Coherent Compound Motion: Corners & Non-Rigid Configurations

Steven W. Zucker* Lee Iverson Robert A. Hummel+

TR–89-9 1989/6/27

Computer Vision and Robotics Laboratory
McGill Research Centre for Intelligent Machines
McGill University
Montréal, Québec, Canada

* Senior Fellow, Canadian Institute for Advanced Research.

+ Courant Institute of Mathematical Sciences and Center for Neural Science, New York University; and Vrije University, Amsterdam

Postal Address: 3480 University Street, Montréal, Québec, Canada H3A 2A7

Telephone: (514) 398 6319 Telex: 05 268510 FAX: (514) 398 4470

Network Address: mcrcim@larry.mcrcim.mcgill.edu

Coherent Compound Motion:
Corners & Non-Rigid Configurations

Steven W. Zucker Lee Iverson Robert A. Hummel

Abstract

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 non-uniform, non-rigid, and non-planar percepts. Most significantly, they also can define percepts with singularities (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 which 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: (i) that singularities are represented within the motion system as multiple directions at the same retinotopic location; and (ii) 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.

Acknowledgements

This research was sponsored by NSERC grant A4470 and by AFOSR grant 89-0260. We thank Allan Dobbins for comments.

1989/6/27

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 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, Dobbins, and Iverson, 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 1-dimensional contours coming together, and we now extend the notions of singular points and coherent motion to 2-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 semi-transparent 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 undergoing flexible or different motions, and suggests that natural coherent motion should not be limited to that arising from rigid, planar objects; non-rigid and singular configurations should arise as well. Second, different layers of forest will interact to create textures of coherent motion under both local motion and an observer's movement; distinguishing these coherent motion displays from (planar) single textures (e.g., a wallpaper pattern) can also utilize information about depth.

Thus non-rigid pattern deformations, discontinuities, and depth matter, and our first contribution in this paper is to introduce a new class of visual stimuli for exhibiting them. The stimuli build on the planar, rigid ones previously studied by Adelson and Movshon [1982], but significantly enlarge the possibilities for psychophysical, physiological, and computational studies. In particular, the perceptual salience of singular "corners" within them implies that algorithms for the neural computation of coherent motion require significant modification from those currently available. We propose that a multiple tangent representation, known to be sufficent to represent tangent discontinuities in orientation selection, can be extended to handle them, and show how such ideas are consistent with the physiology of visual area MT. Finally, the interaction of direction-of-motion and depth is briefly considered.

2. Nonuniform Coherent Motion Displays

Compound motion displays are created from two patterns, denoted P_1 and P_2 where, for the Adelson and Movshon displays, the P_i were sinsusoidal gratings oriented at θ_1° and θ_2°, respectively. Since we shall be primarily interested in the geometry of these patterns, observe that patterns of parallel curves work as well as the sinusoidal gratings, so the components can be thought of as square waves (alternating black and white stripes) oriented at different angles. Random dot Moiré patterns ("Glass patterns") work as well [Zucker and Iverson, 1987], and we now show that patterns which are not constant in the direction orthogonal to their orientation also work. It is this new variation (in the orthogonal direction) that introduces non-uniformities into the coherent motion display. We consider two non-uniform patterns, one based on a sinusoidal variation, and the other on a triangular variation. As we show, these illustrate the variety of non-rigid and singular patterns that can arise naturally.

2.1 Sinusiodal Variation

The first non-uniform compound motion pattern is made by replacing one of the constant patterns with a variable one, say a grating composed of displaced sinusoids rather than lines (Fig. 1). Note that this is different than the Adelson and Movshon display, because now the sinusoidal variation is in position and not in intensity. Sliding the patterns across one another, the result is a non-constant motion field (Fig. 2), for which three coherent interpretations are possible (in addition to the non-coherent, transparent one):

1. 2-dimensional sliding swaths, or a flat display in which the compound motion pattern appears on a non-rigid but flat rubber sheet which 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. 3-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 gratings have been painted to form a cross-hatched grating. Now, let a sinusoidally-shaped set of rollers be brought in from behind, and let the rubber sheet be stretched over the rollers. The apparent motion corresponds to the rollers moving back and forth under the sheet.

3. 3-dimensional individual patterns, in which the display appears as in (2), but only with the sinusoidal component painted onto the staircase surface. The second, linear grating appears separate, as if it were projected from a different angle. To illustrate with an intuitive example, imagine a sinusoidal hill, with trees casting long, straight shadows diagonally across it. The sinusoidal grating then appears to be rigidly attached to the hill, while the "shadow" grating appears to move across it.

