

Motion detectors and motion segregation

JUN ZHANG

*Department of Psychology, University of Michigan, 525 East University Ave.,
Ann Arbor, MI 48109, USA*

Received 15 October 1993; revised 25 October 1994; accepted 13 December 1994

Abstract—The response of motion detectors necessarily confound image velocity with image structure. In particular, even a rigidly moving image (with a uniform velocity field) will give rise to non-uniform detector responses. A mathematical framework has been proposed on how to intrinsically compare motion detectors' responses so that their differences will reflect the true differences in image velocity (Zhang and Wu, *Proc. Natl. Acad. Sci. USA* **87**, 7819–7823, 1990). Here, this notion of 'intrinsic differentiation' was implemented by introducing a γ -matrix determined by the image spatial gradients. The perceptual phenomenon of random-dot motion segregation was successfully simulated.

1. INTRODUCTION

Motion detectors are pre-processors for the motion system. Their outputs are used by the visual system to perform motion-based figure-ground segregation, that is, to define the object, and to generate optic flow field, that is, to define the velocity. However, motion detectors compute the variation (over space and time) of image luminance, where neither object nor velocity is explicitly given. Moreover, the appearance of an object (spatial characteristics) and its motion (temporal characteristics) are confounded in the response of any individual detector. This is illustrated in the so-called 'aperture' problem (Marr and Ullman, 1981; Adelson and Movshon, 1982; Hildreth, 1984; Poggio *et al.*, 1989), in which local motion measurements do not coincide (even in direction) with the velocity of a rigidly translating 2-D image owing to the locally oriented (1-D like) luminance profile of the image. In this paper, we will first demonstrate this problem in general, that is, how a rigidly translating image (an image with uniform velocity across the space) will give rise to non-uniform responses of motion detectors. We will then provide a solution to this generalized aperture problem by requiring a 'correct' reading out and 'intrinsic' comparison of these non-uniform detector responses. Finally, we will apply this idea to a simple but non-trivial testing case: figural segregation from a two-frame random-dot kinematogram where both object shape and velocity are the result (rather than the premise) of correct comparison of motion-detector outputs.¹ Mathematically minded readers are directed to Zhang and Wu (1990) for a more elaborate account of this framework based on non-Euclidean differential geometry.

2. RESPONSE PROPERTIES OF MOTION DETECTORS

2.1. Classes of motion detector

There are two different classes of motion detectors recognized in the current literature: (1) *Motion energy detectors* (van Santen and Sperling, 1984; Adelson and Bergen, 1985; Watson and Ahumada, 1985), which extract the spatial-temporal frequency power spectrum of an image based on the filtering characteristics of the detectors; (2) *Motion field detectors* (Horn and Schunck, 1981; Nagel, 1983; Haralick and Lee, 1983; Uras *et al.*, 1988), which extract the vector field of image velocities based on calculations of spatial and temporal gradients of an image. An important link between these two different classes of detectors was established in a paper by Reichardt and Schlögl (1988). They showed that in the limiting case of infinitesimally small spatial separations and temporal delays between the sub-units which make up the detector, the response of a Reichardt-type Elementary Motion Detector or EMD (Hassenstein and Reichardt, 1956) is analogous in form to that of a motion field detector. Specifically, the response of the EMD under these approximations is shown to be related to the image velocity through an image gradient matrix of second-order spatial derivatives (as we shall discuss below). Since the EMD is commonly regarded as a prototypical motion energy detector when its component sub-units have elaborated spatial and temporal filtering properties (van Santen and Sperling, 1985), the mathematical demonstration of Reichardt and Schlögl (1988) provides a unified picture of the essence of motion computation. We may, therefore, focus on this simplified description of motion-detector response while conveniently ignoring for the moment the filtering properties of the sub-units, which reflect the spatiotemporal range and peak values to which an exemplar detector is tuned.

2.2. Elementary motion detector

Motivated by Reichardt and Schlögl (1988), the following form of motion detector is considered (see also Poggio *et al.*, 1989):

$$V = -\frac{\partial}{\partial t}(\nabla_l f), \quad (1)$$

where $f = f(x, y, t)$ is the image intensity function, and ∇_l is the gradient operator (partial derivative) along direction l , called the 'preferred' direction or the axis of the detector. The detector computes the spatial and then the temporal gradients of the image *in succession*. There are many motion detectors at each location, with their preferred directions covering the whole range of 360° . An orthogonal pair is formed for two individual detectors bearing orthogonal axes, and the response of the detector pair can be described as a two-component vector $\vec{V} = [V_1(x, y, t), V_2(x, y, t)]^T$. The reason we choose this form of motion detector is as follows. For a rigid object $f = f(x(t), y(t))$ undergoing 2-D translatory motion with velocity $\vec{v} = [dx(t)/dt, dy(t)/dt]^T = [v_1(t), v_2(t)]^T$, the response of the orthogonal pair of detectors (simply

called the detector response) is given by (using image equation $df(x(t), y(t))/dt = 0$)

