International Journal of Computer Vision, 3, 155–175 (1989)

A Mathematical Analysis of the Motion Coherence Theory

ALAN L. YUILLE and NORBERTO M. GRZYWACZ
Harvard University Division of Applied Sciences, G12e Pierce Hall, Cambridge, MA 02138; and Center for Biological Information Processing, Massachusetts Institute of Technology, E25-201, Cambridge, MA 02139

Abstract

In motion perception, there are a number of important phenomena involving coherence. Examples include motion capture and motion cooperativity. We propose a theoretical model, called the motion coherence theory, that gives a possible explanation for these effects [1, 2]. In this framework, the aperture problem can also be thought of as a problem of coherence and given a similar explanation. We propose the concept of a velocity field defined everywhere in the image, even where there is no explicit motion information available. Through a cost function, the model imposes smoothness on the velocity field in a more general way than in previous theories. In this paper, we provide a detailed theoretical analysis of the motion coherence theory. We discuss its relations with previous theories and show that some of them are approximations to it. A second paper [3] provides extensions for temporal coherence and comparisons to psychophysical phenomena. The theory applies to both short-range and long-range motion. It places them in the same computational framework and provides a way to define interactions between the two processes.

1 Introduction

When humans see motion, the perception of how one image feature moves depends on the movement of other features. In particular, the Gestaltists proposed the law of shared common fate. In it, features tend to be perceived as moving coherently [4]. This law is supported by recent psychophysical findings such as motion capture [5] and the cooperativity of the motion system [6, 7]. Furthermore, these percepts of coherent motion can be justified on two computational grounds. First, if two features are close, then they probably belong to the same object, and thus tend to move together. (Work by Yuille and Ullman [8] suggests that there is a statistical relation between the smoothness of motion in the image plane and the rigidity of the object in three-dimensional space.) Second, the measurement of local motion may be inaccurate and an integration of motion information over large areas may help to improve the performance.

These coherent motion percepts are not fully accounted for by the extant theories of visual motion. These theories are limited, since they do not integrate local measurements (see [9], for example), they integrate motion only when the measurements are dense (for example, [10, 11]), or they do not reward coherence (for example, [12]).

In this paper, we propose an alternative theory to account for the perception of coherent motion. We call this theory the *motion coherence theory* [1, 2]. It is shown that, without further assumptions, the theory provides a qualitative solution for the aperture problem [13], which can also be considered as a problem of coherence [14]. Moreover, the theory agrees with experiments by Nakayama and Silverman [15, 16] that investigate variations of the aperture problem, and which are not easily explained by current theories.

(It has been suggested that there are two processes used by humans for motion measurement—see [17–19]. The first, the short-range pro-

cess, deals with motions that span short distances and times. The second, the long-range process, deals with long distances and times. However, the determination of whether a motion is short or long depends complexly on the image—see [20–23]).

This paper is mainly theoretical. It describes the theory for long and short-range motion, proves a number of theoretical results, and discusses the relationship with other theories. A second paper [3] compares the theory to psychophysical experiments in more detail and provides an extension to the theory that includes temporal coherence.

In its present format, this theory does not incorporate mechanisms for motion segmentation and temporal integration of motion information. Theoretical extensions to include motion inertia are described in [3] and are related to psychophysical experiments by [24, 25]. These issues are discussed later in this paper.

This paper's organization is as follows. The next section describes the two stages of the theory, namely, the local motion measurements and their spatial integration. Also, the section explains that this integration, that is, the smoothing, is constrained if the problems solved by the theory are to be well posed and its output spatially restricted. Then, in section 3, we present three short-range motion examples of the theory: motions of isolated features and contours, and optical flows. In the same section, other strategies for short-range motion measurement are discussed. Section 4 discusses how the motion coherence theory deals with long-range motions and compares it to Ullman's minimal mapping theory [12]. Next, in section 5, the combination of different forms of motion measurement is discussed. This suggests a way to combine short- and long-range processes. The following section describes the motion coherence theory's relevance for two important psychophysical phenomena: the solution for the aperture problem and motion capture. Section 7 sketches extensions of the theory; and finally, in section 8, we discuss our conclusions.

2 The Theory

The theory divides the computation of motion into two stages: the measuring and the smoothing stages. In the measuring stage, motion is measured as in any of the previous theories of visual motion and some restrictions are put on the velocity field. For example, in Hildreth's work on short-range motion [11], the normal component of the velocity on a contour is measured and the tangential component is undetermined. In an example of long-range motion, Ullman's cover principle [12] puts a constraint on the possible velocities of long-range motion. The measured velocity might also be computed by spatiotemporal filters or by a correlation model [9,26–30,56], see section 3.3. (For an alternative approach see [31]). In the smoothing stage, velocity field is constructed over the entire visual field; even where no estimates of motion have been made. It is postulated that this velocity field obeys, as strictly as possible, the restrictions found in the measuring stage, and simultaneously, is as smooth as possible. These requirements force the velocities of neighboring features to be similar, while "respecting" the restrictions of the measuring stage.

We now present the theory formally. Let the velocity measurement obtained by the measuring stage at point $\vec{r}_i$ be $M(\vec{U}_i)$, where $\vec{U}_i$ is the true image velocity. The measurement operator M will depend on the measurements. If for example, isolated point-like features undergo short-range motion, then both their velocity components may be measured. For contour motion, on the other hand, the measurements may correspond to the normal component of the velocity field along edge contours. Section 3 discusses several examples of the operator M for short-range motions. For long-range correspondence, the measurements are more complicated and are described in section 4.

The motion coherence theory proposes that the smoothing stage constructs a velocity field, $\vec{v}(\vec{r})$, such that the following functional is minimized for both components of $\vec{v}$

$$\varepsilon(\vec{v}(\vec{r}), \vec{U}_i) = \sum_i [M(\vec{v}(\vec{r}_i)) - M(\vec{U}_i)]^2 + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \qquad (1)$$

where $\lambda > 0$ and $c_m > 0$ are constants. The derivative operator (D^m) is a scalar operator if m is even and a vector operator if m is odd. More precisely

$$(D^{2m}\vec{v}) = \nabla^{2m}\vec{v} \tag{2a}$$

$$(D^{2m+1}\vec{v}) = \vec{\nabla}(\nabla^{2m}\vec{v}) \tag{2b}$$

where $\vec{\nabla}$ is the gradient operator and ∇^2 is the Laplacian operator. Smoothness operators of this type correspond to Tikhonov stabilizers [32–34]. The sum of the first term of the right-hand side of equation (1) is taken over all points where explicit motion information is available.

The precise form of the smoothness operator is important. It determines the form of the interactions between measurements in different parts of the visual field and, in particular, the way these interactions fall off with distance.

2.1 The Form of the Smoothness Function

The parameter λ sets the strength of the smoothing and the c_m ($c_m > 0$) determine its form. There are many possible forms for the smoothing function. In this section, we discuss several conditions that this function must satisfy. We then argue that a good choice is a smoothness function that yields an interaction with the spatial behavior of a Gaussian function.

We impose two criteria on our smoothness function: (i) It must impose enough smoothness to make the problem well posed, and (ii) the interactions between different measurements must fall off to zero at large distances. (See the Discussion section for the case where boundary conditions are used.)

We now state theorems that give necessary and sufficient conditions for (i) and (ii).

THEOREM 1. A necessary and sufficient condition for (i) is that derivatives of higher than first order exist in the smoothing operator.

Proof. This theorem is proved in [35]. Intuitively the amount of smoothness required depends on the dimensionality of the space and the dimensionality of the data. For isolated data points (such as features) in two dimensions an operator with only first-order derivatives does not supply enough smoothness. This would correspond to fitting a membrane surface to isolated data points. The human visual system, however, obtains coherent percepts even for discrete features.

THEOREM 2. A necessary and sufficient condition for the interaction to fall faster than 1/r, where r is the distance from a motion measurement site, is that $c_0 > 0$.

Proof. This theorem is proven in appendix A. (In practice, if $c_0 = 0$, the interaction does not fall at all. Also, for the total contribution of a data point to the computed velocity field to be finite, it is critical that the interaction falls faster than $1/r$.)

We can deduce a corollary

COROLLARY. Smoothing operators involving only first-order derivatives do not satisfy either of the conditions (i) and (ii).

This corollary is important, since previous theories that integrated motion over space [10, 11] used a similar mathematical structure as in equation (1), with only first-order derivatives (although Horn and Schunck suggested the possibility of using Laplacians). Thus, any attempts to extend these theories directly to deal with sparse data will encounter difficulties. It can also be shown, see appendix B, that these difficulties will start to occur even for dense data.

The two theorems say that to satisfy our criteria, the smoothing operator must have positive c_0 and one or more derivatives higher than first order.

For this paper, we choose the smoothing so that the interaction is a Gaussian. This form of interaction was chosen for four reasons: First, it meets the criteria above; second, it generates analytic solutions; third, it has a natural spatial scale; and fourth, it may be an optimal smoothing filter [36–38]. Moreover, variations of the higher-order terms, c_m for $m > 2$, seem to have little effect on the interaction, which thus usually resembles a Gaussian [39]. To obtain a Gaussian interaction we set $c_m = \sigma^{2m}/(m!2^m)$ (see appendix A. This result has also been obtained by Shulman and Aloimonos—private communication).

Smoothing has been traditionally justified in vision as a method for dealing with noise or filling

in sparse data. We would like to stress that smoothing also gives rise to computational efficiency. If computation is done at a coarse scale, then the interpolated result can be used to reduce significantly the complexity of calculation at finer scales.

3 Examples of Short-Range Motions

In this section, we start with three examples of the theory's predictions for short-range motions. The first is given in section 3.1 and deals with motions of isolated features. This example is the simplest, because isolated features may be tracked and their motion determined with precision. Nevertheless, this is not a trivial example, since the phenomenon of motion cooperativity [6, 7] shows that these motions interact.

Section 3.2 gives two other short-range motion examples: motions of contours and optical flows. For the former, we compare the predictions of the motion coherence theory with a computational theory proposed by Hildreth [11]. For the latter, a comparison is made with a computational theory proposed by Horn and Schunck [10].

Finally, section 3.3 discusses other, more biological, examples of how short-range motion may be measured.

3.1 Isolated Features and Motion Cooperativity

This section deals with the simplest form of motion: that of isolated features undergoing short-range motions. The example that we present here shows that the motion coherence theory may account for the cooperativity of the motion system [6, 7]. This cooperativity is demonstrated by an experiment as follows. The stimulus consists of random-dot patterns in which the dots make a random walk, with their direction of motion at each step taken from some distribution. If there is a uniform distribution over all directions, one sees dynamic noise. However, only a slight bias of the range of motion directions may lead to a percept of global motion in the direction of the bias. The theory predicts this behavior, because the small variance of the constructed velocity field may enable the mean motion to be detected.

