Communicated by Richard A. Andersen

Coherent Compound Motion: Corners and Nonrigid Configurations

Steven W. Zucker*
Lee Iverson
Robert A. Hummel†
Computer Vision and Robotics Laboratory,
McGill Research Centre for Intelligent Machines,
McGill University, Montréal, Québec H3A 2A7 Canada

Consider two wire gratings, superimposed and moving across each other. Under certain conditions the two gratings will cohere into a single, compound pattern, which will appear to be moving in another direction. Such coherent motion patterns have been studied for sinusoidal component gratings, and give rise to percepts of rigid, planar motions. In this paper we show how to construct coherent motion displays that give rise to nonuniform, nonrigid, and nonplanar percepts. Most significantly, they also can define percepts with corners. Since these patterns are more consistent with the structure of natural scenes than rigid sinusoidal gratings, they stand as interesting stimuli for both computational and physiological studies. To illustrate, our display with sharp corners (tangent discontinuities or singularities) separating regions of coherent motion suggests that smoothing does not cross tangent discontinuities, a point that argues against existing (regularization) algorithms for computing motion. This leads us to consider how singularities can be confronted directly within optical flow computations, and we conclude with two hypotheses: (1) that singularities are represented within the motion system as multiple directions at the same retinotopic location; and (2) for component gratings to cohere, they must be at the same depth from the viewer. Both hypotheses have implications for the neural computation of coherent motion.

1 Introduction

Imagine waves opening onto a beach. Although the dominant physical direction is inward, the visual impression is of strong lateral movement. This impression derives from the interaction between the crests of waves

*Fellow, Canadian Institute for Advanced Research.
†Courant Institute of Mathematical Sciences and Center for Neural Science, New York University, New York, NY, and Vrije University, Amsterdam, The Netherlands.

Neural Computation **2, 44–57** (1990)

adjacent in time, and is an instance of a much more general phenomenon: whenever partially overlapping (or occluding) objects move with respect to one another, the point where their bounding contours intersect creates a singularity (Zucker and Casson 1985). Under certain conditions this singularity represents a point where the two motions can cohere into a compound percept, and therefore carries information about possible occlusion and relative movement. Another example is the motion of the point of contact between the blades of a closing scissors; the singular point moves toward the tip as the scissors are closed.

The scissors example illustrates a key point about coherent motion: hold the scissors in one position and observe that it is possible to leave the singular point in two different ways, by traveling in one direction onto one blade, or in another direction onto the other blade. Differentially this corresponds to taking a limit, and intuitively leads to thinking of representing the singular point as a point at which the contour has two tangents. Such is precisely the representation we have suggested for tangent discontinuities in early vision (Zucker *et al.* 1989), and one of our goals in this paper is to show how it can be extended to coherent motion computations.

The previous discussion was focused on two one-dimensional contours coming together, and we now extend the notions of singular points and coherent motion to two-dimensional (texture) patterns. In particular, if a "screen" of parallel diagonal lines is superimposed onto a pattern of lines at a different orientation, then a full array of intersections (or singular points) can be created. The proviso, of course, is that the two patterns be at about the same depth; otherwise they could appear as two semitransparent sheets. Adelson and Movshon (1982) extended such constructions into motion, and, using sinusoidal gratings, showed that coherent compound motion can arise if one pattern is moved relative to the other.

To illustrate, suppose one grid is slanted to the left of the vertical, the other to the right, and that they are moving horizontally in opposite directions. The compound motions of each singular point will then cohere into the percept of a rigid texture moving vertically. Thus the compound pattern can be analyzed in terms of its component parts.