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = - \begin{bmatrix} \partial(\partial f / \partial x) / \partial t \\ \partial(\partial f / \partial y) / \partial t \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad (2)$$

where subscripts of f denote partial derivatives with respect to coordinates x and y , so that $f_{xx} = \partial^2 f / \partial x^2$, etc., and $f_{xy} = f_{yx}$. This vectorial relationship between detector response \vec{V} and image velocity \vec{v} is analogous to that obtained from an EMD (c.f. Reichardt and Schlögl, 1988) and thus may adequately represent the nature of motion computations in general. For this reason, we will use this form of detector response to present our analysis.

2.3. Consistency and completeness

The detector response \vec{V} as expressed in Eqn (2) demonstrates great mathematical clarity. In particular, the following desirable properties are observed:

(1) *Self-consistency*. \vec{V} is *covariant*; that is, the response of any detector pair at a specific location represents the same vector² regardless of the choice of the preferred axis. Consider the new coordinate axes $x'-y'$ obtained via a counterclockwise rotation of the $x-y$ coordinate axes by an angle α :

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The Cartesian components of vector \vec{v} should co-vary with the change of coordinates

$$\begin{bmatrix} v'_1 \\ v'_2 \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

while the second-order derivatives (the Hessian matrix \mathcal{F}) will be transformed (through the chain rule of differentiation) according to

$$\begin{bmatrix} \partial^2 f / \partial x'^2 & \partial^2 f / \partial x' \partial y' \\ \partial^2 f / \partial y' \partial x' & \partial^2 f / \partial y'^2 \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \partial^2 f / \partial x^2 & \partial^2 f / \partial x \partial y \\ \partial^2 f / \partial y \partial x & \partial^2 f / \partial y^2 \end{bmatrix} \\ \times \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}.$$

Therefore, the vector \vec{V} in the new coordinates, as given by the multiplication of the Hessian matrix \mathcal{F} and image velocity \vec{v} , becomes

$$\begin{bmatrix} V'_1 \\ V'_2 \end{bmatrix} = \begin{bmatrix} f_{x'x'} & f_{x'y'} \\ f_{y'x'} & f_{y'y'} \end{bmatrix} \begin{bmatrix} v'_1 \\ v'_2 \end{bmatrix} \\ = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \partial^2 f / \partial x^2 & \partial^2 f / \partial x \partial y \\ \partial^2 f / \partial y \partial x & \partial^2 f / \partial y^2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix};$$

that is, the components of the vector \vec{V} also transform in a covariant fashion. The biological significance of this covariant relationship is that all detectors give mutually compatible information regardless of the individual 'label' of their preferred directions. The computational significance is that we only need, at least in theory, one orthogonal pair of detectors (two orthogonal detector axes) to implement this type of motion detector. Of course in practice, one might want to utilize more than one orthogonal pair to deal with noisy inputs.

(2) *Self-completeness.* \vec{V} is *invertible*; that is, the local image velocity \vec{v} can be unambiguously determined by the detector response \vec{V} through the inversion of the 2×2 image gradient (Hessian) matrix \mathcal{F}

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}^{-1} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad (3)$$

where

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}^{-1} = \frac{1}{\det \mathcal{F}} \begin{bmatrix} f_{yy} & -f_{xy} \\ -f_{yx} & f_{xx} \end{bmatrix}. \quad (4)$$

The only exception occurs where the determinant $\det \mathcal{F} = f_{xx}f_{yy} - f_{xy}^2$ is zero, a restriction closely related to the so-called 'aperture problem' (Adelson and Movshon, 1982; Hildreth, 1984; Reichardt *et al.*, 1988) or at least in its weak form (Poggio *et al.*, 1989); namely, it is physically impossible to specify the image velocity along the direction where the image gradient matrix degenerates. This is a restriction due to the nature of the image, not due to the biological processing of it. Any physically extractable information about image motion is thus completely preserved in the detector response. However, to represent multiple velocities at one location (as in motion transparency), one needs to invoke simultaneous motion computations at different spatial scales, a topic we will not go into in this paper.

