Adventures with gelatinous ellipses—constraints on models of human motion analysis

Yair Weiss, Edward H Adelson
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; e-mail: yweiss@cs.berkeley.edu

Abstract. An ellipse rotating rigidly about its center may appear to rotate rigidly or to deform nonrigidly so that it appears gelatinous. We use this ambiguous stimulus to study how motion information is propagated across space. We find that features that are quite far from the contour of the ellipse may have a strong influence on the percept of the ellipse, provided they move in a way consistent with the motion of the ellipse. We show that the percept cannot be accounted for by computational models that pool constraints over a local area only, or by models that propagate information along contours, or by models that indiscriminately propagate information across space. However, the percept can be accounted for by a class of models that assume smoothness in a layered representation.

1 Introduction
An ellipse rotating rigidly in the image plane may be perceived in at least three different ways (Musatti 1924; Wallach et al 1956; Vallortigara et al 1988). Sometimes, the veridical rigid rotation is perceived but two other 'illusory' motions can be seen as well. The ellipse may be perceived as deforming nonrigidly in the image plane, as if it were made of rubber. Alternatively, the ellipse may be perceived as executing a rigid motion in depth, as if it were a coin twisting in 3-D.

Wallach et al (1956) and Hildreth (1983) have pointed out that these additional percepts should not really be called 'illusions'—they are interpretations that are fully consistent with the retinal input. Since the ellipse is bounded by a smooth contour, there exists an infinite number of velocity fields that would generate the same retinal sequence. In accordance with the well-known aperture problem (Wallach 1935; Marr and Ullman 1981; Hildreth 1983), at any point along the contour the local motion information is consistent with an infinite number of possible motions (figure 1a). Graphically, this ambiguity corresponds to a 'constraint line' in velocity space (Adelson and Movshon 1982; Nakayama and Silverman 1988a) (see figure 1b). To illustrate this ambiguity, figure 2 shows velocity fields that have been computed analytically such that they are consistent with the local constraint lines of a rotating ellipse. Figures 2a and 2c show the rotational velocity fields for ellipses of two aspect ratio, and figures 2b and 2d show the deforming normal flow; at every location the flow is perpendicular to the contour of the ellipse. Each field is fully consistent with the local constraint lines.

Despite the infinite number of possible interpretations, observers typically report seeing one dominant interpretation at a time. For example, as Wallach et al first reported, a 'narrow' ellipse tends to be perceived as rigidly rotating, while a 'fat' ellipse tends to be perceived as nonrigidly deforming (at short presentation times) or as a coin rotating in depth (after extended viewing) (Vallortigara et al 1988). As in other examples of ambiguous stimuli, the consistent choice of one possible interpretation over others can provide insight into the computation done during perception. In this paper, we focus on the choice between the two 2-D interpretations. We report some new variants of rotating-ellipse stimuli and discuss the constraints these phenomena place on models of human motion analysis.
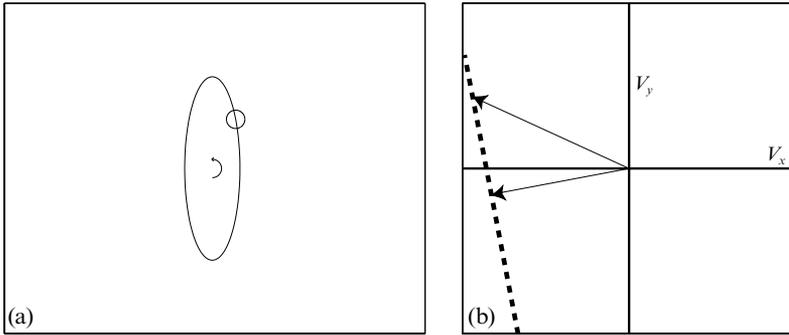
Figure 1. The aperture problem. The motion of a smooth curve cannot be resolved locally. Motion along the curve's tangent cannot be detected. (a) A rotating ellipse. (b) A velocity-space plot of the motions consistent with the information in the window in (a). All velocities lying on the dashed line are consistent. The arrow orthogonal to the constraint line denotes the normal velocity. The other arrow denotes the rotational velocity.
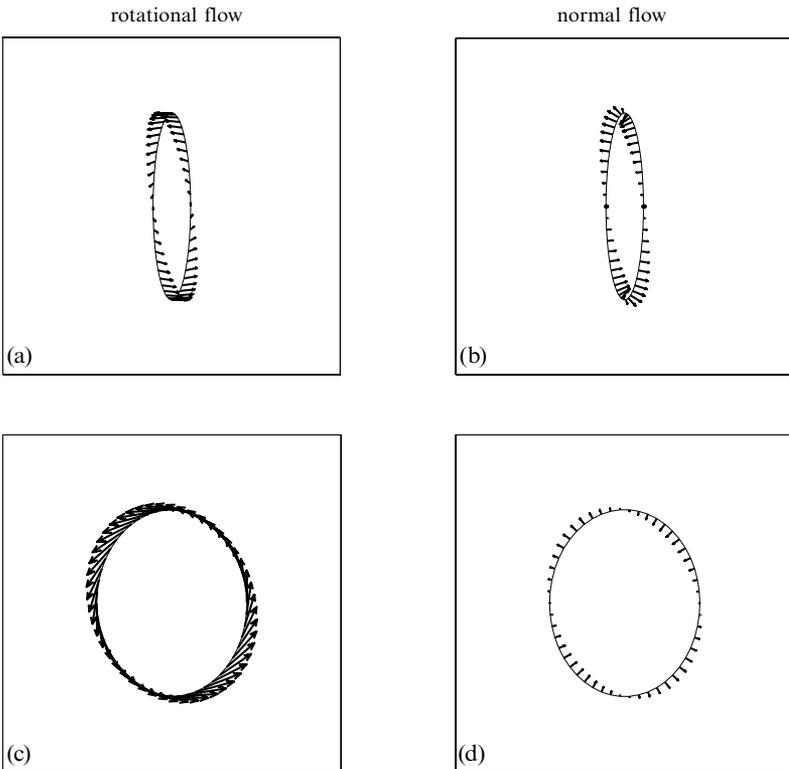