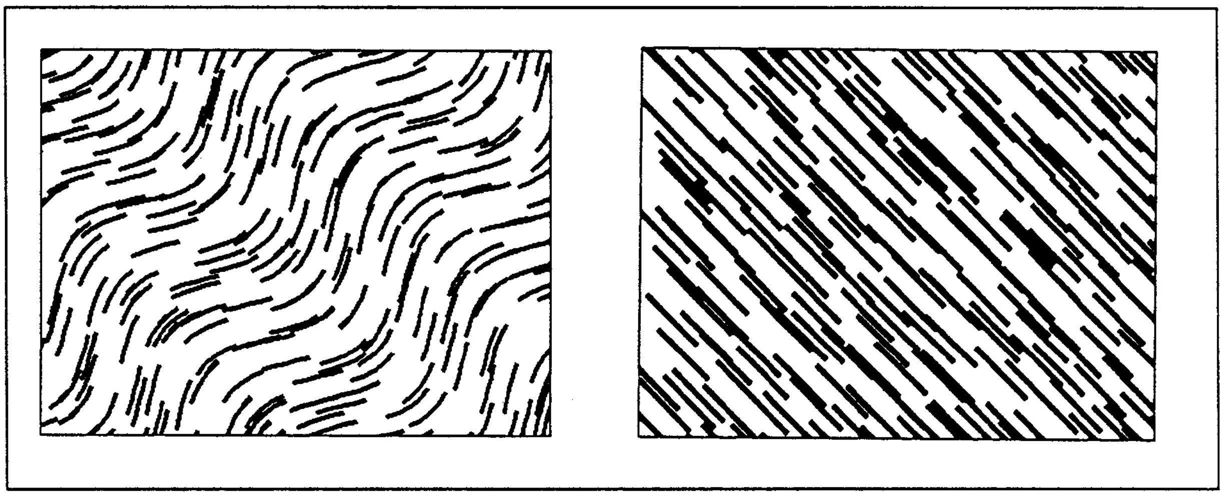

Figure 1 Illustration of the construction of smooth but non-uniform coherent motion patterns. The first component pattern (left) consists of a field of displaced sinusoinal curves, oriented at a positive angle (with respect to the vertical), while the second component consists of displaced parallel lines (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.

2.2 Triangular Variation

Replacing the sinusiodal grating with a triangular one illustrates the emergence of percepts with singularities (tangent discontinuities). 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 Fig. 3 and 4. Such discontinuity boundaries are particularly salient, and differ qualitatively from patterns with high curvatures in them (e.g., high-frequency sinusoids).