3. INTRINSIC COMPARISON OF DETECTOR RESPONSES

The detector response \vec{V} clearly depends on the spatial and temporal gradients of the image intensity function $f(x, y)$. As can be seen from Eqn (2), the directions of \vec{V} do not usually coincide with the directions of \vec{v} . Moreover, even if an object moves rigidly (i.e. \vec{v} is not a function of space variable x and y), the detector response is still a function of spatial locations. This is because the spatial derivatives (gradients) of $f(x, y)$ are generally not constant over the entire space — each image has its own structure. It is of great importance to consider how to correctly read out the detector response. The difference between adjacent detector responses $d\vec{V}$ (in the case of $\vec{v} = \text{constant}$, i.e. rigid motion) is calculated according to Eqns (2) and (3)

$$\begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} = \begin{bmatrix} df_{xx} & df_{xy} \\ df_{yx} & df_{yy} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} df_{xx} & df_{xy} \\ df_{yx} & df_{yy} \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}^{-1} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix},$$

or simply

$$\begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad (5)$$

whereby we define

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} = \begin{bmatrix} df_{xx} & df_{xy} \\ df_{yx} & df_{yy} \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}^{-1}. \quad (6)$$

Note that the γ -matrix is related only to the image gradient, and its elements are differential quantities. Equation (5) expresses what the differences between adjacent detector responses ought to be if the image were moving with constant velocity (rigid motion), and the right-hand side of Eqn (5) is directly calculable from the image intensity function $f(x, y)$ and the detector response \vec{V} . As we noted earlier, the detector response is not a uniform vector field (i.e. $d\vec{V} \neq 0$) even for a rigidly moving image, since the γ -matrix elements do not identically vanish in general.

In order to unconfound the information of image structure (spatial gradient) from the information of image velocity in the motion-detector response, we introduce the D -derivative

$$\begin{bmatrix} DV_1 \\ DV_2 \end{bmatrix} = \begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} - \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \quad (7)$$

In addition to the usual (component-wise) differentiation, the D -derivative has a second term which represents the adjustment of the differential detector response due to the confounding information contributed by any non-uniform image. Thus, $D\vec{V}$ calculates the intrinsic differences of the response vector \vec{V} ('intrinsic' in the sense that \vec{V} reflects solely the image velocity \vec{v} , with the contamination from the image structure removed). Explicitly stated, $D\vec{V} = 0$ for a rigidly translating image, no matter what the image structure is.

The idea of intrinsic differentiation is rooted in the long-established theory of differential geometry. In a general non-Euclidean space, the intrinsic constancy of a vector field at different spatial locations — intrinsic to the observer in that geometry — is not defined by looking at each of their Cartesian vector components. (Recall how one compares vectors on a unit sphere, for example.) Instead, it is achieved by transplanting a vector from one location to another while adjusting the Cartesian components by some prescribed amount so as to preserve the parallelism (an intrinsic measurement under that geometry) of the vector under transplantation. Therefore, two vectors at nearby points may be deemed intrinsically constant despite the fact that their Cartesian components are not usually equal, but differ exactly by some amount. This amount is determined by a set of numbers called the affine connection Γ (equivalent to the γ -matrix defined above), and is needed in order to take into account the particular geometry under which a vector field is defined. In this sense, the non-Euclidean space is unequivocally determined by its affine connection Γ . In particular, it specifies how to 'correctly' compare the intrinsic difference (in the geometrical sense) between vectors at two nearby locations, namely the intrinsic differentiation (equivalent to the D -derivative here). That the intrinsic difference (in the sense of discounting image

structure) of the detector response vector $D\vec{V}$ is identically zero for a rigidly moving object can now be rephrased: \vec{V} is an intrinsically constant vector field under this (generally speaking) non-Euclidean geometry. Obviously, the particular geometry to be used is determined by the particular image structure. Furthermore, since human observers do actually segregate those points of $D\vec{V} = 0$ to form a single percept of a visual object (as we shall discuss in the next section), it is all too appropriate a metaphor to say that the higher visual centers of the brain interpret the motion-detector response \vec{V} as if it were situated in a non-Euclidean space. The fascinating application of these geometric concepts has led to a novel mathematical framework to describe visual perception (Zhang and Wu, 1990).

To summarize, we propose that neighboring detector responses should be compared intrinsically (using operator D) rather than veridically (using operator d). The intrinsic differentiation is the natural way of comparing neighboring vectors under a given geometry; it reduces to ordinary differentiation when the geometry is Euclidean. Here we have treated the image luminance profile as if it provided an underlying geometry whereby this intrinsic differentiation of detector responses is to be applied. In other words, an image determines a geometry (through the γ -matrix); detector responses are to be compared intrinsically (using D) under that geometry. Figural segregation can be achieved when nearby motion-detector responses are 'equal' ($D\vec{V} = 0$) under this intrinsic comparison (intrinsic in that it discounts the image luminance structure).