Figure 2. Velocity fields calculated analytically to be consistent with the local motion information of a rotating ellipse. (a) and (b) Rotational and normal flow for a rotating ellipse of aspect ratio 0.2. Owing to the aperture problem, the local data are equally consistent with both velocity fields. (c) and (d) Rotational and normal flow for a rotating ellipse of aspect ratio 0.8.

2 Phenomena

The phenomena described here are sufficiently robust that they work for naive and experienced observers under different viewing conditions. Results reported in this section have been obtained under the following conditions. Image sequences were generated on a Silicon-Graphics O2 computer and were recorded on an sVHS video cassette via the SGI O2 s-video

output board. The rotational motion of the ellipse was sampled at a rate of $1°$ per frame and all contours were anti-aliased with the use of the SGI hardware to create the percept of a smooth motion. Using this method we generated eight short movie clips (referred to below as clips 1 through 8). Five subjects (including one of the authors) free-viewed these clips and were instructed to verbally report the perceived motion of the ellipse including "what material the ellipse appears to be made of". The percepts reported by the observers were in agreement for all of the clips except clip 8 (discussed below). None of the observers reported multiple interpretations. We have also shown these displays to hundreds of observers at scientific meetings (Weiss and Adelson 1995a, 1995b, 1996a) who reported similar percepts. We urge the reader to view the clips (available online at www.perceptionweb.com/perc0300/weiss.html) to gain a firsthand experience of the percept. Details of the parameters used in the experiments are given in the appendix.

2.1 *Narrow versus fat*
Figure 3 and movie clip 1 illustrate the effect of ellipse aspect ratio on the perceived rigidity, first reported by Wallach et al (1956). Observers tend to perceive narrow ellipses, in which the aspect ratio is far from unity, as rigidly rotating in the image plane. As the aspect ratio becomes closer to unity, the percept becomes less and less rigid and subjects describe the ellipse as gelatinous—it appears to deform nonrigidly with motion, qualitatively similar to the normal flow in figure 2d.
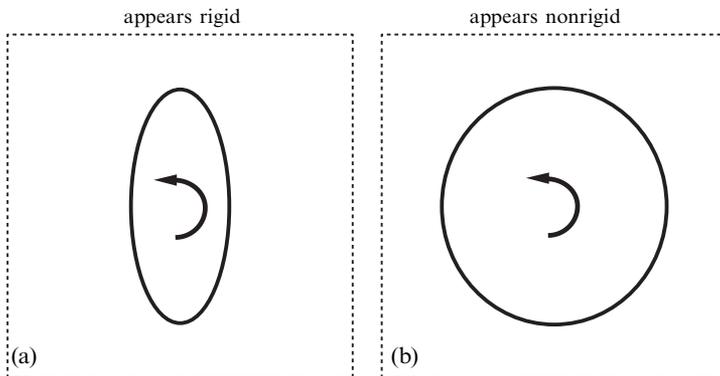

Figure 3. Narrow versus fat. A narrow rotating ellipse (a) appears rigid while a fat rotating one (b) appears nonrigid.

