. W. (2003). Sketches with curvature: The curve indicator random field and Markov processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4), 387–400.
- Barbieri, D., Citti, G., Cocci, G., & Sarti, A. (2014). A cortical-inspired geometry for contour perception and motion integration. *Journal of Mathematical Imaging and Vision*, 49(3), 511–529.
- Barbieri, D., Citti, G., & Sarti, A. (2014). How uncertainty bounds the shape index of simple cells. *Journal of Mathematical Neuroscience*, 4(1), 1–15.
- Barnett, L., Buckley, C. L., & Bullock, S. (2009). Neural complexity and structural connectivity. *Physical Review E*, 79(5), 051914.
- Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 1373–1396.
- Ben-Shahar, O., & Zucker, S. (2004). Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Computation*, 16(3), 445–476.
- Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6), 2112–2127.
- Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2002). What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14(3), 473–491.
- Chisum, H. J., Mooser, F., & Fitzpatrick, D. (2003). Emergent properties of layer 2/3 neurons reflect the collinear arrangement of horizontal connections in tree shrew visual cortex. *Journal of Neuroscience*, 23(7), 2947–2960.

- Citti, G., & Sarti, A. (2006). A cortical based model of perceptual completion in the roto-translation space. *Journal of Mathematical Imaging and Vision*, 24(3), 307–326.
- Citti, G., & Sarti, A. (2014). *Neuromathematics of Vision: Lecture Notes in Morphogenesis*. New York: Springer.
- Cocci, G., Barbieri, D., & Sarti, A. (2012). Spatiotemporal receptive fields of cells in v1 are optimally shaped for stimulus velocity estimation. *JOSA A*, 29(1), 130–138.
- Coifman, R. R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1), 5–30.
- Coifman, R. R., Maggioni, M., Zucker, S. W., & Kevrekidis, I. G. (2005). Geometric diffusions for the analysis of data from sensor networks. *Current Opinion in Neurobiology*, 15(5), 576–584.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7), 1160–1169.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, 18(10), 451–458.
- Duits, R., & Van Almsick, M. (2008). The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2D Euclidean motion group. *Quarterly of Applied Mathematics*, 66(1), 27–68.
- Faugeras, O., Veltz, R., & Grimbert, F. (2009). Persistent neural states: Stationary localized activity patterns in nonlinear continuous N-population, Q-dimensional neural networks. *Neural Computation*, 21(1), 147–187.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33(2), 173–193.
- Geisler, W., Perry, J., Super, B., & Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41(6), 711–724.
- Graham, C., & Talay, D. (2013). *Stochastic simulation and Monte Carlo methods*. New York: Springer.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86(5), 1698–1702.
- Hess, R. F., & Ledgeway, T. (2003). The detection of direction-defined and speed-defined spatial contours: One mechanism or two? *Vision Research*, 43(5), 597–606.
- Hoffman, W. C. (1989). The visual cortex is a contact bundle. *Applied Mathematics and Computation*, 32(2), 137–167.
- Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Scientific American.
- Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 198(1130), 1–59.
- Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6), 1233–1258.
- Kannan, R., Vempala, S., & Vetta, A. (2004). On clusterings: Good, bad and spectral. *Journal of the ACM*, 51(3), 497–515.
- Kisvarday, Z., & Eysel, U. T. (1992). Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17). *Neuroscience*, 46(2), 275–286.



- Kisvarday, Z., Toth, E., Rausch, M., & Eysel, U. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cerebral Cortex*, 7(7), 605–618.
- Koenderink, J. J., & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics*, 55(6), 367–375.
- Lafon, S., & Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9), 1393–1403.
- Ledgeway, T., & Hess, R. F. (2002). Rules for combining the outputs of local motion detectors to define simple contours. *Vision Research*, 42(5), 653–659.
- Ledgeway, T., Hess, R. F., & Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of association field models of contour integration. *Vision Research*, 45(19), 2511–2522.
- London, M., & Häusser, M. (2005). Dendritic computation. *Annu. Rev. Neurosci.*, 28, 503–532.
- Malach, R., Schirman, T., Harel, M., Tootell, R., & Malonek, D. (1997). Organization of intrinsic connections in owl monkey area MT. *Cerebral Cortex*, 7(4), 386–393.
- Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5), 1127–1147.
- Meila, M., & Pentney, W. (2007). Clustering by weighted cuts in directed graphs. In *Proceedings of the SIAM International Conference on Data Mining* (Vol. 7, pp. 135–144). Philadelphia: SIAM.
- Meila, M., & Shi, J. (2001). A random walks view of spectral segmentation. In *Proceedings of the 8th International Workshop on Artificial Intelligence and Statistics*. <http://www.gatsby.ucl.ac.uk/aistats/aistats2001/index.html>
- Mel, B. W., & Schiller, J. (2004). On the fight between excitation and inhibition: Location is everything. *Sci. STKE*, 250, 1–3.
- Mumford, D. (1994). *Elastica and computer vision*. New York: Springer.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, 14 (pp. 849–856).
- Pavan, A., Casco, C., Mather, G., Bellacosa, R. M., Cuturi, L. F., & Campana, G. (2011). The effect of spatial orientation on detecting motion trajectories in noise. *Vision Research*, 51(18), 2077–2084.
- Pentney, W., & Meila, M. (2005). Spectral clustering of biological sequence data. In *Proceedings of the 20th National Conference on Artificial Intelligence* (pp. 845–850). Cambridge, MA: AAAI Press.
- Perona, P., & Freeman, W. (1998). A factorization approach to grouping. In *Computer Vision/ECCV'98* (pp. 655–670). New York: Springer.
- Petitot, J., & Tondut, Y. (1999). Vers une neurogéométrie: Fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques informatique et sciences humaines*, 145, 5–102.
- Platen, E. (1999). An introduction to numerical methods for stochastic differential equations. *Acta Numerica*, 8, 197–246.