4. RANDOM-DOT MOTION SEGREGATION: AN EXAMPLE

As an illustration of this formulation, we simulate the phenomenon of figural segregation resulted from *successively* presented random-dot stimuli. Two frames of random-dot patterns are alternated in temporal sequence at the same spatial location (Fig. 1). The two patterns are uncorrelated in their distributions of dots, except for a region (subset) of identical but uniformly displaced dots. A vivid percept of object shape corresponding to the uniformly displaced region 'pops out', so long as the magnitude of this uniform displacement is less than a certain amount, often referred to as the Braddick limit (Braddick, 1974).

The random-dot kinematogram is a stimulus that is devoid of positional cues and therefore *exclusively* drives the luminance-based short-range motion mechanism (Anstis, 1980; Braddick, 1980). The classes of motion detectors that mediate this psychophysically-revealed 'true' motion mechanism are the motion energy detectors and motion field detectors, described in Section 2. On the other hand, the aperture problem suffered by these motion detectors poses an acute problem for understanding the vivid and almost instantaneous perceptual segregation in a random-dot kinematogram. We perform a simulation on this test case as a way to illustrate our proposed solution of the aperture problem, that is, through the application of the intrinsic differentiation notion advanced in Section 3.

In the simulation, the responses of motion detectors \vec{V} are calculated by taking the (temporal) differences of the spatial gradients at each point between the two frames (see Eqn (1)). These response vectors are non-uniform (in terms of Cartesian

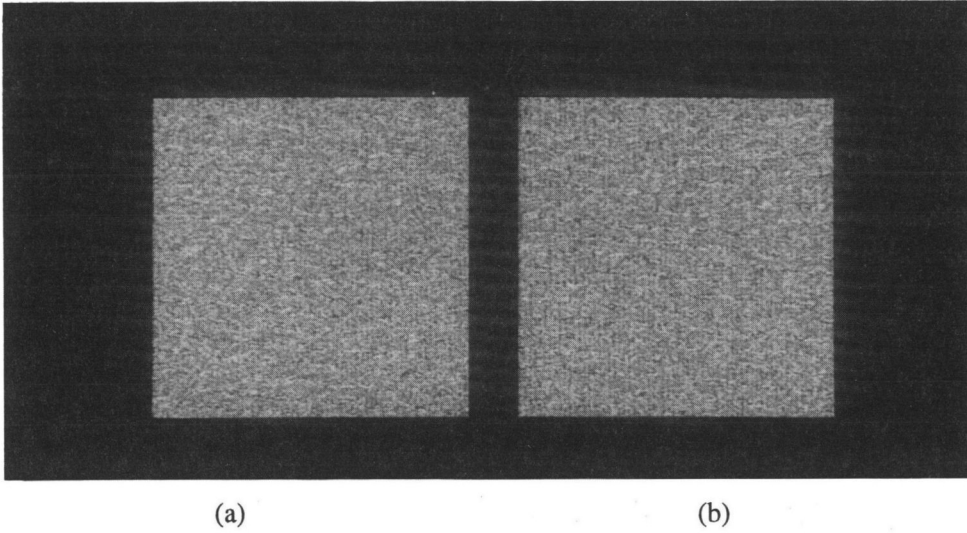


Figure 1. Random-dot kinematogram. The two random-dot patterns are presented at the same spatial location while alternating in temporal sequence. The frame size is 256×256 in pixels, and each pixel has equal probabilities of being 'bright' or 'dark'. The target consists of a central region with a uniform (correlated) displacement between the frames, and the background consists of dynamical noise (random variations). Here the displacement of the target T is one pixel toward the left in (b) as it is in (a).

components) across the image. The responses of neighboring detectors are then intrinsically compared using a γ -matrix to decide the rigidity in the motion. The γ -matrix is calculable from the image gradient matrix \mathcal{F} of the *first* frame through Eqn (6). It is used for specifying the amount to be adjusted toward the difference of neighboring detectors' responses $d\vec{V}$ so that the adjusted or intrinsic difference $D\vec{V}$ faithfully reflects the actual difference of the image velocity $d\vec{v}$. The region corresponding to the uniform displacement in the two random-dot frames has the same image velocity ($d\vec{v} = \vec{0}$), and thus can be segregated through the intrinsic differentiation operation on \vec{V} (in other words, the points satisfying $D\vec{V} = 0$ are to be segregated).