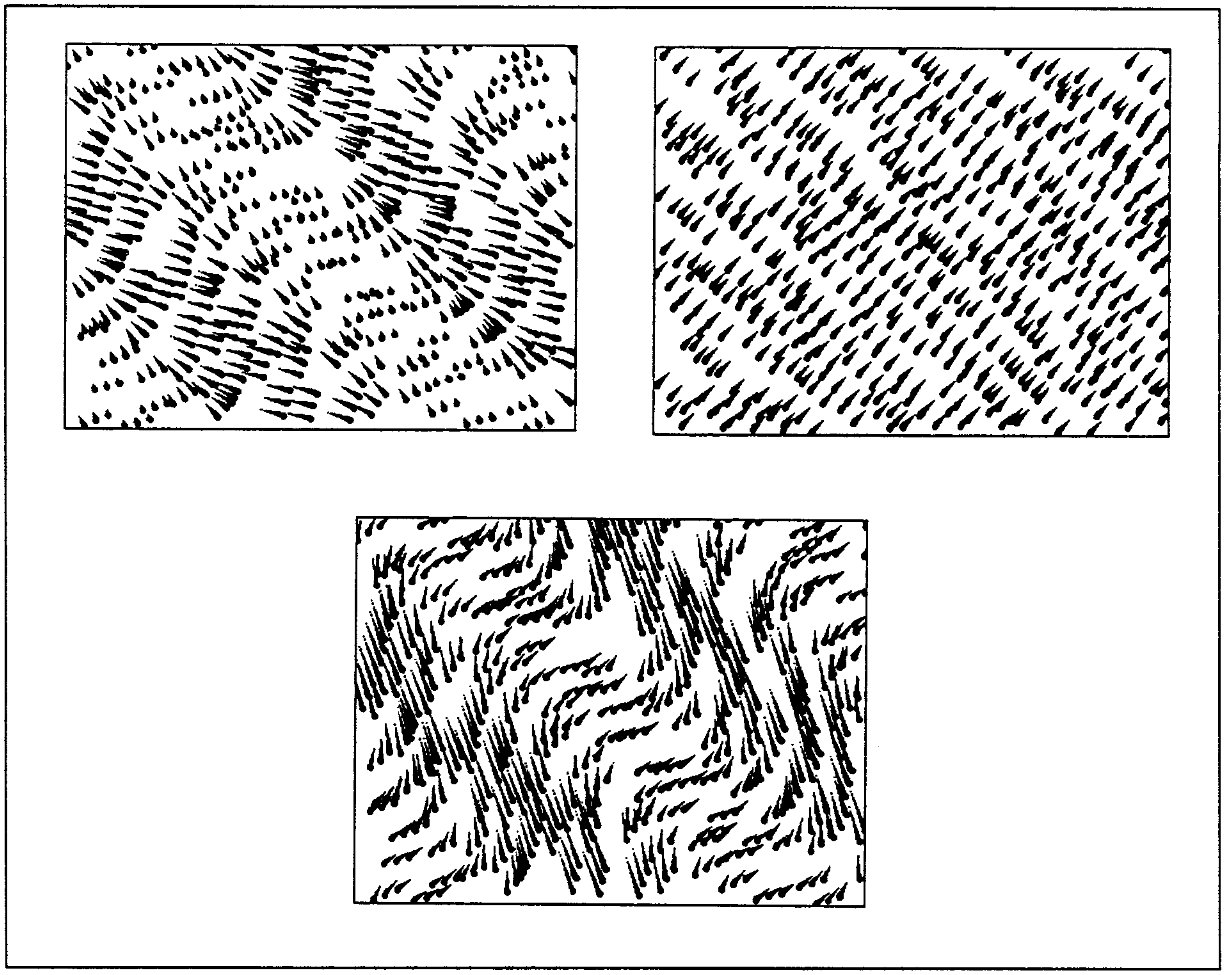

Figure 2 Calculation of the flow fields for the patterns in Fig. 1: (upper left) the normal velocity to the sinusoidal pattern; (upper right) the normal velocity to the line pattern; (bottom) the compound velocity.

3. Local Analysis of Moving Intersections

Given the existence of patterns that exhibit non-uniform 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 now 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:

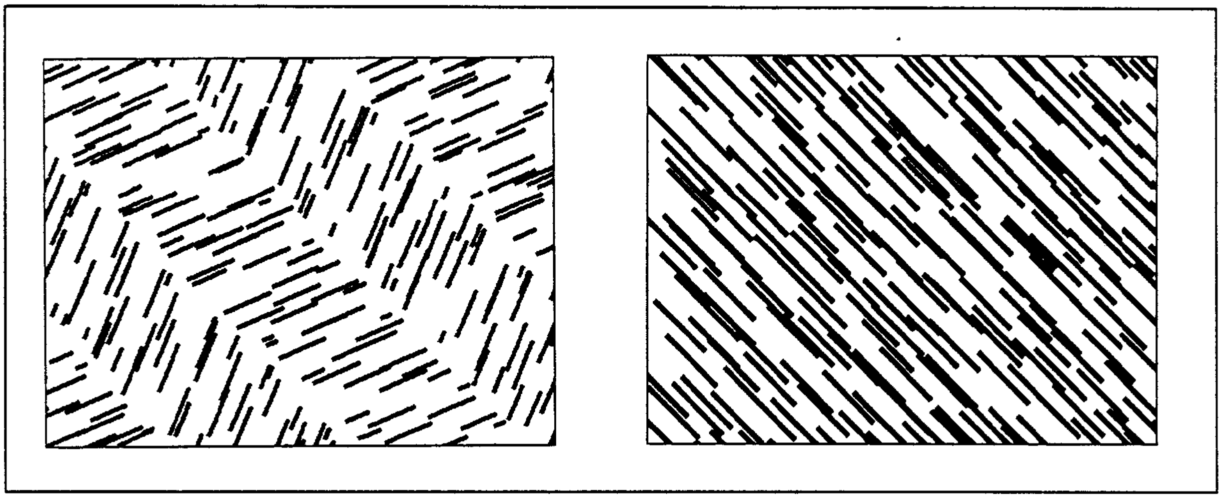

Figure 3 Illustration of the construction of non-uniform coherent motion patterns with discontinuities. The first component pattern (left) consists of a field of displaced triangular curves, oriented at a positive angle, while the second component again consists of displaced parallel lines (right) oriented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