These experiments are examples of short-range motion. In them, each dot can be tracked and there is no correspondence problem. We assume that in this case, the velocity measurements are the velocities of the dots. Suppose that the dots are at points $\vec{r}_i$ with velocities $\vec{U}_i$. The motion coherence theory suggests minimizing

$$E(\vec{v}(\vec{r}),\vec{U}_i) = \sum_i [(\vec{v}(\vec{r}_i)) - (\vec{U}_i)]^2 + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \qquad (3)$$

Then, it is shown in appendix C that the solution of (3), obtained by standard calculus of variations, has the form

$$\vec{v}(\vec{r}) = \sum_i \frac{\vec{\beta}_i}{2\pi\sigma^2} \exp \frac{-(\vec{r} - \vec{r}_i)^2}{2\sigma^2} \qquad (4)$$

where the β_i are solutions of

$$(\lambda\delta_{ij} + G_{ij})\vec{\beta}_j = \vec{U}_i \qquad (5)$$

where

$$G_{ij} = \frac{1}{2\pi\sigma^2} \exp \frac{-(\vec{r}_j - \vec{r}_i)^2}{2\sigma^2} \qquad (6)$$

The solution will depend on the values of σ and λ. To understand these dependencies, we took two approaches. First, simulations were made of motion cooperativity examples, by solving equation (5) with fixed values of λ and changing values of σ (figure 1), and also, with fixed σ and changing λ (figure 2). Second, we performed analytical calculations with a simplified version of the motion cooperativity paradigms.

If we fix λ (figure 1) and make σ large, then the magnitude of the velocity vectors in the constructed field is small. This is as if the interactions drag empty space, which slows the motions down. On the other limit, that is, for small σ, there is no interaction and the constructed field coincides with the data in the measurement sites. However, over a large intermediate range of σ, the solution is close to the data bias.

If σ is fixed (figure 2) and λ is large, the constructed field is again small. This is due to the c_0 of equation (3), which enforces the interactions'

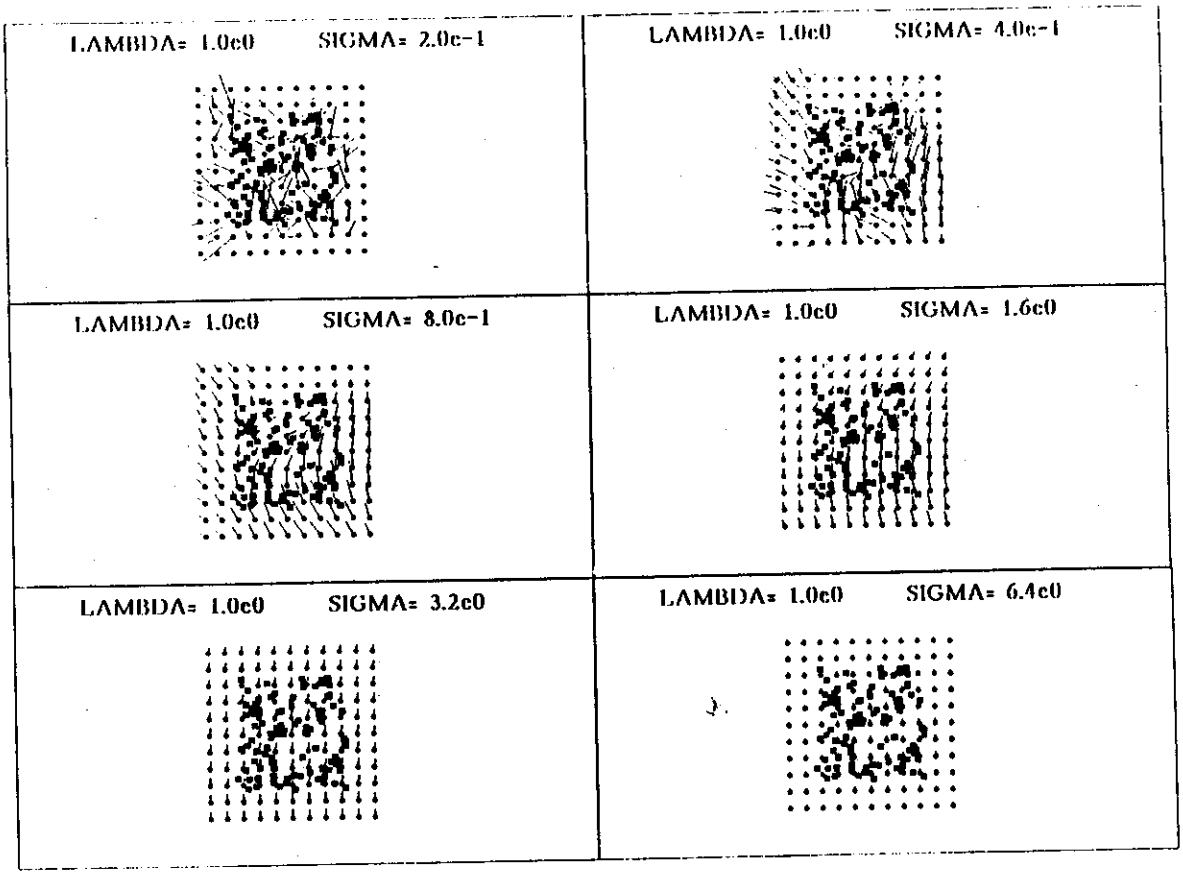

Fig. 1. Motion cooperativity and the spatial extent of the interactions. One hundred dots were randomly placed inside a square region of unit sides; the dot positions are indicated by the little solid squares. Each of these dots was assumed to be moving. The horizontal components of their velocity were obtained from a homogeneous random distribution whose values ranged from −1 to 1. The vertical components had the same distribution plus a constant bias upward of 0.25. The dots' velocities are not shown, but are close to those shown in the upper left panel. The figure shows the velocity field computed by the motion coherence theory by minimizing equation (3) with fixed λ and six values of σ. The computed field is shown as lines attached to little circles positioned in a grid. (However, the solution is continuous, because it is computed analytically as in equation (4); the grid is shown for purposes of illustration.) For small σ, no interactions occur and the output is similar to the input. On the other hand, for large σ, the computed motion is small due to interactions with empty space. However, in a wide range of σ, the computed field at the position of the dots is close to the 0.25 bias. The output coherence in this range is similar to the psychophysical phenomenon of motion cooperativity [6, 7].

locality and leads to small velocities. On the other hand, if λ is small, the solution is noisy, because with little smoothing the problem of minimizing equation (3) becomes ill-conditioned. Once again, for a large intermediate range of λ, a solution that is close to the data bias is obtained. We also verified that these close-to-bias solutions are maintained if σ and λ are varied simultaneously, provided that they are in their correct range.

We analyzed a simplified version of the motion cooperativity paradigms. In this version, the dots are randomly and homogeneously distributed all over the space, that is, they are not constrained to a small region as in figures 1 and 2. Also, this analysis assumes that the density of the dots, ρ, is high ($\rho\pi\sigma^2 \gg 1$) such that they strongly interact.

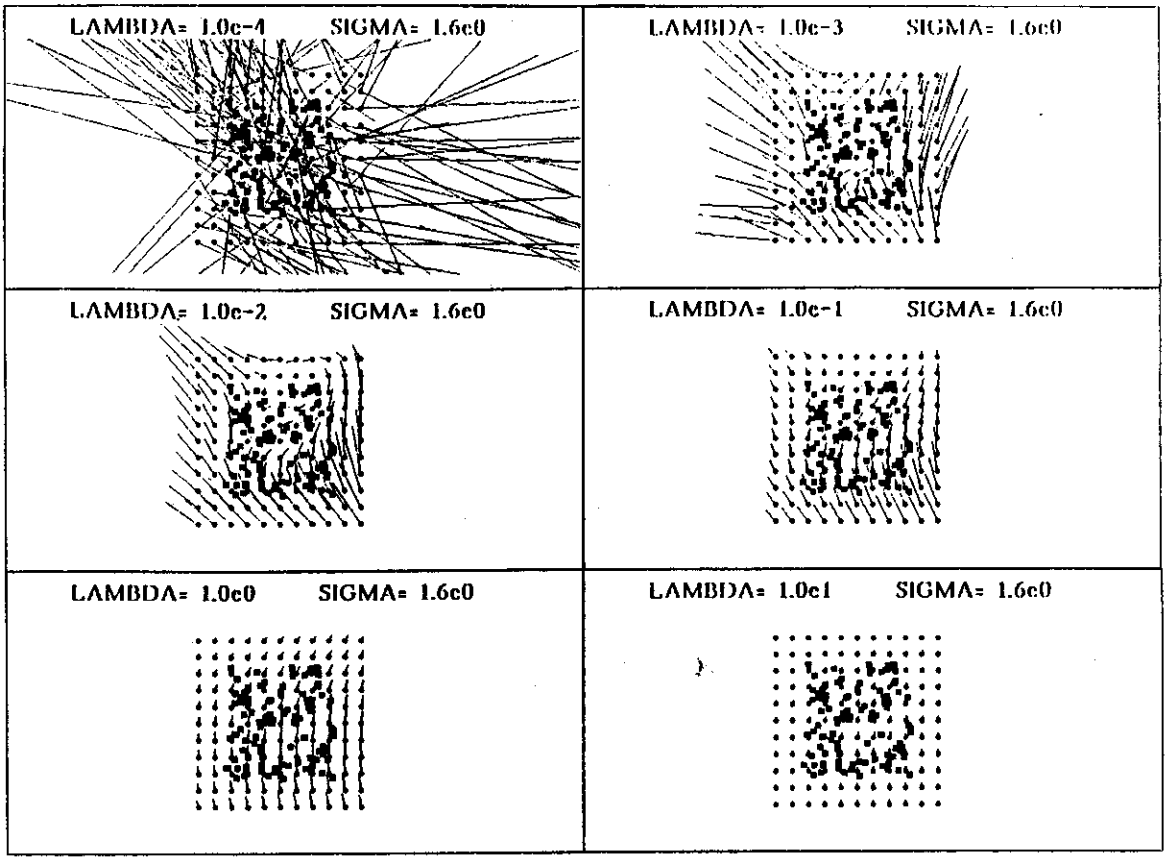

Fig. 2. Motion cooperativity and smoothness. This figure is identical to figure 1, except that this time we fixed σ and varied λ. For small λ, the computed field is noisy, because with little smoothing the problem of minimizing equation (3) becomes ill-conditioned. However, if λ is large, the motion is small. This is due to the c_0 of equation (3) which enforces the interactions' locality and leads to small velocities. Once again, for a large intermediate range of λ, a solution that is close to the input's bias is obtained.