The procedures are outlined step by step in the following (Fig. 2): (1) The binary-valued random-dot patterns are first blurred using a Gaussian (low-pass) kernel to yield continuous images. Here the Gaussian kernel has a standard deviation of 3 pixels. (2) We then calculate \vec{V} and γ -matrix by taking the difference of neighboring pixels (spatial gradient) and adjacent frame (temporal gradient). The second derivatives were obtained by taking the difference of the difference of pixel intensities. Note that the calculation of γ -matrix only makes use of the first frame, while the calculation of \vec{V} makes use of both frames. (Actually, it is the γ -matrix multiplied by $\det\mathcal{F}$ that is calculated; see below.) (3) The differential vector $d\vec{V}(P)$ at a point P with respect to one of its four neighbors $P + \delta P$ is obtained by straightforward, component-wise subtraction of the detector response at P from that at $P + \delta P$: $d\vec{V}(P) = \vec{V}(P + \delta P) - \vec{V}(P)$. (4) From $\vec{V}(P)$, $d\vec{V}(P)$, and the γ -matrix value at P , the intrinsic difference $D\vec{V}(P)$ of the response vector between P and $P + \delta P$ may be obtained. To avoid $\det\mathcal{F} = 0$ for the calculation of γ -matrix, both sides of Eqn (7)

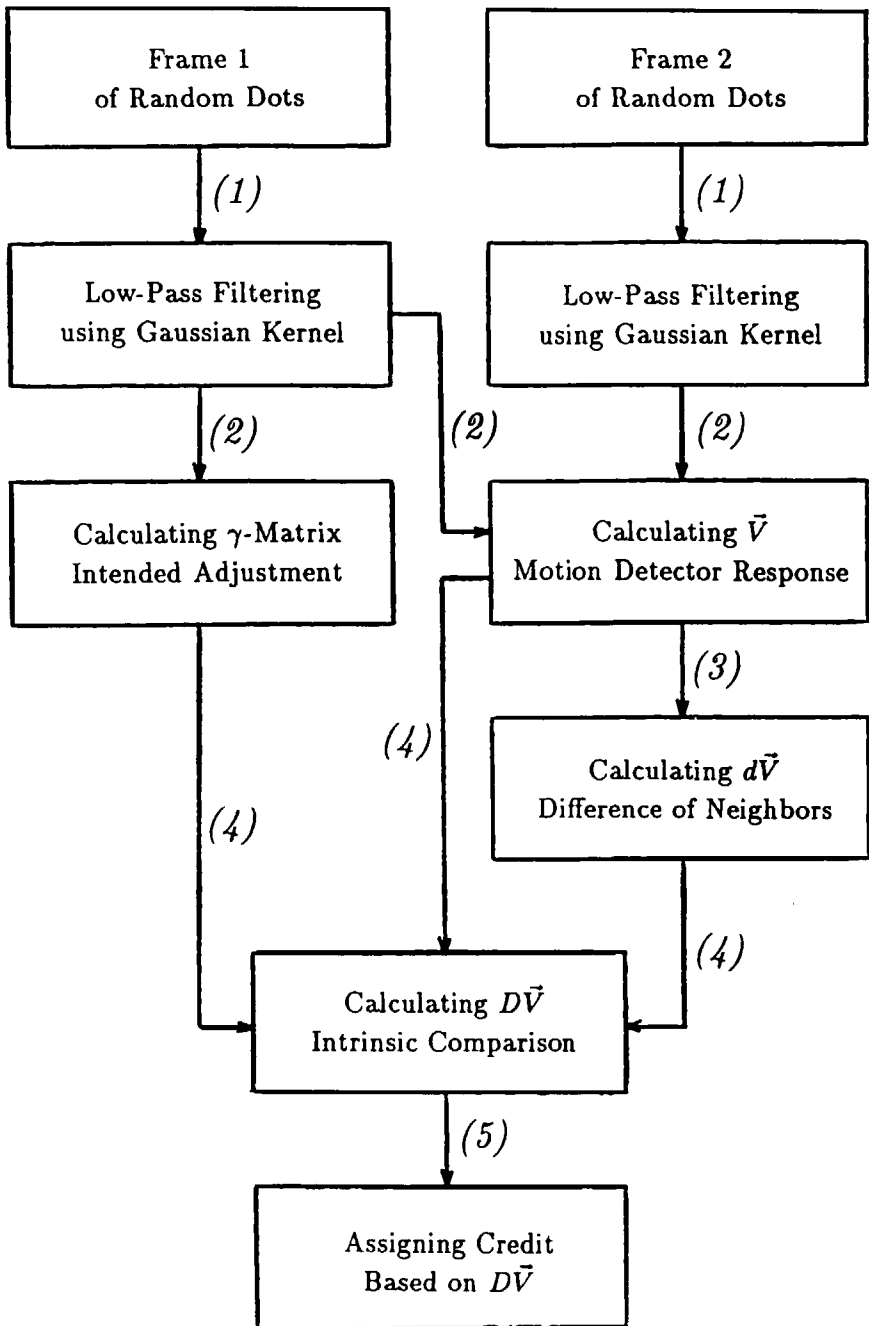


Figure 2. Procedures of the implementation. After Gaussian blurring (step 1), the motion-detector response \vec{V} and the image gradient based γ -matrix are calculated (step 2). The γ -matrix is intended for calculating an adjustment to the differential response of neighboring detectors $d\vec{V}$ (step 3), so that their intrinsic difference $D\vec{V}$ reflects faithfully the differences in image velocity of neighboring points (step 4), and can be used for assigning an additive credit of image rigidity at that point (5).