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

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

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:

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

where $(n_{11}, n_{12}) = \mathbf{n}_1$ and $(n_{21}, n_{22}) = \mathbf{n}_2$. 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):

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

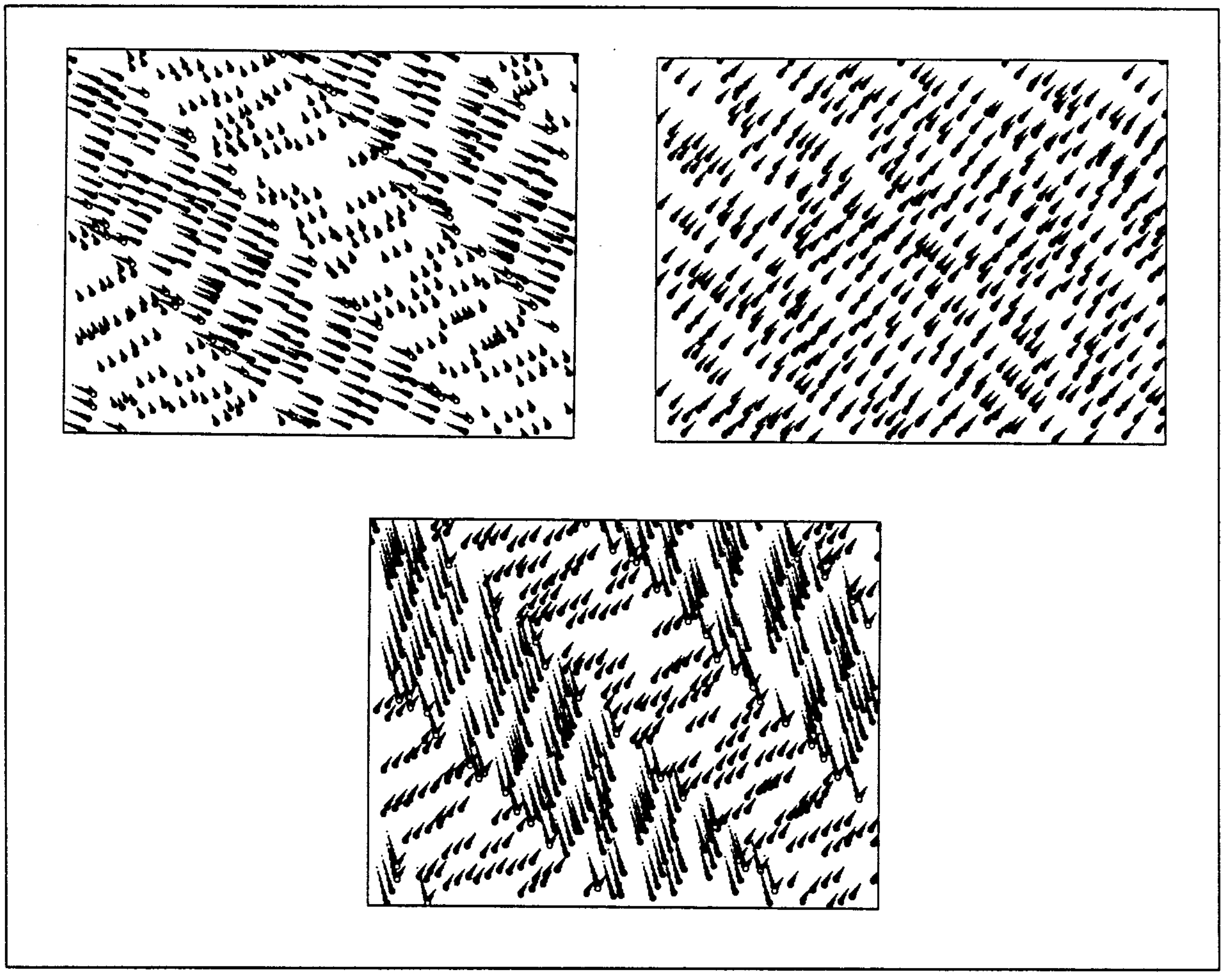

Figure 4 Calculation of the flow fields for the patterns in Fig. 3: (upper left) the normal velocity to the triangular pattern; (upper right) the normal velocity to the line pattern; (bottom) the compound velocity. Velocity vectors at the singular points of the triangle component are shown with small open circles, indicating that two directions are associated with each such point. In (c) the open circles indicate what we refer to as the singular points of coherent motion, or those positions to which two compound motion vectors are rect.