The dots' velocity fields are distributed with a given mean and standard deviation. In appendix C, we show how to minimize (3) given these assumptions. In particular, we analyze how the mean and the variance of the output velocity field depend on the statistics of the input and its dot density. We find that the mean velocity of the output is

$$\langle \vec{v} \rangle = \frac{\rho \langle \vec{U}_i \rangle}{\lambda + \rho} \tag{7}$$

Thus, the constructed field's mean velocity is smaller than the input's mean. But if λ is sufficiently small as compared to the dot density, then these means are arbitrarily close. (This provides a first guiding rule for the choice of λ in a given computation.) This equation is consistent with the high λ limit of figure 2 but not with the high σ limit of figure 1. The discrepancy results because, in this figure, the features were restricted to a small region and the large σ led to interactions with empty space.

The dependence of the output variance, $\mathrm{Var}(\vec{v})$, on the statistics of the input is:

$$\mathrm{Var}(\vec{v}) = \frac{1}{2\pi\sigma^2\rho} \left[\log\left(\frac{\rho + \lambda}{\lambda}\right) - \frac{\rho}{\rho + \lambda} \right] \langle \vec{U}_i^2 \rangle \tag{8}$$

This equation confirms the findings of figures 1 and 2. Not surprisingly the output variance falls with smoothing, λ, and the range of interactions, σ. Also, the variance increases unbounded as the smoothing decreases. Unfortunately, the limit of small σ is not valid in this equation due to the assumption $\rho\pi\sigma^2 \gg 1$, which was used in the derivation. The dependence of the output variance on dot density is weak, but it should be detectable experimentally.

Our model may need to be modified to estimate dynamically (adaptation) λ and σ from the image by a process that decides how much smoothing is necessary and how wide the range of interaction should be. This adaptation may be fooled in experiments by quick changes of dot density.

3.2 Relations with the Theories of Hildreth, and Horn and Schunck

We now discuss how the theory compares to two previous theories for short-range motion: one that deals with contour motions [11] and another that deals with optical flow [10].

Hildreth assumes that contours have been extracted from the image by a previous process, for example, by edge detection. In this case, the velocity information is known only along a contour. Moreover, only the normal component of the velocity field can be measured directly without further assumptions. This is an example of the well-known aperture problem [13], which will be discussed further in section 5. The aperture problem is defined by saying that local measurements of the velocity field on a contour cannot determine the velocity field. Hildreth proposes determining the velocity field by minimizing an energy function defined on the contour C:

$$E_H[\vec{v}(s)] = \int_C [\vec{v}(s)] \cdot \vec{n}(s) - u(s)]^2 + \lambda \int_C \frac{\partial \vec{v}(s)}{\partial s} \cdot \frac{\partial \vec{v}(s)}{\partial s} \qquad (9)$$

with respect to $\vec{v}(s)$, where $\vec{n}(s)$ is the normal vector to the curve and $u(s)$ is the measured normal velocity field as functions of the arc length s. In this formulation the velocity is only smoothed along the contour.

The motion coherence theory can also deal with contour motion. The measurement of the velocity field is the component of velocity perpendicular to the contour $M[\vec{v}(\vec{x})] = \vec{v}(\vec{x}) \cdot \vec{n}(\vec{x})$. The theory then gives

$$E[\vec{v}(\vec{x})] = \int_C [\vec{v}(\vec{x}) \cdot \vec{n}(\vec{x}) - u(\vec{x})]^2 + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \qquad (10)$$

where the first integral is taken along the contour C and the second integral is over all space. The motion coherence theory allows for the motions on nearby objects to affect the perception of the motion on the contour, an effect demonstrated by Nakayama and Silverman [15, 16]. This will be discussed in section 5.

Horn and Schunck [10] consider measuring motion directly from the intensity field. Thus, an image can be considered as a function, $I(\vec{x},t)$, and points are assumed to have the same intensity over time. If the velocity field is taken to be $\vec{v}(\vec{x})$ this gives the constraint

$$\nabla I \cdot \vec{v}(\vec{x}) = -\frac{\partial I}{\partial t} \qquad (11)$$

where ∇I is the gradient of the image intensity. Note that again only one component of the velocity field is available, namely, the component in the direction of the image gradient. Horn and Schunck proposed minimizing a function

$$E_{HS}[\vec{v}(\vec{x})] = \int \left[\nabla I \cdot \vec{v}(\vec{x}) + \frac{\partial I}{\partial t} \right]^2 + \lambda \int \nabla \vec{v}(\vec{x}) \cdot \nabla \vec{v}(\vec{x}) \qquad (12)$$

with respect to $\vec{v}(\vec{x})$, where the integrals are taken over space.

The motion coherence theory generalizes Horn and Schunk's method as follows:

$$E[\vec{v}(\vec{x})] = \int \left[\nabla I \cdot \vec{v}(\vec{x}) + \frac{\partial I}{\partial t} \right]^2 + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \qquad (13)$$

3.3 Motion Measurement, Correlation, and Spatiotemporal Filters

In the theory, as described in the last section, we have been assuming simple types of motion

measurement. We have assumed that the normal components of the velocity field, and spatial and temporal derivatives, can be measured locally and instantaneously for edges and images respectively. Since a camera, or an eye, takes time to capture the intensity of an image, this will not always be true; and if the image is changing quickly, motion blurring will result. A second problem arises for noisy images: the local measurements of the velocity field may be very noisy.

Two alternative ways to reduce the severity of these problems are to use spatiotemporal filters either to look for correlations of the image over space and time or to measure motion energy [9]. There have been a large number of theories for short-range motion based on these approaches [9, 26–30, 56]. These theories assume that motion is constant and in a straight line over the space-time support of the filters. Thus, they provide an implicit solution to the aperture problem, although they can only summate data over the range of the filter. These theories emphasize time and the use of many semi-independent channels. (Recently, we proposed a complete theory to combine the information from the different channels [56]). The concept of different channels allows for simple explanations of transparency effects [56]. A present disadvantage is the limit on spatiotemporal interaction given by the size of the filter.

In any case, methods of this type could be used to make more accurate estimates as input for the motion coherence theory.

4 The Theory for Long-Range Motion

For long-range correspondence, the theory becomes more complicated. For the short-range motion described in previous sections, one component of the velocity field was always known, or could be directly measured. However, in the long-range case, we have only the restriction that features should correspond. Note that if two features correspond, then a velocity vector can be assigned at the position of the first feature based on the difference in spatial positions of the features and the time between frames. Thus for N features we have $N!$ possible matches between frames corresponding to $N!$ possible velocity assignments at the features. In the motion coherence theory, some assignments will correspond to smoother velocity fields than others. The matching between the features is found by requiring the velocity field to be smooth. The smoothing and the correspondence are done simultaneously (figure 3).

To make this precise, suppose we have N features in the two frames at positions $\vec{x}_i$ in the first frame and $\vec{x}_a$ in the second frame. We define a set of binary matching elements V_{ia}, where $V_{ia} = 1$ if the ith feature in the first frame is matched to the ath feature in the second frame, and is 0 otherwise. The constraints on the velocity field can be expressed by a term

$$E_{match} = \sum_{i\,a} V_{ia}[\vec{v}(\vec{x}_i) - (\vec{x}_a - \vec{x}_i)]^2 \qquad (14)$$

where we have normalized the time difference between frames to be unity. If $V_{ia} = 1$, then the feature at $\vec{x}_i$ in the first frame travels to feature $\vec{x}_a$ in the second; and therefore it corresponds to a velocity of $\vec{x}_a - \vec{x}_i$ (with the time normalization). The velocity field at $\vec{x}_i$ is therefore constrained by this value. If $V_{ia} = 0$, then the features are unmatched and no contribution is made to the energy. To obtain the full energy function we add a smoothing term for the velocity field as before to obtain

$$E[V_{ia}, \vec{v}(\vec{x})] = \sum_{i\,a} V_{ia}[v(\vec{x}_i) - (\vec{x}_a - \vec{x}_i)]^2 + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \qquad (15)$$

We minimize this function over V_{ia} and $\vec{v}(\vec{x})$ simultaneously. Thus, the smoothness requirement directly affects the matching in this case. The V_{ia} must be constrained so that features typically make exactly one match. There are two possible ways to do this: To use the cover principle [12] and require that all features have at least one match, or to incorporate a cost in the energy function to bias against too many or too few matches, for example $E_c = \sum_a (\sum_i V_{ia} - 1)^2 + \sum_i (\sum_a V_{ia} - 1)^2$ [40]. The latter may be more consistent with psychophysics, since there are situations in which the cover principle is violated (for example, in the splitting and fusion phenomena see [12, 21])

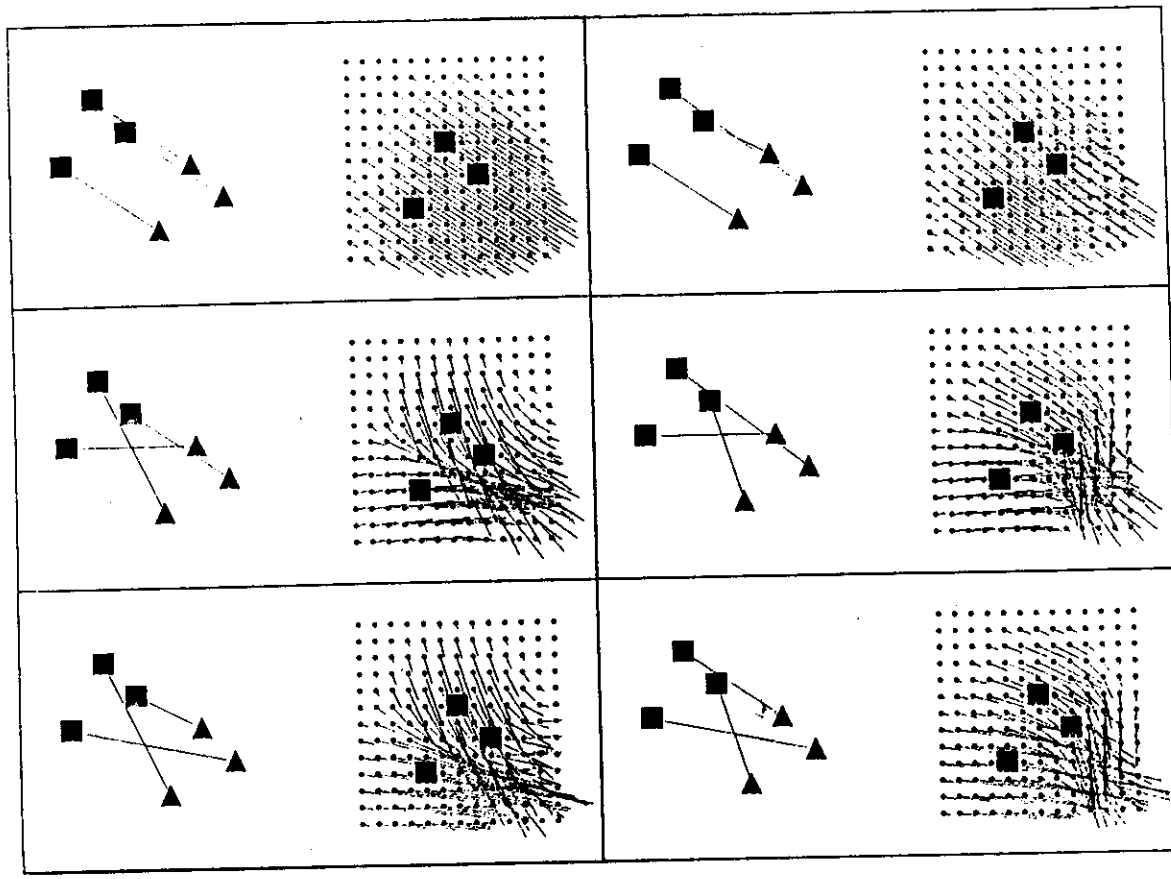

Fig. 3. The theory for long-range motion. Three features underwent an apparent motion in the downward and rightward direction. The positions of the features in the first and second frames of the motion are indicated by the squares and triangles respectively. This motion generates a correspondence problem: To which feature of the first frame does a given feature in the second frame correspond? If one imposes a one-to-one match, there are six ways in which the features could correspond; these are indicated in the six panels. For each correspondence the velocity field that minimizes equation (15) was found and displayed in a grid as in figure 1. The motion coherence theory suggests that the correspondence that leads to the deepest of these minima is what the visual system adopts. In this figure, the adopted correspondence is the one in the upper left panel, which is also the smoothest.