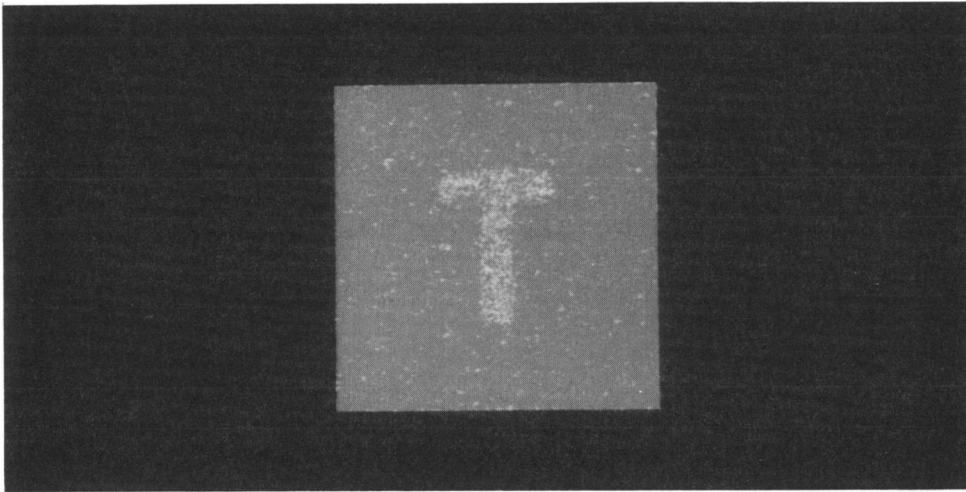


Figure 3. The simulation result of figural segregation of the random-dot motion stimuli of Fig. 1. A pixel is rendered white if the overall credit of image rigidity meets a certain criterion value. Note the background noise, which corresponds to various 'ghost' matches. The Gaussian convolution kernel has $\sigma = 3$ pixels.

are multiplied by $\det \mathcal{F}$. (5) Ideally, $D\vec{V}(P)$ should be zero for the region of rigid motion, and a relatively small value of $D\vec{V}(P)$ indicates that there is proportionally little difference in the image velocity between pixels P and $P + \delta P$. In the simulation, some 'credits' were assigned to pixels P and $P + \delta P$, which were proportional to the square-root of $|D\vec{V}(P) \cdot D\vec{V}(P)| = DV_1^2(P) + DV_2^2(P)$. The credits from all four neighbors of a given point P were then added. This value measures the likelihood that point P shares a common motion (same velocity \vec{v}) with its neighboring points, and therefore its likelihood of being perceptually segregated along with them. The individual credit at each image point is thus computed in parallel, and a map of such credit can be displayed after proper thresholding and rendering (Fig. 3).

This simulation algorithm is massively parallel and strictly local, since identical operations are carried out for each image point involving only the four neighbors. The processing time would not scale with image size if it were to run on a parallel machine, which may be of practical interest for potential applications. Secondly, this algorithm does not require an explicit recovery of image velocity. Indeed, for the pixels at the background, their velocity is ill-defined anyway, either computationally or psychologically. This aspect is different from previous algorithms of motion computation (e.g. Haralick and Lee, 1983; Nagel, 1983; Tretiak and Pastor, 1984; Uras *et al.*, 1988).

5. COMPARING DETECTOR RESPONSES USING GEODESICS

The algorithm presented above is essentially a second-order method in which image Hessian (second-order spatial derivatives) needs to be evaluated. Like all second-order

methods, it may suffer from image noise because of successive spatial differentiations. To circumvent this problem, additional algorithms are sought to implement the notion of intrinsic differentiation.

Indeed, as provided by Riemannian differential geometry, nearby vectors may be compared 'intrinsically' through the use of the so-called geodesics. Geodesics are special curves of a geometry; a geodesic connects points via a shortest path. They are an extension of (and indeed reduce to) straight lines of Euclidean geometry. Geodesics are unambiguously defined in a geometry and, in the present case where the geometry is induced by an image, by image luminance structure. Analogous to what happens under the Euclidean geometry, two vectors at nearby points are considered parallel if they make the same angle with respect to the geodesic passing through the two points. This suggests that we may intrinsically compare detector responses through their projection onto the image-induced geodesics.