- Rainville, S. J., & Wilson, H. R. (2005). Global shape coding for motion-defined radial-frequency contours. *Vision Research*, 45(25), 3189–3201.
- Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, 28(2), 147–166.
- Rodman, H. R., & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research*, 27(12), 2035–2048.
- Roerig, B., & Kao, J. (1999). Organization of intracortical circuits in relation to direction preference maps in ferret visual cortex. *Journal of Neuroscience*, 19(24), RC44.
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326.
- Sanguinetti, G., Citti, G., & Sarti, A. (2010). A model of natural image edge co-occurrence in the roto-translation group. *Journal of Vision*, 10(14), 1–16.
- Sarti, A., & Citti, G. (2014). The constitution of visual perceptual units in the functional architecture of V1. *Journal of Computational Neuroscience*, 1–16. doi:10.1007/s10827-014-0540-6
- Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge: Cambridge University Press.
- Series, P., Georges, S., Lorenceau, J., & Frégnac, Y. (2002). Orientation dependent modulation of apparent speed: A model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. *Vision Research*, 42(25), 2781–2797.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.
- Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11), 5033–5037.
- Verghese, P., & McKee, S. P. (2006). Motion grouping impairs speed discrimination. *Vision Research*, 46(8), 1540–1546.
- Verghese, P., McKee, S. P., & Grzywacz, N. M. (2000). Stimulus configuration determines the detectability of motion signals in noise. *JOSA A*, 17(9), 1525–1534.
- Verghese, P., Watamaniuk, S. N., McKee, S. P., & Grzywacz, N. M. (1999). Local motion detectors cannot account for the detectability of an extended trajectory in noise. *Vision Research*, 39(1), 19–30.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217.
- Watamaniuk, S. N. (2005). The predictive power of trajectory motion. *Vision Research*, 45(24), 2993–3003.
- Weiss, Y. (1999). Segmentation using eigenvectors: A unifying view. In *Proceedings of the Seventh IEEE International Conference on Computer Vision* (Vol. 2, pp. 975–982). Piscataway, NJ: IEEE.
- Weliky, M., Bosking, W. H., & Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature*, 379(6567), 725–728.

- Wertheimer, M. (1938). "Laws of organization in perceptual forms," In W. B. Ellis (Ed.), *A Sourcebook of Gestalt Psychology* (pp. 71–88). London: Routledge and Kegan Paul.
- Williams, L. R., & Jacobs, D. W. (1995). Stochastic completion fields: A neural model of illusory contour shape and salience. In *Proceedings of the Fifth International Conference on Computer Vision* (pp. 408–415). Piscataway, NJ: IEEE.
- Wu, W., Tiesinga, P. H., Tucker, T. R., Mitroff, S. R., & Fitzpatrick, D. (2011). Dynamics of population response to changes of motion direction in primary visual cortex. *Journal of Neuroscience*, 31(36), 12767–12777.
- Zelnik-Manor, L., & Perona, P. (2004). Self-tuning spectral clustering. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, 17 (pp. 1601–1608). Cambridge, MA: MIT Press.
- Zhang, J., Duits, R., & Romeny, B. M. (2014). On numerical approaches for linear left-invariant diffusions on  $se(2)$ , their comparison to exact solutions, and their applications in retinal imaging. arXiv preprint arXiv:1403.3320.

---

Received June 14, 2014; accepted January 8, 2015.