4.1 The Comparison with Minimal Mapping

We now show that the energy function can be transformed into a form that includes the minimal mapping theory [12] as a limiting case. This is desirable, since it is known that minimal mapping theory agrees with a reasonable number of perceptual experiments, and it is plausible that the motion coherence theory will give similar predictions in these cases. There will be situations [1-3], however, involving motion capture for which the motion coherence theory will give different predictions. This is because the minimal mapping theory does not encourage coherence as strongly as the motion coherence theory. We performed psychophysical experiments [3] that were consistent with the motion coherence theory and inconsistent with minimal mapping.

THEOREM. For long-range motions, for which the number of features over time is constant, the motion coherence theory with the cover principle is equivalent to minimizing a cost function

$$E(V_{ia}) = \lambda(\sum_a V_{ia}\, d_{ia})(\lambda\delta_{ij} + G_{ij})^{-1}$$

$$(\sum_b V_{bj} d_{bj})$$

with respect to the V_{ia}, where $d_{ia} = |\vec{r}_i - \vec{r}_a|$. This cost function is similar to the square of the cost function for minimal mapping, $E(V_{ia}) = \sum_{ia} V_{ia} d_{ia}$. These cost functions have the same minima in the limits $\sigma \to 0$ and $\lambda \to \infty$.

Proof. The key observation is that $E(V_{ia}, \vec{v}(\vec{x}))$ is quadratic in $\vec{v}(\vec{x})$. The Euler-Lagrange equations for $\vec{v}(\vec{x})$ will, therefore, be linear and can be solved for as a function of V_{ia}. We now express $E_M(V_{ia})$ in terms of the V_{ia} only.

We first note that the x and y components of the velocity field in (15) do not interact so that we can treat each component separately. Henceforth, we consider the x component only. The Euler-Lagrange equations become

$$\lambda \sum_{m=0}^{\infty} c_m \nabla^{2m} v(\vec{x}) = -\sum_{i\,a} V_{ia}[v(\vec{x}_i) - d_{ia}]\,\delta(\vec{x} - \vec{x}_i) \tag{16}$$

where δ is the Dirac delta function. The c_m are given by $c_m = \sigma^{2m}/(m!2^m)$. This choice ensures that the Green function of the operator $\sum_{m=0}^{\infty} c_m \nabla^{2m}$ is the Gaussian $G(\vec{x}, \sigma)$, in other words

$$\sum_{m=0}^{\infty} \frac{\sigma^{2m}}{(m!2^m)} \nabla^{2m} G(\vec{x},\sigma) = \delta(\vec{x}) \tag{17}$$

Using (17) we see that the solution to (16) is of the form

$$v(\vec{x}) = \sum_i \beta_i G(\vec{x} - \vec{x}_i) \tag{18}$$

Substituting this back into (16) we obtain

$$\lambda \sum_i \beta_i \delta(\vec{x} - \vec{x}_i) = -\sum_{i\,a} V_{ia}[v(\vec{x}_i) - d_{ia}]\delta(\vec{x} - \vec{x}_i) \tag{19}$$

Equating coefficients of the delta functions gives us a set of equations

$$\lambda\beta_i = -\sum_a V_{ia}\, v(\vec{x}_i) + \sum_a V_{ia} d_{ia} \tag{20}$$

Now, we can use $\sum_a V_{ia} = 1$ and substitute for v from (18)

$$\lambda\beta_i = -\sum_j G_{ij}\beta_j + \sum_a V_{ia} d_{ia} \tag{21}$$

where $G_{ij} = G(\vec{x}_i - \vec{x}_j)$. Using the summation convention ($a_i b_i = \sum_{i=1}^{N} a_i b_i$ and $a_i H_{ij} b_j = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i H_{ij} b_j$), and the Kroenecker delta function we find

$$(\lambda\delta_{ij} + G_{ij})\beta_j = \sum_a V_{ia} d_{ia} \tag{22}$$

Now we substitute back for $\vec{v}$ from equations (18) and (20) into the energy function. This gives, using the summation convention,

$$E(V_{ia}) = \lambda^2\beta_i\beta_i + \lambda\beta_i G_{ij}\beta_j \tag{23}$$

Using (22) gives

$$E(V_{ia}) = \lambda\Big(\sum_a V_{ia} d_{ia}\Big)(\lambda\delta_{ij} + G_{ij})^{-1}\Big(\sum_b V_{bj} d_{bj}\Big) \tag{24}$$

In the limit as $\sigma \to 0$, we obtain $\lambda\delta_{ij} + G_{ij} \to \delta_{ij}/2\pi\sigma^2$. Therefore

$$E_{\text{limit}} = \lambda 2\pi\sigma^2\Big(\sum_{ia} V_{ia} d_{ia}\Big)^2 \tag{25}$$

This has the same minimum as a function of V_{ia}, and thus makes the same predictions as the minimal mapping theory, which is given by

$$E_{mm} = \sum_{ia} V_{ia} d_{ia} \tag{26}$$

This similarity provides a new interpretation for the minimal mapping theory. As figure 1 illustrates, the limit $\sigma \to 0$ corresponds to the situation where the locally measured data are most respected. In this case, the constructed velocity field corresponds to sharp peaks centered around the data points and large valleys between these peaks. Such a landscape is most smooth when the peaks are small. Thus, the smoothest choice of correspondence occurs for the smallest velocities, that is, for minimal mapping.

Curiously, we also found that if $\lambda \to \infty$, then the correspondences of the motion coherence theory and the minimal mapping theory coincide. Indeed, in this limit, $\lambda\delta_{ij} + G_{ij} \to \lambda\delta_{ij}$. The reason for the similarity between the theories in this case is that since $c_0 \neq 0$, the constructed field tends to zero. Thus, the energy (12) depends only on the distances between the matched features and is minimal when these distances are minimal.

There is an interesting parallel between this method of eliminating the velocity field and cer-

tain theories of physics. We have shown that the velocity field, which is local, can be eliminated to obtain a nonlocal interaction between the features. The velocity field can therefore be considered merely a fictitious field, which is not essential for the interaction and can be eliminated. The concept of the velocity field is important, however, since it acts as a unifying principle for all types of motion. In a similar way, Wheeler and Feynman [42, 43] considered a system of charged particles interacting via an electromagnetic field. They showed that the electromagnetic field could be eliminated from the system, and hence considered fictitious, leaving a nonlocal interaction between the particles.

5 Combining Different Forms of Measurement

We have illustrated how the motion coherence theory applies to situations where the motion measurements take different forms. We can combine different types of motion measurement into one cost function. Suppose, for example, that we can measure simultaneously the short-range motion of some isolated features (equation 3), motion of contours (10), spatial and temporal derivatives of image intensity (12), and long-range positions of salient features (14). The motion coherence theory would then propose minimizing (through the combination of these equations):

$$\begin{aligned} E(\vec{v}(\vec{x}), V_{ia}) = {} & \psi_1 \sum_i [\vec{v}(\vec{r}_i)) - (\vec{U}_i)]^2 \\ & + \psi_2 \int_C [\vec{v}(\vec{x}) \cdot \vec{n}(\vec{x}) - u(\vec{x})]^2 \\ & + \psi_3 \int \left(\nabla I \cdot \vec{v}(\vec{x}) + \frac{\partial I}{\partial t} \right)^2 \\ & + \psi_4 \sum_{ia} V_{ia} [\vec{v}(\vec{x}_i) - (\vec{x}_a - \vec{x}_i)]^2 \\ & + \lambda \int \sum_{m=0}^{\infty} c_m (D^m \vec{v})^2 \end{aligned} \tag{27}$$

The only difficulty would lie in assigning the relevant weighting factors ψ_1, ψ_2, ψ_3, and ψ_4 to the measurement terms. These weights should depend on the relative reliability of the information given by the different terms and might be determined experimentally.

The theory gives a natural way of combining short-range and long-range motion; psychophysical evidence for this combination exists [44].

6 The Aperture Problem and Motion Capture

There are two important problems in motion perception with which the motion coherence theory can deal; they are the topics of this section: the aperture problem and motion capture.

An example of the aperture problem consists of finding the motion of a contour; local motion detectors can only detect motion that is perpendicular to the contour. The motion coherence theory for contour motion appears in equation (10). We obtain the smoothest velocity field in the two-dimensional space, in contrast to Hildreth, who obtains the smoothest velocity field along the contour (figure 4).

For our solutions, the computed speed is smaller than the true one (see equation (7)). This is due to the term c_0, which forces the overall speed down, but is necessary to ensure that spatial interactions drop off at infinity. If $c_0 = 0$, we can prove that the correct solution may be obtained for a certain class of stimuli, and in particular, the correct answer is always obtained for translational motion (see appendix D). Similar results hold for Hildreth's method (see [45]). Psychophysical experiments, such as the rotating-ellipse illusion [11], show that humans may not estimate the correct velocity fields for other stimuli. In a number of examples, Hildreth showed that when her method gave the incorrect velocity field, it was often close to the perceived motion.