It has been proved (Zhang and Wu, 1990, Zhang 1992) that a geodesic induced by the image structure is given implicitly as the solution curve $x = x(s)$, $y = y(s)$ of the following equation (with s the parameter of the curve):

$$\begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} s + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad (8)$$

where b_1, b_2, c_1, c_2 are all constants specifying a particular geodesic; f_x, f_y are first-order image gradients. The projection of \vec{V} onto a given geodesic, i.e. a curve $(x(s), y(s))$ that satisfies the above equation, is

$$V_{\text{proj}} = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} dx/ds \\ dy/ds \end{bmatrix}.$$

In the case of rigid translation, the detector response vector $\vec{V} = [V_1, V_2]^T$ is given by Eqn (2). Therefore

$$\begin{aligned} V_{\text{proj}} &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} dx/ds \\ dy/ds \end{bmatrix} \\ &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} df_x/ds \\ df_y/ds \end{bmatrix} \\ &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = v_1 b_1 + v_2 b_2 \end{aligned}$$

(the first line made use of Eqn (2), the second line the chain rule of differentiation, and the third line, Eqn (8)), or

$$V_{\text{proj}} = \text{constant}. \quad (9)$$

This is to say, the projection of detector responses \vec{V} onto the geodesics (defined by Eqn (8)) will be constant for those points that correspond to rigid translation. It

follows that to correctly read out detector responses, we need only construct geodesics and project the motion vector field onto those geodesics that are determined solely by the image structure.

Explicit calculation of geodesics can be facilitated by modifying Eqn (8) to yield

$$b_2 f_x - b_1 f_y = b_2 c_1 - b_1 c_2.$$

Since the b and c are arbitrary constants, the above expression may be recast using the directional derivative ∇_l (gradient operator along a direction l)

$$\frac{d}{ds}(\nabla_l f) = \text{constant}. \quad (10)$$

Equation (10) says that a geodesic is formed by those points whose first-order spatial gradient along given direction l remains constant. This gives rise to a first-order algorithm for calculating geodesics and, through geodesics, for intrinsically comparing detector responses. Again, figural segregation is achieved if detector responses are deemed identical under this intrinsic comparison. Preliminary results were reported (Zhang, 1993), and a detailed description will appear elsewhere.

6. DISCUSSION

The motion-detector response in Eqn (1) confounds image velocity with image spatial gradient. As such, even to a rigidly translating object with spatially uniform velocity field, the response of motion detectors are non-uniform. This aperture problem calls for proper interpretation of local motion measurements performed by the detectors. This problem is particularly acute when the object itself has to be defined by a uniform velocity field, such as in the random-dot kinematogram. This paper proposes an intrinsic comparison of detector responses based on a measure (geometry) that is purely and completely determined by static image luminance structure. The intrinsic differentiation of motion-detector responses discounts the contamination by image spatial structure and thus is (or is at least proportional to) a true differentiation performed on image velocity.

One may ask, why not try to design in the first place a detector that is purely selective for image velocity vector? The answer lies in the nature of motion computations. To speak of velocity, one must be referring to, whether explicitly or not, the velocity of something (some identified feature). Therefore a pure velocity detector requires feature detectors as its input and feature matching for its computation. This can become quite cumbersome, if not entirely impossible, in cases where the feature itself is to be identified by motion computation, such as a textured figure moving on a similar textured background (see earlier section on dynamical random-dot patterns). It is quite unlikely that identifying and matching each individual element in the random-dot or other complex pattern is needed for motion computations. On the other hand, both motion energy detectors and motion field detectors calculate spatiotemporal changes of the image function *without* the overhead of feature detection. As a compromise, these

forms of motion detectors necessarily confound image velocity with image structure, be it image Fourier power spectrum or image gradient structure. To read out motion detector responses correctly (less any uncertainty due to the aperture problem, of course), it is of paramount importance to discount the image structure information.

Tretiak and Pastor (1984) adopted the same form of detector response given by Eqn (2). They then proceeded directly with matrix inversion and obtained an explicit velocity representation, as was common in conventional approaches. It is true that in our present implementation of intrinsic differentiation (D -derivative), we also need to invert the Hessian matrix (see Eqns (6) and (7)), an operation unsatisfactory both biologically and computationally. However, while matrix inversion is unavoidable for obtaining an *explicit* velocity representation, it can in principle be avoided in our formulation, since intrinsic comparisons of vectors can be (or are perhaps more naturally) described and performed under an appropriate geometrical context (see Section 5). Indeed, we regard detector responses V as *the* internal representation of any motion stimuli, and proceed to compare them directly. Although the intrinsic comparison of their Cartesian components may seem complicated (see Eqn (7)), it is nevertheless the only meaningful way of comparing these detector responses. The net result of unconfounding the image velocity from the image structure (by the use of D -derivative) is equivalent to treating \dot{V} as a vector field in a certain non-Euclidean space, whereby the intrinsic differentiations become the most natural operations. It is for this reason that we suggested that our motion perception may be adequately described under a general (non-Euclidean) geometry, that the motion-detector responses resulting from a rigid motion stimulus are regarded by higher processing centers as intrinsically constant vectors, and that the particular form of the non-Euclidean geometry is determined by the image structure through a metric tensor (Zhang and Wu, 1990). This novel geometrical description of motion perception provides a clarified picture for understanding the essence of object segregation and object representation in the motion system.