But compound motion arises in more natural situations as well, and gives rise to coherent motion that is neither rigid nor uniform. Again to illustrate, superimposed patterns often arise in two different ways in densely arbored forest habitats (e.g., Fig. 2 in Allman *et al.* 1985). First, consider an object (say a predator) with oriented surface markings lurking in the trees; the predator's surface markings interact with the local orientation of the foliage to create a locus of singular points. A slight movement on either part would create compound motion at these points, which would then cohere into the predator's image. Thus, singular points and coherent motion are useful for separating figure from ground. More complex examples arise in this same way, e.g., between nearby trees

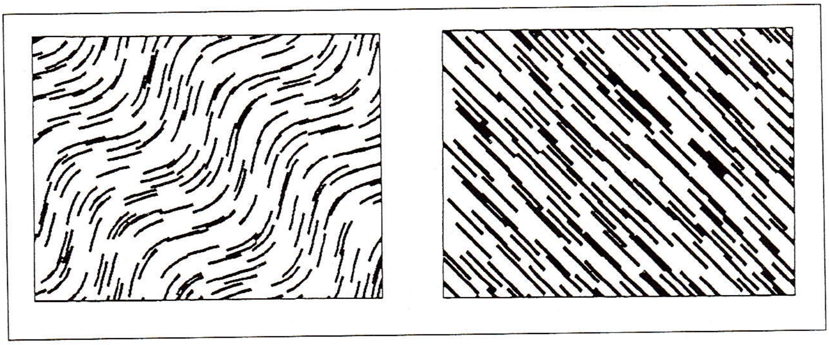

Figure 1: Illustration of the construction of smooth but nonuniform coherent motion patterns. The first component pattern (left) consists of a field of displaced sinusoinal curves, oriented at a positive angle (with respect to the vertical), while the second component consists of displaced parallel lines (right) oriented at a negative angle. The two patterns move across each other in opposite directions, e.g., pattern (left) is moved to the left, while pattern (right) is moved to the right. Other smooth functions could be substituted for either of these.

(Fig. 2), for which three coherent interpretations are possible (in addition to the noncoherent, transparent one):

1. *Two-dimensional sliding swaths*, or a flat display in which the compound motion pattern appears to be a flat, but nonrigid rubber sheet that is deforming into a series of alternating wide strips, or swaths, each of which moves up and down at what appears to be a constant rate with "elastic" interfaces between the strips. The swath either moves rapidly or slowly, depending on the orientation of the sinusoid, and the interfaces between the swaths appear to stretch in a manner resembling viscous flow. The situation here is the optical flow analog of placing edges between the "bright" and the "dark" swaths on a sinusoidal intensity grating.

2. *Three-dimensional compound grating*, in which the display appears to be a sinsusoidally shaped *staircase* surface in depth on which a cross-hatched pattern has been painted. The staircase appears rigid, and the cross-hatched pattern moves uniformly back and forth across it. Or, to visualize it, imagine a rubber sheet on which two bar

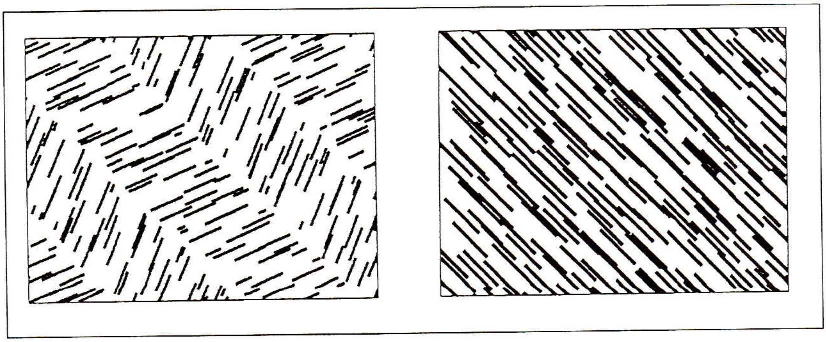

Figure 3: Illustration of the construction of nonuniform coherent motion patterns with discontinuities. The first component pattern (left) consists of a field of displaced triangular curves, oriented at a positive angle, while the second component again consists of displaced parallel lines (right) oriented at a negative angle. The two patterns move across each other as before. Again, other functions involving discontinuities could be substituted for these.

2.2 Triangular Variation. Replacing the sinusiodal grating with a triangular one illustrates the emergence of percepts with discontinuities, or sharp corners. The same three percepts are possible, under the same display conditions, except the smooth patterns in depth now have abrupt changes, and the swaths in (1) have clean segmentation boundaries between them (see Figs. 3 and 4). Such discontinuity boundaries are particularly salient, and differ qualitatively from patterns with high curvatures in them (e.g., high-frequency sinusoids). The subjective impression is as if the sinusoidal patterns give rise to an elastic percept, in which the imaged object stretches and compresses according to curvature, while the triangular patterns give rise to sharp discontinuities.