The motion coherence theory also agrees with experiments by Nakayama and Silverman [15, 16] that investigated variations of the aperture problem that are not easily explained by current theories. These experiments displayed curves translating over time (figure 5) and showed that the correct motion is not always perceived even when the curves are moving rigidly, that terminators in the curves enhance the rigid percepts, and that terminators near the curves may also lead to this enhancement. The motion coherence

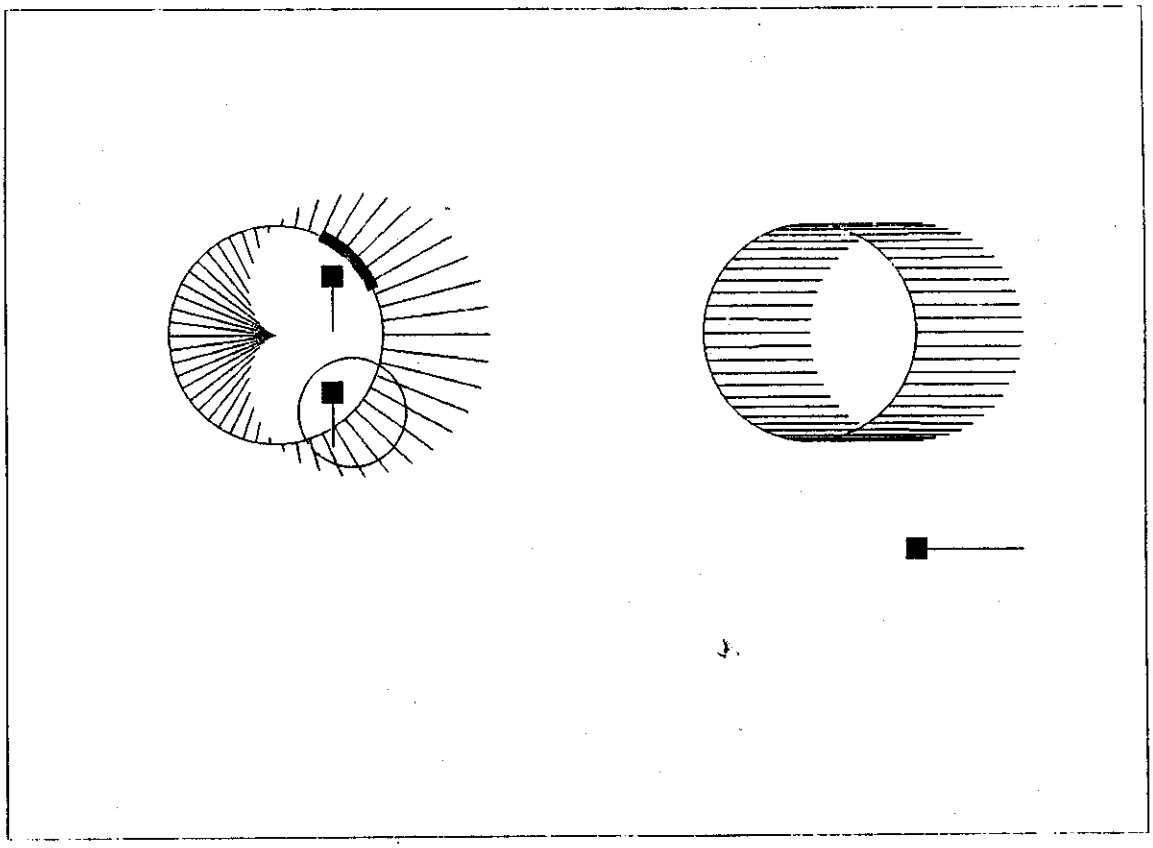

Fig. 4. The aperture problem. A circle of unit radius was assumed to be moving to the right with a velocity of 1.0. If the motion is measured with local detectors, then only the components of the motion perpendicular to the contour can be measured. This is an example of the aperture problem; and in the left figure, these perpendicular components are shown. The velocity field along the contour, as predicted by the motion coherence theory, is shown in the right figure. This field was computed by solving equation (10) with $\lambda = 0.001$ and $\sigma = 3$. The isolated dot moving at the bottom of the right figure has the correct velocity of the circle, implying that the aperture problem has been solved. The average speed of the computed velocity field along the contour was 0.996, which is incorrect by only 0.4%. The motion coherence theory solves the aperture problem in a different way than what Hildreth proposed [11]. She suggested that the visual system integrates motion information along contours, as illustrated in the left figure by the contour's highlight. In this case, the contour motion does not interact with off-the-contour information. The motion coherence theory, on the other hand, integrates motion over space (left figure's small circle). Thus, this theory predicts that the solution of the aperture problem may interact with off-the-contour motion.

theory accounts for the nonrigidity (figure 5), because if $c_0 \neq 0$ the aperture problem is not completely solved (appendix D). However, if the display includes terminators, then they provide extra information that helps to improve the solution. The theory controls the relative importance of on-the-curve and off-the-curve terminators through the ψ of equation (27), which must be found through experiments.

The theory gives a possible explanation for motion capture [5]. In this phenomenon, randomly moving dots are captured and move coherently with a superimposed grating or a surrounding contour. The motion coherence theory simultaneously solves the aperture problem of the surrounding contour and captures the internal dots [1-3]. The theory also predicts that capture may happen when only a few dots move. An example

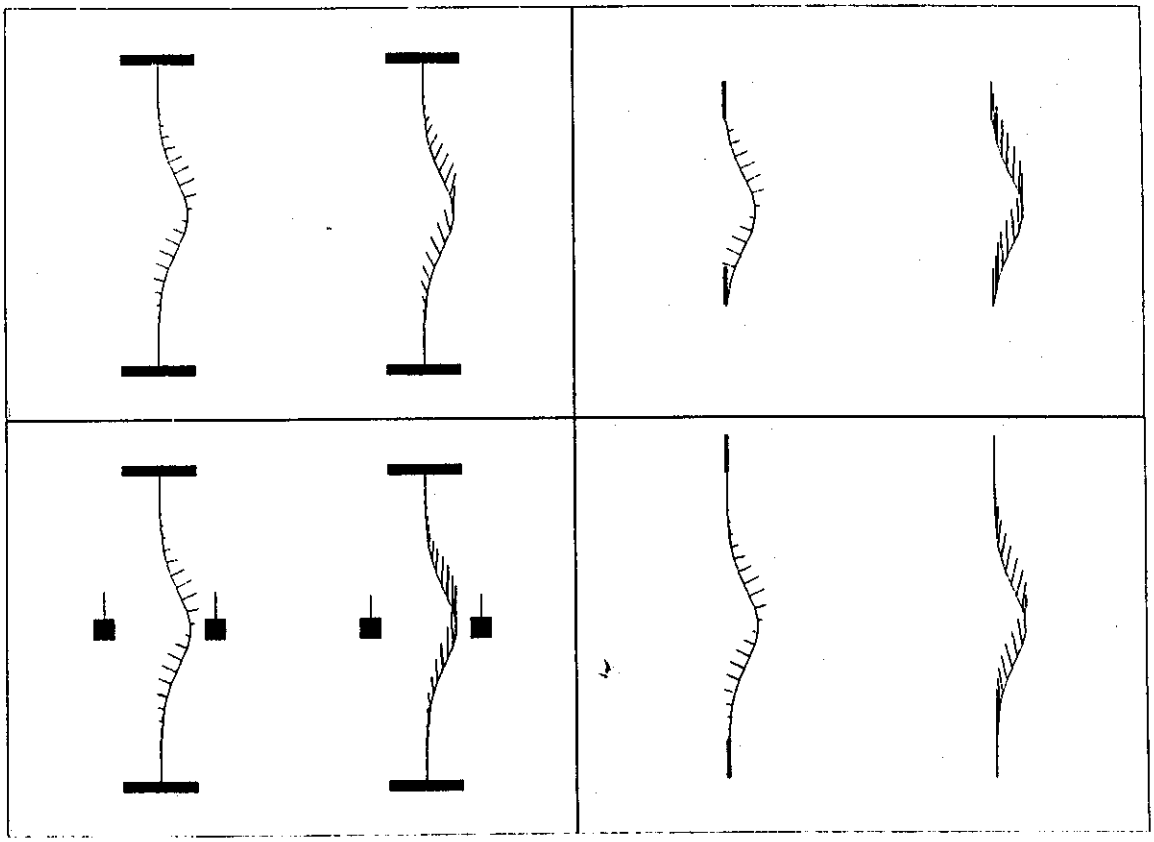

Fig. 5. Variations on the aperture problem. In this figure, Gaussians move upward with the speed highlighted in the right panels. In all panels, the left and the right displays show the motion coherence theory respectively. Nakayama and Silverman [15, 16] studied these paradigms, and their results are consistent with the theory. In the left panels, the Gaussians' extremities were occluded. In this case, the theory predicts a nonrigid percept (upper left panel). However, when features that move with the same velocity as the curve flank it, they capture it and make it look rigid (lower left panel). This is consistent with the motion coherence theory but not with Hildreth's theory [11] (see figure 4). Furthermore, if we break the curve, then the extremities' information help to solve the aperture problem (upper right panel). It is solved less well if the extremities are far from the center (lower right panel).

of this occurs when a central ambiguous motion is captured by a peripheral unambiguous motion. Psychophysical experiments with such paradigms were performed. The results were consistent with the motion coherence theory and inconsistent with the minimal mapping theory [3].

7 Extensions to Motion Discontinuity, Time Integration, and Transparency

In this section, we discuss possible extensions of the theory which deal with three problems the theory currently does not handle: discontinuities, temporal integration, and transparency.

There are several possible ways to extend this work to deal with motion discontinuities. These include the following three alternatives: First, using line processors to break the smoothing [46–48], second, detecting edges in the measuring stage's velocity field and then smoothing inside the edge region again [56] (see section 8), and third, using estimation of the velocity field whereby neighboring elements support each other if they are similar, but do not if they are dissimilar.

Psychophysical experiments have shown [24, 25, 49, 50] that for long-range motion there are inertia effects whereby the history of the motion affects the correspondence between features. It is also likely, on engineering grounds, that the velocity field is not measured instantaneously, but is estimated over time. There are several methods by which the motion coherence theory could be extended to include temporal integration and some of these are discussed elsewhere [3, 51].

The phenomena of transparency is relevant here. They occur when two or more fields of motion with very different spatial and temporal characteristics are superimposed, in which case they do not interact, but slide past each other [14]. There are at least two possible, and maybe complementary, hypotheses to explain these effects. In the first, if the motions are too dissimilar, then no interaction between them occurs (this may be the case if their measurements are performed by different spatiotemporal filters [56 and Adelson, private communication]). In the second, motion effects over time are included: If the estimate of velocity changes too quickly with time, then this may be due to integrating over different surfaces and two or more regions are created. (However, this hypothesis cannot explain all transparency phenomena, because they may occur when the moving features are short-lived [52].)

8 Discussion

In conclusion, we propose a motion coherence theory that deals simultaneously with the aperture problem and the phenomena of motion capture and motion cooperativity (see also [30]). The theory differs from some other works that deal with similar problems [9, 26–30, 53, 54]. While these works are based on "neural" considerations, the motion coherence theory is a computational theory based on an optimization process. Thus, systems based on this method can directly incorporate realistic constraints of motions in the visual world [55]. Nevertheless, in spite of the computational approach, the theory's elements, even for long-range motion [40], can be interpreted in terms of neural processes.