Acknowledgement

Part of this research was conducted while the author was a graduate student at University of California, Berkeley and was supported by research assistantships from NSF Grant BNS-8819867 and PHS Grant EY-00014 to Dr Russell L. De Valois and Dr Karen K. De Valois.

NOTES

1. This is a simplest case of motion-based figure-ground segregation. It has long been established that random-dot segregation occurs relatively early (and pre-attentively) in visual perception (Anstis, 1980; Braddick, 1980) and is mediated by the short-range or 'true' motion mechanism (Braddick, 1974), corresponding to the class of motion detectors to be discussed. The 'aperture problem', however, is manifested to its fullest even though we restrict our discussions to 2-D translation only, as in the random-dot kinematogram.

2. A 2-vector is an entity (of the two-dimensional linear space) represented by a duplex of numbers that are dependent on a particular selection of and therefore co-vary with a change of the coordinates. Here we do not go into the technical difference between a *covariant* vector and a *contravariant* vector.

REFERENCES

- Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299.
- Adelson, E. H. and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature (London)* **300**, 523–525.
- Anstis, S. M. (1980). The perception of apparent motion. *Phil. Trans. Roy. Soc. Lond. B* **290**, 153–168.
- Braddick, O. (1974). A short-range process in apparent motion. *Vision Res.* **14**, 519–527.
- Braddick, O. (1980). Low-level and high-level processes in apparent motion. *Phil. Trans. Roy. Soc. Lond. B* **290**, 137–151.
- Haralick, R. M. and Lee, J. S. (1983). The facet approach to optic flow. In: *Proceedings Image Understanding Workshop*. L. S. Baumann (Ed.). Science Applications, Arlington, VA, pp. 84–93.
- Hassenstein, B. and Reichardt, W. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenbewertung bei der Bewegungsperzeption der Rüsselkäfers. *Chlorophanus. Z. Naturforsch* **11b**, 513–524.
- Hildreth, E. C. (1984). *The Measurement of Visual Motion*. MIT Press, Cambridge, MA.
- Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.* **17**, 185–203.
- Marr D. and Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proc. Roy. Soc. Lond. B* **208**, 385–387.
- Nagel, H.-H. (1983). Displacement vectors derived from second-order intensity variations in image sequences. *Comp. Vision Graph. Image Process.* **21**, 85–117.
- Poggio, T., Yang, W. and Torre, V. (1989). Optical flow: computational properties and networks, biological and analog. In: *The Computing Neuron*. R. Durbin, C. Miall and G. Mitchison, G. (Eds). Addison-Wesley Publishing, pp. 355–370.
- Reichardt, W. and Schlögl, R. W. (1988). A two dimensional field theory for motion computation. *Biol. Cybernet.* **60**, 23–35.
- Reichardt, W., Schlögl, R. W. and Egelhaaf, M. (1988). Movement detectors provide sufficient information for local computation of 2-D velocity field. *Naturwissenschaften* **75**, 313–315.
- van Santen, J. P. H. and Sperling, G. (1984). A temporal covariance model of human motion perception. *J. Opt. Soc. Am. A* **1**, 451–473.
- van Santen, J. P. H. and Sperling, G. (1985). Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300–321.
- Tretiak, O. and Pastor, L. (1984). Velocity estimation from image sequences with second order differential operators. In: *Proc. Int. Conf. Pattern Recognition*. Montreal, Quebec, pp. 16–19.
- Uras, S., Girosi, F., Verri, A. and Torre, V. (1988). A computational approach to motion perception. *Biol. Cybernet.* **60**, 79–87.
- Watson, A. B. and Ahumada, Jr., A. J. (1985). Models of human visual-motion sensing. *J. Opt. Soc. Am. A* **2**, 322–341.
- Zhang, J. (1992). On the Perception of Visual Motion. Unpublished doctoral dissertation, University of California, Berkeley.
- Zhang, J. (1993). A motion segregation model implemented by V1 mechanisms. *Soc. Neurosci. Abstr.* **19**, 868.
- Zhang, J. and Wu, S. (1990). Structure of visual perception. *Proc. Natl. Acad. Sci. USA* **87**, 7819–7823.