2.3 Perceptual Instability. To determine which of these three possible percepts are actually seen, we implemented the above displays on a Symbolics 3650 Lisp Machine. Patterns were viewed on the console as black dots on a bright white background, with the sinusoid (or triangular wave) constructed as in Figure 1. The patterns were viewed informally by more than 10 subjects, either graduate students or visitors to the laboratory, and all reported a spontaneous shift from one percept to another. Percepts (1) and (2) seemed to be more common than (3), but individual variation was significant. The spontaneous shifts from

3 Local Analysis of Moving Intersections

Given the existence of patterns that exhibit nonuniform compound motion, we now show how the characterization of rigid compound motion can be extended to include them. To begin, observe that one may think of compound motion displays either as raw patterns that interact, or as patterns of moving "intersections" that arise from these interactions. Concentrate now on the intersections, and imagine a pattern consisting of gratings of arbitrarily high frequency, so that the individual undulations shrink to lines. Each intersection is then defined by two lines, and the distribution of intersections is dense over the image. (Of course, in realistic situations only a discrete approximation to such dense distributions of intersections would obtain.)

Now, concentrate on a typical intersection, whose motion we shall calculate. (Observe that this holds for each point in the compound image.) The equations for the lines meeting at a typical intersection $\mathbf{x} = (x, y)$ can be written

$$\mathbf{n_1} \cdot \mathbf{x} = c_1 + v_1 t$$

$$\mathbf{n_2} \cdot \mathbf{x} = c_2 + v_2 t$$

where $\mathbf{n}_i$, $i = 1, 2$ are the normals to the lines in the first and the second patterns, respectively, c_i are their intercepts, and v_i are their (normal) velocities. Observe that the simultaneous solution of these equations is equivalent to the Adelson and Movshon (1982) "intersection of constraints" algorithm (their Fig. 1). In matrix form we have

$$\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 + v_1 t \\ c_2 + v_2 t \end{pmatrix}$$

We can rewrite this equation as

$$N\mathbf{x}(t) = \mathbf{c} + t\mathbf{v}$$

Differentiating both sides with respect to t, we obtain

$$N\dot{\mathbf{x}}(t) \; = \; \mathbf{v}$$

where $\mathbf{v} = (v_1, v_2)^t$.

Thus, the velocity of the intersection of two moving lines can be obtained as the solution to a matrix equation, and is as follows (from Cramer's rule):

$$\dot{x}(t) \;=\; \frac{v_1 n_{22} - v_2 n_{12}}{\Delta}$$

$$\dot{y}(t) \;=\; \frac{-v_1 n_{21} + v_2 n_{11}}{\Delta}$$

where Δ is the matrix determinant, $\Delta = n_{11}n_{22} - n_{12}n_{21}$, and $\dot{\mathbf{x}}(t) \;=\; [\dot{x}(t), \dot{y}(t)]$.

Before beginning, however, we must stress that there is not yet sufficent information to state precisely how the computation of compound motion is carried out physiologically, or what the precise constraints are for coherence. The analysis in the previous section represents an idealized mathematical competence, and its relationship to biology remains to be determined. Nevertheless, several observations are suggestive. First, it indicates that one need not try to implement the graphical version of the Adelson and Movshon (1982) "intersection of constraints" algorithm literally, but, now that the mathematical requirements are given, any number of different implementations become viable formally. Biologically it is likely that the computation involves several stages, and the evidence is that initial measurements of optical flow are provided by cells whose receptive fields resemble space-time filters, tuned for possible directions of (normal) motion (Movshon *et al.* 1985). Abstractly the filters can be thought of as being implemented by (e.g.) Gabor functions, truncated to local regions of space–time. Such filters provide a degree of smoothing, which is useful in removing image quantization and related affects, but which also blurs across distinctions about which filter (or filters) is (are) signaling the actual motion at each point. In fact, because of their broad tuning, many are usually firing to some extent. An additional selection process is thus required to refine these confused signals, and it is in this selection process that the inappropriate regularization has been postulated to take place.