The exact choice of the smoothing filter is an important experimental question. We have chosen the Gaussian because of its many fine properties, but there are many alternatives. The situation is slightly analogous to action-at-a distance theories of physics, such as Newton's theory of gravity, where the force between two points falls off inversely proportional to the square of the distance. For the motion coherence theory, using the Gaussian filter, we have an interaction between features that falls off as the negative exponential of the square of the distance. This decreases rapidly with distance. Psychophysical experiments to determine what the exact fall off law is, and in particular, if it is a Gaussian, can be done.

We have emphasized the importance of requiring that the interaction falls off with distance. If this condition is not satisfied then the velocity field will become arbitrarily large the further away we are from the data. Perhaps more seriously, a feature is more likely to be captured by an object that is far away from it, rather than by an object close by. In our theory the interaction is controlled by the c_0 term in the cost function.

An alternative way to prevent the velocity field from blowing up at infinity is to impose boundary conditions on the field. However, in this case (see appendix E), the resulting velocity field will depend strongly on the choice of the boundary. In some situations, a boundary may be found naturally (for example, by image segmentation through brightness, texture, color, or motion) and natural boundary conditions on the velocity field would occur.

Thus, if the boundaries can be found reliably, then the c_0 term can be dropped. (In appendix D we show that if c_0 is dropped, then the correct velocity field would always be computed for 2-D translations.) In this case the boundary conditions would have a stronger effect on the velocity field than if $c_0 \neq 0$. It is an open experimental question to see if this occurs.

However, if the boundaries cannot be found reliably and are set arbitrarily, then if the c_0 term is dropped, the resulting velocity field will depend strongly on this arbitrary choice (unless the data is dense everywhere). *The boundary terms behave like measurement sources that influence (and sometimes capture) the motion of the internal features.*

There is a subtle difference between our interpretations of the short and the long versions of the

motion coherence theory. In the former case, the perception is of the constructed velocity field. In the latter case, what is perceived may sometimes be the correspondence. This issue matters, because in general, a given correspondence and its constructed velocity field do not coincide. If the number of moving features is small, then there is strong evidence that correspondence is established. However, if this number is large, then precise correspondence may not be computed.

An approach similar to the motion coherence theory can be applied to stereo [Yuille and Gennert, private communication]. Stereo can be considered as a "special case" of long-range motion and so most of the mathematics of the motion coherence theory can be directly applied. Yuille and Gennert describe a theory of stereo that matches image intensities and salient features.

This theory is specified at an abstract level in terms of minimizing a cost functional. We stress that the minimization of such a cost functional is inherently a parallel process and could be performed on a parallel computer or a neural network. Elsewhere, we describe a parallel network that implements our long-range theory under certain conditions [40].

Acknowledgements

Comments from Ted Adelson, Jim Clark, Dave Heeger, Ellen Hildreth, Dan Kersten, Tommy Poggio, and Jim Smith were very helpful. We would also like to acknowledge Consuelo Correa, Ivan Correa, and Michelle Borde for inspiration and encouragement. A.L.Y. was supported by United States Army Research Office grant DAAL03-86-K-0171. N.M.G. was supported by grant BNS-8809528 from the National Science Foundation, by the Sloan Foundation, and by a grant to Tomaso Poggio and Ellen Hildreth from the Office of Naval Research, Cognitive and Neural Systems Division.

Appendix A

This appendix proves two main results. First, we can choose the c_m so that the interaction is mediated by a Gaussian. Second, the interaction falls to zero at infinity if and only if the c_0 term is positive.

Suppose we have a set of measurements U_i at points $\vec{x}_i$. The Euler-Lagrange equations of equation (1) are:

$$\lambda \sum_{m=0}^{\infty} (-1)^m c_m \nabla^{2m} \vec{v}(\vec{x}) = -\sum_i \delta(\vec{x} - \vec{x}_i)[\vec{v}(\vec{x}) - \vec{U}_i] \quad \text{(A1)}$$

The solutions of these equations are of the form

$$\vec{v}(\vec{x}) = \sum_i \vec{\beta}_i f(\vec{x}) - \vec{x}_i) \quad \text{(A2)}$$

where $f(\vec{x})$ is the Green's function of the differential operator $\sum_{m=0}^{\infty}(-1)^m c_m \nabla^{2m}$.

More precisely

$$\sum_{m=0}^{\infty} (-1)^m c_m \nabla^{2m} f(\vec{x}) = \delta(\vec{x}) \quad \text{(A3)}$$

To solve this equation we apply Fourier transform theory. If the Fourier transform of $f(\vec{x})$ is given by $g(\vec{\omega})$, then we obtain

$$g(\vec{\omega}) = \frac{1}{\sum_m c_m (\vec{\omega} \cdot \vec{\omega})^m} \quad \text{(A4)}$$

To show that the Gaussian is a solution we note that its Fourier transform is

$$\frac{1}{\exp[-(\vec{\omega} \cdot \vec{\omega})\sigma^2/2]} \quad \text{(A5)}$$

We expand the denominator in powers of $\vec{\omega} \cdot \vec{\omega}$ and choose the coefficients c_m so they are the coefficients of this expansion. Thus the Gaussian can be chosen as the interaction.

Now consider integrating equation (A3) over a region bounded by a large circle centered on the origin. The integral of the right-hand side will be 1, because of properties of the delta function. The integral from the left-hand side will have a number of different contributions from all the terms in the summation. If we apply the constraint that f and all its derivatives fall to zero at infinity faster than $1/r$, where r is the circle's radius, then in the limit, as the size of the circle goes to infinity, the contributions from the terms with $n > 0$ will vanish. This is because by Stokes theorem:

$$\int_{\text{space}} \nabla^{2m} f(\vec{x}) = \int_{\text{boundary}} \vec{\nabla}[\nabla^{2m-2} f(\vec{x})] \quad \text{(A6)}$$

and the right hand side integrand goes to zero faster than $1/r$ on the boundary. So, if $c_0 = 0$ we obtain a contradiction. Thus, $c_0 > 0$ is a necessary condition for the interaction to fall off fast enough.

Conversely if $c_0 \geq 0$ then the energy function contains a term

$$c_0 \int \vec{v}(\vec{x}) \cdot \vec{v}(\vec{x}) d\vec{x} \tag{A7}$$

For this term to be finite, the velocity field, and hence $f(\vec{x})$, must fall off faster than $1/r$ as $r \to \infty$.

Appendix B

We now consider the spatial fall-off of interactions for Horn and Schunck's theory [10].

This theory proposes minimizing a cost function

$$E(\vec{v}) = \int (\vec{\nabla} I \cdot \vec{v} + I_t)^2 + \lambda \int \left(\frac{\partial \vec{v}}{\partial x} \cdot \frac{\partial \vec{v}}{\partial x} + \frac{\partial \vec{v}}{\partial y} \cdot \frac{\partial \vec{v}}{\partial y} \right) \tag{A8}$$

The Euler-Lagrange equations are

$$\lambda \nabla^2 \vec{v} = \vec{\nabla} I (\vec{v} \cdot \vec{\nabla} I + I_t) \tag{A9}$$

We can write this in coordinate notation

$$\lambda \nabla^2 v_i = \sum_j \vec{\nabla} I_i \vec{\nabla} I_j v_j + \vec{\nabla} I_i I_t \tag{A10}$$

where the indexes i,j correspond to the x,y components of the vectors.

In (A10) the term $\vec{\nabla} I_i \vec{\nabla} I_j$ effectively corresponds to a c_0 term, because it is the coefficient of the linear term in $\vec{v}$ in the Euler-Lagrange equations, and because it is nonnegative definite. If the data is dense, then this term will prevent the interactions from becoming long range. Observe, however, that the smaller $|\vec{\nabla} I_i|$ is, the longer the range of the interactions. Thus, if there are regions in the image where $|\vec{\nabla} I_i|$ is very small, which is often the case, the interactions become very long range.

Appendix C

We discuss a simple version of the motion cooperativity experiments. It consists of a set of dots randomly and homogeneously distributed all over the space. They are also assigned velocities with a probability distribution $P(\vec{U})$. We want to calculate the mean and the standard deviation of the velocity field predicted by the motion coherence theory. To perform this calculation, we assume that the distribution of the dots' positions and velocities in small regions of the image is representative of the overall probability distributions. By small we mean on the spatial scale, σ, of the interaction. This assumption will only hold if the number of dots is large on this scale. More precisely, if $\rho\pi\sigma^2 \gg 1$, where p is the density of the dots.

The Euler-Lagrange equation is

$$\lambda \sum_{m=0}^{\infty} (-1)^m \frac{\sigma^{2m}}{(m!2^m)} \nabla^{2m} \vec{v}(\vec{x}) = -\sum_i \delta(\vec{x} - \vec{x}_i)[v(\vec{x}) - U_i] \tag{A11}$$

Then, the velocity field (again we can consider each component separately) is given by

$$v(\vec{x}) = \sum_i \beta_i G(\vec{x} - \vec{x}_i) \tag{A12}$$

where the β_i obey (obtained by substituting equation (A12) into (A11))

$$(\lambda \delta_{ij} + G_{ij}) \beta_j = U_i \tag{A13}$$

where G_{ij} was defined in equation (6), the summation convention is used (see its definition before Equation (22)), and the U_i are the components of the velocity data. This can be written as

$$\beta_i = A_{ij} U_j \tag{A14}$$

where

$$A_{ij} = (\lambda \delta_{ij} + G_{ij})^{-1} \tag{A15}$$

Note that A_{ij} depends only on the dots' spatial distribution and U_i depends only on $P(\vec{U})$. We use the notation $(U_i)_v$ to denote the expectation U_i with respect to the velocity distribution P_v and define $H_i = G(\vec{x} - \vec{x}_i)$. Then, we obtain

$$v(\vec{x}) = H_i A_{ij} U_j \tag{A16}$$

Thus, the mean $\langle v(\vec{x}) \rangle_{S,v}$ is given by

$$\langle v(\vec{x}) \rangle_{S,v} = \langle H_i A_{ij} U_j \rangle_{S,v} \tag{A17}$$

To find A_{ij} we expand it in an infinite Taylor series. Setting $g = 1/(2\pi\sigma^2)$ we find

$$A_{ij} = [(\lambda + g)\delta_{ij} + G_{ij} - g\delta_{ij})]^{-1} \quad \text{(A18)}$$

$$A_{ij} = \frac{1}{\lambda + g}\sum_{n=0}^{n=\infty}\frac{(-1)^n}{(\lambda+g)^n}(G_{ij} - g\delta_{ij})^n \quad \text{(A19)}$$

$$A_{ij} = \frac{1}{\lambda + g}\sum_{n=0}^{n=\infty}\frac{(-1)^n}{(\lambda+g)^n} \sum_{m=0}^{n}\frac{n!}{m!(n-m)!}(-1)^{(n-m)}g^{n-m}G_{ij}^m \quad \text{(A20)}$$