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

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

Several special cases deserve comment. Suppose that $\mathbf{n}_1$ and $\mathbf{n}_2$ are perpendicular, so that the two lines meet at a right angle. Once again, assume that the normal velocities of the two lines are v_1 and v_2 respectively. Then the velocity of the intersection is readily seen to be the vector sum of the velocities of the two lines:

$$\dot{\mathbf{x}}(t) \;=\; v_1 \cdot \mathbf{n}_1 \;+\; v_2 \cdot \mathbf{n}_2.$$

The simplest case involving a distribution of intersections is two sets of parallel lines, each set having orientations given by $\mathbf{n}_1$ and $\mathbf{n}_2$, and moving with a uniform (normal) velocity of v_1 and v_2, respectively. Then all of the intersections will have the same velocity, given by the solution $\dot{\mathbf{x}}$ to the matrix equation. As Adelson and Movshon found, the overall percept in this situation is a uniform motion of $\dot{\mathbf{x}}$.

More generally, as we showed there may be many lines and edges, oriented and moving as a function of their position. Thus there will be many intersections moving according to the above matrix equation. If the line elements (with normals $\mathbf{n}_1$ and $\mathbf{n}_2$) are associated with objects that are themselves moving with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, then the normal velocities of the lines are obtained from the projections $v_1 = \mathbf{v}_1 \cdot \mathbf{n}_1$ and $v_2 = \mathbf{v}_2 \cdot \mathbf{n}_2$. The velocity of the intersection point then satisfies the matrix equation $N\dot{\mathbf{x}} = (v_1, v_2)$ at each instant t.

4. Implications for Neural Computation

It it widely held that theories of motion need not treat discontinuities directly, and that "segmentation" is a problem separate from motion. This view has lead to a rash of "regularized" algorithms with three key features: (1) smoothing is uniform and unconditional; (2) single, unique values are demanded as the solution at each point; and (3) discontinuities, if addressed at all, are relegated to an adjoint process [Bulthoff, Little, and Poggio, 1989; Wang, Mathur, and Koch, 1989; Yuille and Grzywacz, 1988]. We believe that all three of these features need modification, and submit that the current demonstrations are evidence against them; if regularization-style algorithms were applied to the sine- and the triangle-wave coherent motion patterns, the smoothing would obscure the differences between them. To properly treat these examples, algorithms must be found in which discontinuities are naturally localized and smoothing does not cross over them. Furthermore, single values should not be required at each position, but rather representations that permit multiple values at a position should be sought. Such multiple-valued representations are natural for transparency, and, as we show below, are natural for representing discontinuities as well.

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 use the Adelson and Movshon [1982] "intersection of constraints" algorithm literally, but, now that the mathematical requirements are given, any number of different 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, Adelson, Gizzi, and Newsome, 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 is to signal the actual motion. An additional selection process is thus required to refine these confused signals, and it is in this selection process that the inappropriate regularization takes 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 non-linear type of smoothing, and is designed to confront discontinuities directly. It is best introduced by analogy with orientation selection [Zucker and Iverson, 1987]. Consider a static triangle wave. Zucker *et al.* [1988, 1989] propose that the goal of orientation selection is a coarse description of the local differential structure, through curvature, at each position. It is achieved by postulating an iterative, possibly inter-columnar, process to refine the initial orientation estimates (analogous to the initial motion measurements) by maximizing a functional that captures how the local differential estimates fit together. This is done by partitioning all possible curves into a finite number of equivalence classes, and then evaluating support for each of them independently. An important consequence of this algorithm is that, if more than one equivalence class is supported by the image data at a single point, then both enter the final representation at that point. This is precisely what happens at a tangent discontinuity, with the supported equivalence classes containing the curves leading into the discontinuity (example: a static version of the scissors example in the Introduction). Mathematically this corresponds to the Zariski tangent space [Shafarevich, 1977]); and physiologically the multiple values at a single point could be implemented by multiple neurons (with different preferred orientations) firing within a single orientation hypercolumn.