To illustrate, a selection procedure for compound motion was proposed by Heeger (1988) from the observation that a translating compound grating occupies a tilted plane in the frequency domain. (This comes from the fact that each translating sinusoidal grating occupies a pair of points along a line in spatial-frequency space; the plane is defined by two lines, one from each component grating.) After transforming the Gabor filters' responses into energy terms, Heeger's selection process reduces to fitting a plane. However, the fitting cannot be done pointwise; rather, an average is taken over a neighborhood, effectively smoothing nearby values together. This is permissible in some cases, e.g., for the planar, rigid patterns that Adelson and Movshon studied. But it will fail for the examples in this paper, rounding off the corners within the triangle waves. It cannot handle transparency either, because a single value is enforced at each point (only one plane can be fit). Other variations in this same spirit, based on Tikhonov regularization or other ad hoc (e.g., "winner-take-all") ideas, differ in the averaging that they employ, but still impose smoothness and single-valuedness on the solution (Bulthoff *et al.* 1989; Wang *et al.* 1989; Yuille and Grzywacz 1988). They cannot work in general.

A different variation on the selection procedure relaxes the requirement that only a single value be assigned to each position, incorporates a highly nonlinear type of smoothing, and is designed to confront discontinuities directly. It is best introduced by analogy with orientation selection compound motion as well. That they further provide the basis for interpolation (Zucker and Iverson 1987) and for defining regions of coherence also seems likely.

The triangle-wave example deserves special attention, since it provides a bridge between the orientation selection and optical flow computations. In particular, for nonsingular points on the triangle wave, there is a single orientation and a single direction-of-motion vector. Thus the compound motion computation can run normally. However, at the singular points of the triangle wave there are two orientations (call them n_α and n_β); each of these defines a compound motion with the diagonal component (denoted simply n). Thus, in mathematical terms, there are three possible ways to formulate the matrix equation, with $(n_\alpha, n), (n_\beta, n)$, and (n_α, n_β). The solutions to the first two problems define the two compound motion vectors that define the corner, while the third combination simply gives the translation of the triangle wave at the singular point. In summary, we have:

Conjecture 1. *Singularities are represented in visual area MT analogously to the way they are represented in V1; that is, via the activity of multiple neurons representing different direction-of-motion vectors at about the same (retinotopic) location.*

We thus have that coherent pattern motion involves multiple data concerning orientation and direction at a single retinotopic location, but there is still a remaining question of depth. That depth likely plays a role was argued in the Introduction, but formally enters as follows. Recall that the tilted plane for rigid compound motion (e.g., in Heeger's algorithm) resulted from the combination of component gratings. But a necessary condition for physical components to belong to the same physical object is that they be at the same depth, otherwise a figure/ground or transparency configuration should obtain. MT neurons are known to be sensitive to depth, and Allman *et al.* (1985) have speculated that interactions between depth and motion exist. We now refine this speculation to the conjecture that

Conjecture 2. *The subpopulation of MT neurons that responds to compound motion agrees with the subpopulation that is sensitive to zero (or to equivalent) disparity.*

There is some indirect evidence in support of this conjecture, in that Movshon *et al.* (1985) (see also Rodman and Albright 1988) have reported that only a subpopulation of MT neurons responds to compound pattern motion, and Maunsell and Van Essen (1983) have reported that a subpopulation of MT neurones is tuned to zero disparity. Perhaps these are the same subpopulations. Otherwise more complex computations relating depth and coherent motion will be required.

As a final point, observe that all of the analysis of compound motion was done in terms of optical flow, or the projection of the (3-D) velocities

Zucker, S. W., and Iverson, L. 1987. From orientation selection to optical flow. *Comput. Vision, Graphics, Image Process.* **37**, 196–220.

Zucker, S. W., David, C., Dobbins, A., and Iverson, L. 1988. The organization of curve detection: Coarse tangent fields and fine spline coverings. *Proc. Second Int. Conf. Comput. Vision*, IEEE catalog no. 88CH2664-1, 568–577. Tarpon Springs, FL.

Received 10 July 1989; accepted 8 January 1990.