So we must evaluate the expectations of terms like $G_{ij}^m G(\vec{x} - \vec{x}_i)U_j$. We assume that exactly N dots are distributed uniformly in a square box of side L, which has an area similar or larger than $\pi\sigma^2$. The density ρ is therefore given by $\rho = N/L^2$. The fixed-N assumption is valid, because the number of dots inside the box is large ($> \rho\pi\sigma^2$) and follows a Poisson distribution, which in this case, has a small standard deviation to mean ratio. After doing these expectation calculations, we will take the limit as $N \to \infty$ and $L \to \infty$. The expectation of $G_{ij}^m G(\vec{x} - \vec{x}_i)U_j$ is given by

$$\sum_{\substack{i,i_1,\ldots,i_{m-1}\\ j=1}}^{N}\int G(\vec{x}_i - \vec{x}_{i_1})G(\vec{x}_{i_1} - \vec{x}_{i_2})\times \ldots \times G(\vec{x}_{i_{m-1}} - \vec{x}_j)G(\vec{x} - \vec{x}_i)U_j \frac{d\vec{x}_i}{L^2}\frac{d\vec{x}_{i_1}}{L^2}\ldots \frac{d\vec{x}_{i_{m-1}}}{L^2}\frac{d\vec{x}_i}{L^2}P(U_j)dU_j \quad \text{(A21)}$$

We evaluate the integral (the Gaussians integrate out to 1) and perform the summations. Using $\rho = N/L^2$ this yields

$$\rho^{m+1}\langle U\rangle \quad \text{(A22)}$$

Thus, whenever we obtain a term of the form $G_{ij}^m G(\vec{x} - \vec{x}_i)U_j$ in (A17) we replace it by (A22). We find

$$\langle v\rangle = \frac{\langle U\rangle\rho}{\lambda + g}\sum_{n=0}^{n=\infty}\frac{(-1)^n}{(\lambda+g)^n} \times \sum_{m=0}^{n}\frac{n!}{m!(n-m)!}(-1)^m g^{n-m}\rho^m \quad \text{(A23)}$$

This can be summed to give

$$\langle v\rangle = \frac{\langle U\rangle\rho}{\lambda + g}\sum_{n=0}^{n=\infty}\frac{(-1)^n}{(\lambda+g)^n}(\rho - g)^n \quad \text{(A24)}$$

Thus, we obtain

$$\langle v\rangle = \frac{\rho}{\lambda + \rho}\langle U\rangle \quad \text{(A25)}$$

To compute the standard deviation we must evaluate terms of the form

$$\langle v^2\rangle = \langle (H_i A_{ij} U_j)^2\rangle \quad \text{(A26)}$$

$$v^2(\vec{x}) = \Big\langle \sum_{i,j=1}^{N}\sum_{p,q=1}^{N} A_{ij}G(\vec{x} - \vec{x}_i) \times U_j A_{pq} G(\vec{x} - \vec{x}_p)U_q \Big\rangle \quad \text{(A27)}$$

We must now evaluate the expectations of terms like $G_{ij}^m G(\vec{x} - \vec{x}_i)U_j G_{pq}^b G(\vec{x} - \vec{x}_p)U_q$. This is

$$\sum_{\substack{i,i_1,\ldots,i_{m-1}\\ p,p_1,\ldots,q=1}}^{N}\int G(\vec{x} - \vec{x}_i)G(\vec{x}_{i_1}) \times G(\vec{x}_{i_1} - \vec{x}_{i_2}) \times \ldots G(\vec{x}_{i_{m-1}} - \vec{x}_j)U_j \times G(\vec{x} - \vec{x}_p)G(\vec{x}_p - \vec{x}_{p_1})G(\vec{x}_{p_1} - \vec{x}_{p_2}) \times \ldots \times G(\vec{x}_{p_{b-1}} - \vec{x}_q)U_q \frac{d\vec{x}_j}{L^2}\frac{d\vec{x}_{i_1}}{L^2} \times \ldots \times \frac{d\vec{x}_{i_{m-1}}}{L^2}\frac{d\vec{x}_j}{L^2}\frac{d\vec{x}_p}{L^2}\frac{d\vec{x}_{p_1}}{L^2} \times \frac{d\vec{x}_{p_{b-1}}}{L^2}\frac{d\vec{x}_q}{L^2}P(U_j)\,dU_j)P(U_q)dU_q \quad \text{(A28)}$$

We evaluate the integral, except for the j, q terms (the Gaussians integrate out to 1), and perform the summations. Using $\rho = N/L^2$ this yields

$$\rho^{m+b}\sum_{j,q=1}^{N}\int G_{t_{1=(m+1)}}(\vec{x} - \vec{x}_j)U_j G_{t_{2=(b+1)}} (\vec{x} - \vec{x}_q)\frac{d\vec{x}_j}{L^2}\frac{d\vec{x}_q}{L^2}U_q P(U_j)\,dU_j P(U_q)dU_q \quad \text{(A29)}$$

We first consider the diagonal terms, where $j = q$. Integrating the Gaussians gives,

$$\rho^{m+b+1}\langle U^2\rangle\frac{1}{2\pi(m + b + 2)\sigma^2} \quad \text{(A30)}$$

The remaining, off-diagonal, terms gives

$$\rho^{m+b}\frac{N(N-1)}{L^4}\langle U\rangle^2 \quad \text{(A31)}$$

which becomes

$$\rho^{m+b+2}\langle U\rangle^2 \tag{A32}$$

in the limit of large N.

Thus, the total contribution of the diagonal terms is

$$A = \langle U^2\rangle \sum_{n,m}\sum_{p,q}\frac{(-1)^n}{(\lambda+g)^{n+1}}(-g)^{n-p} \times \frac{n!}{p!\,(n-p)!}\frac{(-1)^m}{(\lambda+g)^{m+1}}(-g)^{m-q} \times \frac{m!}{q!(m-q)!}\frac{\rho^{p+q+1}}{2\pi(p+q+2)\sigma^2} \tag{A33}$$

To evaluate this formula, we use the identity

$$\frac{\rho^{p+q+1}}{2\pi(p+q+2)\sigma^2} = \frac{1}{2\pi\sigma^2}\frac{1}{\rho}\int_0^\rho ds s^{p+q+1} \tag{A34}$$

We find by substituting (A34) into (A33) and summing over p and q:

$$A = \frac{\langle U^2\rangle}{2\pi\sigma^2}\frac{1}{\rho}\int_0^\rho ds s \sum_{n,m}\frac{(-1)^n}{(\lambda+g)^{n+1}}(s-g)^n \frac{(-1)^m}{(\lambda+g)^{m+1}}(s-g)^m \tag{A35}$$

Therefore

$$A = \frac{\langle U^2\rangle}{2\pi\sigma^2}\frac{1}{\rho}\int_0^\rho ds\,\frac{s}{(\lambda+s)^2} \tag{A36}$$

$$A = \frac{\langle U^2\rangle}{2\pi\sigma^2}\frac{1}{\rho}\left[\log\left(\frac{\rho+\lambda}{\lambda}\right) - \frac{\rho}{\rho+\lambda}\right] \tag{A37}$$

The total contribution of the off-diagonal terms is

$$B = \sum_{n,m}\sum_{p,q}\frac{(-1)^n}{(\lambda+g)^{n+1}}(-g)^{n-p}\frac{n!}{p!(n-p)!} \times \frac{(-1)^m}{(\lambda+g)^{m+1}}(-g)^{m-q}\frac{m!}{q!(m-q)!}\rho^{p+q+2}\langle U\rangle^2 \tag{A38}$$

By comparing (A38) to formula (A23) for the mean we see that

$$B = \frac{\rho^2}{(\rho+\lambda)^2}\langle U\rangle^2 \tag{A39}$$

The variance of $v(\vec{x})$ is found by subtracting the square of the mean from the sum of A and B. This yields

$$\mathrm{var}(v) = \frac{1}{2\pi\sigma^2}\frac{1}{\rho}\left[\log\left(\frac{\rho+\lambda}{\lambda}\right) - \frac{\rho}{\rho+\lambda}\right]\langle U^2\rangle \tag{A40}$$

Appendix D

In this section, we consider the aperture problem. We consider an object translating rigidly in the image plane and prove that if $c_0 = 0$, then the true velocity will always be calculated. If $c_0 \neq 0$, as in the motion coherence theory, the calculated velocity will only approximate the true velocity, although simulations show that the approximation is usually good.

Psychophysical experiments [15, 16] show that the true velocity field is not always perceived for rigidly moving objects. This gives an additional motivation for setting $c_0 \neq 0$.

Suppose we have a contour C, $\vec{r} = \vec{r}(s)$, with normal vector $\vec{n}(s)$, where s is the arc length. The normal components $u(s)$ of the velocity field are measured and the theory predicts the velocity field that minimizes

$$E(\vec{v}) = \int_C [\vec{v}\cdot\vec{n}(s) - u(s)]^2 + \lambda\int\sum_{m=0}^{\infty} c_m(D^m\vec{v})^2 \tag{A41}$$

which was given in equation (A32). The minimum of this energy function satisfies the Euler-Lagrange equation

$$\int_C \delta[\vec{r} - \vec{r}(s)]\vec{n}(s)[\vec{v}(\vec{r})\cdot\vec{n}(s) - u(s)] = \sum_{m=0}^{\infty}(-1)^{m+1}c_m D^{2m}\vec{v}(\vec{r}) \tag{A42}$$

If the object translates rigidly, then the true velocity field is constant, $\vec{a}$. The data $u(s)$ are given by $u(s) = \vec{a}\cdot\vec{n}(s)$. Now substitute $\vec{v}(\vec{r}) = \vec{a}$ into (A42). The left-hand side of the equation will be automatically zero. The right-hand side equals $-c_0\vec{a}$. Thus the true velocity field is a solution if and only if $c_0 = 0$.

Appendix E

We now show that the velocity field will depend strongly on the boundary conditions if $c_0 = 0$. For simplicity we consider the analogous problem of interpolating a function $f(\vec{x})$ from data d_i given at points $\vec{x}_i$, for $i = 1, \ldots, N$. (For example, $f(x)$ could be one of the components of the velocity field.) Suppose we have a cost function

$$E(f) = \int (Lf)^2 d\vec{x} + \sum_{i=1}^{i=N} [f(\vec{x}_i) - d_i]^2 \qquad \text{(A43)}$$

We now choose a boundary $\vec{x} = \vec{x}(s)$, where s parameterizes the curve $0 \leq s \leq S_{max}$, and impose the boundary conditions that $f(\vec{x}(s)) = 0$ (the same analysis can be applied if we choose $f(\vec{x}(s)) = \psi(s)$, where $\psi(s)$ is an arbitrary function, perhaps found by some other process).

We impose these boundary conditions by using Lagrange multipliers. This gives a modified cost function

$$E_M(f) = \int (Lf)^2 d\vec{x} + \sum_{i=1}^{i=N} [f(\vec{x}_i) - d_i]^2 + \int f(\vec{x}(s))\lambda(s)\, ds \qquad \text{(A44)}$$

which is minimized with respect to the field $f(\vec{x})$ and the Lagrange multipliers $\lambda(s)$.