Now, observe that this is precisely the structure obtained for the coherent motion patterns in the Introduction to this paper—singular points are defined by two orientations, each of which could give rise to a compound motion direction. Hence we propose that multiple motion direction vectors are associated with the points of discontinuity, i.e., with the singular points of coherent motion. These points are illustrated in Fig. 4 (bottom) by the open circles. But for such a scheme to be tractable physiologically, we require a neural architecture capable of supporting multiple values at a single retinotopic position. The evidence supports this, since (i) compound motion may well be computed within visual area MT [Movshon *et al.*, 1985; Rodman and Albright, 1988], and (ii) there is a columnar organization (around direction of motion) in MT to support multiple values (i.e., there could be multiple neurons firing within a direction-of-motion hypercolumn) [Albright *et al.*, 1984].

Before such a scheme could be viable, however, a more subtle requirement needs to be stressed. The tuning characteristic for a direction-selective neuron is typically broad, suggesting that multiple neurons would typically be firing within a hypercolumn. Therefore, exactly as in orientation selection, some non-local processing would be necessary to focus the firing activity, and to constrain multiple firings to singularities. In orientation selection we proposed that these non-local interactions be implemented as inter-columnar interactions [Zucker *et al.*, 1989]; and, again by analogy, now suggest that these non-local interactions exist for 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 non-singular points on the triangle wave, there is a single orientation and a single direction-of-motion vector. Thus 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 motions discussed in the paper, while the third combination simple gives the translation of the triangle wave at the singular point.

In summary, we conjecture that

> **Conjecture 1:** *Singularities are represented in visual area MT analogously to the way they are represented in V1; i.e., 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] has 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]) has reported that only a subpopulation of MT neurons respond to compound pattern motion, and Maunsell and Van Essen [1983] have reported that a subpopulation of MT neurones are tuned to zero disparity. Perhaps these are the same subpopulations. Otherwise more complex computations relating depth and coherent motion will be required.

5. References

Adelson, E.H., and Movshon, J.A., 1982, Phenomenal coherence of moving visual patterns, *Nature*, **200**, 523 - 525.

Albright, T.L., Desimone, R., and Gross, C., 1984, Columnar organization of directionally delective cells in visual area MT of the macaque, *J. Neurophysiol.*, **51**, 16 - 31.

Allman, J., Miezin, F., and McGuinness, 1985, Direction- and velocity-specific responses from beyond the classical receptive field in the Middle Temporal area (MT). *Perception* **14**, 85 - 105.

Bulthoff, H., Little, J., and Poggio, T., 1989. A parallel algorithm for real time computation of optical flow *Nature*, **337**, 549 - 553.

Heeger, D., Optical flow from spatio-temporal filters, 1988. *Int. J. Computer Vision*, **1**, 279 - 302.

Maunsell, J.H.R., Van Essen, D., 1983, Functional properties of neurons in middle temporal visual area of macaque monkey. II. Binocular interactions and sensitivity to binocular disparity, *J. Neurophysiol.*, **49**, 1148 - 1167.

Movshon, J.A., Adelson, E.H., Gizzi, M.S., and Newsome, W.T., 1985. The analysis of moving visual patterns, in Chagas, C., Gattass, R., Gross, C. (eds.), *Study Group on Pattern Recognition Mechanisms*, Pontifica Academia Scientiarum, Vatican City.

Rodman, H., and Albright, T., 1988, Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT), preprint, Dept. of Psychology, Princeton University, Princeton, N.J.

Shafarevich, I.R., 1977. *Basic Algebraic Geometry*, Springer-Verlag, New York.

Wang, H.T., Mathur, B., and Koch, C., 1989 Computing optical flow in the primate visual system, *Neural Computation*, **1**, 92 - 103.

Yuille, A., and Grzywacz, N., 1988, The motion coherence theory, *Proc. Second International Conference on Computer Vision*, IEEE Catalog No. 88CH2664-1, Tarpon Springs, Florida, 344 - 353.

Zucker, S.W. and Casson, Y., Sensitivity to change in early optical flow, *Investigative Ophthalmology and Visual Science (Suppl.)*, 1985, **26**(3), 57.

Zucker, S.W., Dobbins, A., and Iverson, L., 1989. Two stages of curve detection suggest two styles of visual computation, *Neural Computation*, **1**, 68 - 81.

Zucker, S.W., and Iverson, L., 1987. From orientation selection to optical flow, *Computer Vision, Graphics, and Image Processing*, **37**, 196 - 220.

Zucker, S.W., David, C., Dobbins, A., and Iverson, L., The organization of curve detection: Coarse tangent fields and fine spline coverings, *Proc. 2nd Int. Conf. on Computer Vision*, Tarpon Springs, Fla, 1988.