Minimizing with respect to the Lagrange multipliers and $f(\vec{x})$ gives

$$f(\vec{x}(s)) = 0 \qquad \text{(A45a)}$$

$$L^2 f(\vec{x}) = \sum_i \delta(\vec{x} - \vec{x}_i)[f(\vec{x}_i) - d_i] + \int \delta[\vec{x} - \vec{x}(s)]\lambda(s)\, ds \qquad \text{(A45b)}$$

The solution is therefore of form

$$f(\vec{x}) = \sum_{i=1}^{N} \alpha_i Gr(\vec{x} - \vec{x}_i) + \int \mu(s) Gr[\vec{x} - \vec{x}(s)]\, ds \qquad \text{(A46)}$$

where $Gr(\vec{x})$ is the Green's function of the operator L^2, and the α_i and $\mu(s)$ are chosen to satisfy (A45a) and (A45b). This gives

$$\sum_{i=1}^{N} \alpha_i Gr[\vec{x}(s) - \vec{x}_i] + \int \mu(s) Gr[\vec{x}(s) - \vec{x}(t)]\, dt = 0 \qquad \text{(A47a)}$$

$$\alpha_k = \sum_j \left\{ \sum_{i=1}^{N} \alpha_i Gr(\vec{x}_i - \vec{x}_j) + \int \mu(s) Gr[\vec{x}_i - \vec{x}(t)]\, dt \right\} + d_i \qquad \text{(A47b)}$$

These equations are similar to those obtained without boundary conditions. *The boundary terms behave like measurement sources that influence (and sometimes capture) the motion of the internal features.* However, there is one important difference: At the boundary, the values of $f(\vec{x})$ are required to be *exactly* 0 (or whatever value $\psi(s)$ has been specified), while at the measurement points, $\vec{x}_i$, the function $f(\vec{x})$ is only required to be near the data d_i. Thus the "data" at the boundary will be imposed more strongly than the data at the measurement points.

From (A47a) and (A47b) we see that the range of influence of boundary points will depend on the spatial behavior of the Green's function $Gr(\vec{x})$. If the Green's function increases with distance (for example, if $L = \vec{\nabla}$ and $Gr(\vec{x}) = \log|\vec{x}|$), then the choice of boundary conditions can have a large effect on the motion. If the Green's function falls off with distance (for example, if $c_0 \neq 0$), then the boundary conditions will usually have less effect, and will give a negligible contribution as the boundary is sent to infinity.

Similar calculations can be done if the measurements are on contours or on dense data. The conclusion is the same: The boundary terms will influence the motion, and the range of this influence is determined by the Green's function of L^2.

References

1. A.L. Yuille and N.M. Grzywacz, "A computational theory for the perception of coherent visual motion," *Nature* 333: 71–74, 1988.
2. A.L. Yuille and N.M. Grzywacz, "The motion coherence theory," *Proc. 2nd Intern. Conf. Computer Vision*, Tampa, FL, pp. 344–353, 1988.
3. N.M. Grzywacz, J.A. Smith, and A.L. Yuille, *Proc. IEEE Workshop on Visual Motion.*, Irvine, CA, pp. 148–155, 1989.
4. K. Koffka, *Principles of Gestalt Psychology*. Harcourt, Brace and Wood: New York, 1935.
5. V.S. Ramachandran and S.M. Anstis, "Displacement thresholds for coherent apparent motion random dot-patterns," *Vision Research* 24:1719–1724, 1983.
6. D.W. Williams and R. Sekuler, "Coherent global motion percepts from stochastic local motions," *Vision Research* 24:55–62, 1984.

7. D.W. Williams, G. Philips, and R. Sekuler, "Hysteresis in the perception of motion direction: evidence for neural cooperativity," *Nature* 324:253-255, 1986.
8. A.L. Yuille and S. Ullman, "Rigidity and smoothness." A.I. Memo 989, MIT Artif. Intell. Lab., Cambridge, MA, 1987.
9. E.H. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am.* A 2:284-299, 1985.
10. B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence* 17:185-203, 1981.
11. E. Hildreth, *The Measurement of Visual Motion.* MIT Press: Cambridge, MA, 1984.
12. S. Ulman, *The Interpretation of Visual Motion.* MIT Press, Cambridge, MA, 1979.
13. D. Marr, and S. Ullman, "Directional selectivity and its use in early visual processing," *Proc. Roy. Soc. London* B 211:151-180, 1981.
14. E.H. Adelson and J.A. Movshon, "Phenomenal coherence of moving visual patterns," *Nature* 300:523-525, 1982.
15. K. Nakayama and G.H. Silverman, "The aperture problem—I. Perception of nonrigidity and motion direction in translating sinusoidal lines," *Vision Research* 28:739-746, 1988.
16. K. Nakayama and G.H. Silverman, "The aperture problem—II. Spatial integration of velocity information along contours," *Vision Research* 28: 747-753, 1988.
17. M. Biederman-Thorson, J. Thorson, and G.D. Lange, "Apparent movement due to closely spaced sequentially flashed dots in the human peripheral field of vision," *Vision Research* 11:889-903, 1971.
18. O.J. Braddick, "A short range process in apparent motion," *Vision Research* 14:519-527, 1974.
19. O.J. Braddick, "Low-level process in apparent motion," *Phil. Trans. Roy. Soc. London* B. 290:137-151, 1980.
20. C.L. Baker and O.J. Braddick, "Does segregation of differently moving areas depend on relative or absolute displacement?" *Vision Research* 22:851-856, 1982.
21. J.J. Chang and B. Julesz, "Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion," *Vision Research* 23:1379-1385, 1983.
22. J.T. Petersik, R. Pufahl, and E. Krasnoff, "Failure to find an absolute retinal limit of a putative short-range process in apparent motion," *Vision Research* 23:1663-1670, 1983.
23. K. Nakayama and G.H. Silverman, "Temporal and spatial characteristics of the upper displacement limit for motion in random dots," *Vision Research* 24:293-299, 1984.
24. N.M. Grzywacz, "Interactions between minimal mapping and inertia in long-range apparent motion," *Invest. Ophthalmol. Vision* 28:300, 1987.
25. N.M. Grzywacz, "The effects of past motion information on long-range apparent motion." In preparation.
26. B. Hassenstein and W.E. Reichardt, "Systemtheoretische analyse de zeit, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des russelkafers *chlorophanus*," *Z. Natureforsch.* 11b:513-524,1956.
27. J.P.H. van Santen and G. Sperling, "A temporal covariance model of motion perception," *J. Opt. Soc. Am.* A 1:451-473, 1985.
28. A.B. Watson and A.J. Ahumada, "Model of human visual-motion sensing," *J. Opt. Soc. Am.* A 2: 322-341, 1985.
29. D. Heeger, "A model for the extraction of image flow," *J. Opt. Soc. Am.* A. 4:1455-1471, 1987.
30. H. Bulthoff, J. Little, and T. Poggio, "A parallel algorithm for real-time computation of optical flow," Nature 337: 549-553, 1989.
31. K. Wohn and A. Waxman, "Contour evolution, neighborhood deformation and local image flow: Curved surfaces in motion," Tech. Rept. 134, University of Maryland's Center for Automation Research, College Park, MD, 1985.
32. A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems.* Winston: Washington, DC, 1977.
33. M. Bertero, T. Poggio, and V. Torre, "Regularization of ill-posed problems," A.I. Memo. 924, MIT Artif. Intell. Lab., Cambridge, MA, 1987.
34. D. Shulman and J. Aloimonos, "Boundary preserving regularization: theory part I," Tech. Rept. 356, University of Maryland's Center for Automation Research, College Park, MD, 1988.
35. J. Duchon, *Lecture Notes in Mathematics,* vol. 571. W. Schempp, and K. Zeller (eds.), Springer-Verlag: Berlin, West Germany, 1979.
36. D. Marr and E.C. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London* B 207:187-217, 1980.
37. A.L. Yuille and T. Poggio, "Scaling theorems for zero-crossings," A.I. Memo. 722, MIT Artif. Intell. Lab., Cambridge, MA, 1983.
38. A. Witkin, "Scale-space filtering," *Proc. 8th Intern. Joint Conf. Artif. Intell.,* Karlsruhe, West Germany, pp. 1019-1021, 1983.
39. T. Poggio, H. Vorhees, and A.L. Yuille, "A regularized approach to edge detection," A.I. Memo. 833, MIT Artif. Intell. Lab., Cambridge, MA, 1985.
40. N.M. Grzywacz and A.L. Yuille, "Massively parallel implementations of theories for apparent motion," *Spatial Vision.* 3:15-44, 1988.
41. P.A. Kolers, *Aspects of Motion Perception.* Pergamon Press: Oxford, England, 1972.
42. J.A. Wheeler and R.P. Feynman, *Rev. Mod. Phys.* 17:159. 1945.
43. J.A. Wheeler and R.P. Feynman, *Rev. Mod. Phys.* 21:425. 1949.
44. M. Green, "Inhibition and facilitation of apparent motion by real motion," *Vision Research* 23: 861-865, 1983.
45. A.L. Yuille, "The smoothest velocity field and token matching schemes," A.I. Memo. 724, MIT Artif. Intell. Lab., Cambridge, MA, 1984.
46. A.L. Yuille, "Energy functions and analog networks," A.I. Memo. 987, MIT Artif. Intell. Lab., Cambridge, MA, 1987.
47. E.Gamble and T. Poggio, "Visual integration and detection of discontinuities: The key role of intensity edges,"

A.I. Memo. 970, MIT Artif. Intell. Lab., Cambridge, MA, 1987.

48. J. Hutchinson et al., "Computing motion using analog and binary resistive networks," *Computer* 21:52–63, 1988.
49. V.S. Ramachandran and S.M. Anstis, "Extrapolation of motion path in human visual perception," *Vision Research* 23:83–85, 1983.
50. S.M. Anstis and V.S. Ramachandran, "Visual inertia in apparent motion," *Vision Research* 27: 755–764, 1987.
51. N.M. Grzywacz, "Properties of generalized minimal mapping theories for apparent motion," *J. Opt. Soc. Am.* A 3:72, 1986.
52. R.M. Siegel and R.A. Andersen, "Perception of three-dimensional structure from motion in monkey and man," *Nature* 331:259–261, 1988.
53. M. Fahle and T. Poggio, "Visual hyperacuity: Spatiotemporal interpolation in human vision," *Proc. Roy. Soc. London* B. 213:451–477, 1981.
54. W. Reichardt, R.W. Schlogl, and M. Egelhaaf, *Die Naturwissenschaften.* In press.
55. D. Marr, *Vission.* Freeman: San Francisco, CA, 1982.
56. A.L. Yuille and N.M. Grzywacz, "A model for the estimate of local image velocity by cells in the visual cortex." In preparation.