2.2 *The influence of satellites*
Braunstein and Andersen (1984) showed that placing a small number of dots on a rotating smooth contour can dramatically change the perceived motion. Nakayama and Silverman (1988a, 1988b) showed a strong influence occurs even if the points are off the contour (see figure 4). They showed that subjects 'misperceived' the motion of an ogive translating horizontally in the image plane—it appeared to deform nonrigidly. When two horizontally translating dots were added to the display, however, the curve appeared rigid. We wanted to see whether the rotating ellipse is subject to a similar influence.

We added four dots, or 'satellites' to the display, and the motion of these satellites could take on one of two possible motions. In the first case, which we call 'rotating sats', the four dots took part in the same rotational motion as the ellipse. In the second case, which we call 'nonrigid sats', each dot executed normal motion: the motion of each dot was calculated to lie in the direction perpendicular to the contour of an ellipse with the same shape as the displayed ellipse, but with a larger radius. The speed was calculated so that the normal component of the velocity was equal to the normal component of rotation of the ellipse. Thus in both cases the motion of the dots was consistent with the constraint lines generated locally by a rotating ellipse.
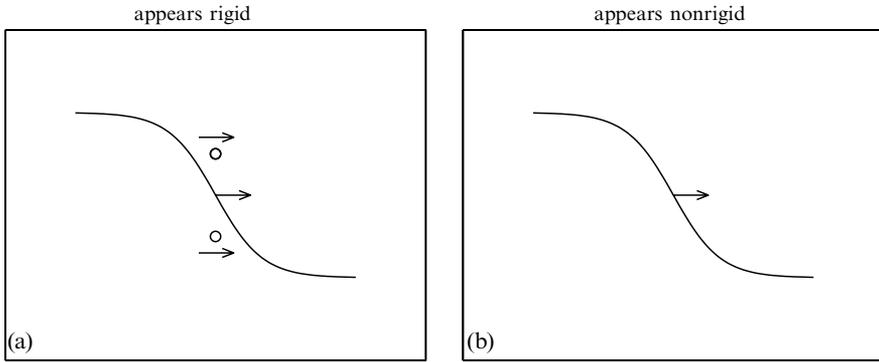
Figure 4. The influence of satellites on the percept of a translating ogive (Nakayama and Silverman 1988a). The rigidly translating ogive appears to be nonrigid (b) but, when a pair of translating dots are added to the display, the curve appears rigid (a).
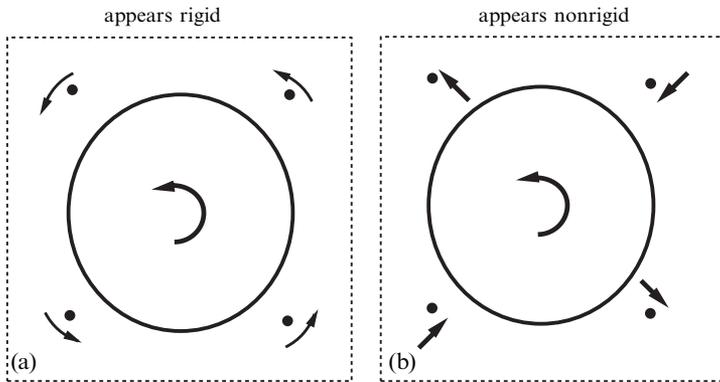

Figure 5. The influence of satellites. A small number of dots profoundly influence the perceived motion of an ellipse with an intermediate aspect ratio. (a) When the dots execute a rigid rotational motion, the ellipse appears rigid. (b) When they execute a nonrigid deformation, the ellipse appears gelatinous—the perceived motion is a nonrigid deformation similar to the normal flow.

Figure 5 and movie clip 2 summarize the percepts. When the satellites move in the direction of rotation, the ellipse appears to rotate rigidly. When they move in the normal direction, it appears to deform nonrigidly. Subjects often refuse to believe that the ellipse is doing the same thing.

2.3 Effect of satellites persists over static background

The strong influence of the satellites on the percept of the ellipse suggests that motion information from the satellites is propagated across space thereby influencing the contour. But does the visual system propagate all motion constraints in the image indiscriminately? Many authors (eg Terzopoulos 1986; Hutchinson et al 1988) have pointed out that such global propagation would lead to estimates that are quite wrong. They have advocated an alternative approach whereby motion boundaries stop the propagation.

In order to create a motion boundary between the ellipse and the dots, we placed the ellipse over a textured static background. This has the effect of surrounding the satellite with unambiguous motion signals corresponding to zero velocity. Since the satellites are unambiguously moving, most motion boundary algorithms would find a boundary surrounding each satellite. Thus a model that stops propagation at motion boundaries would predict that there should be no influence of the satellites in this display. Figure 6 and clip 3 summarize the percept—the effect of the satellites persists. Observers continue to report that the perceived motion of the ellipse changes dramatically as the motion of the satellites is varied.
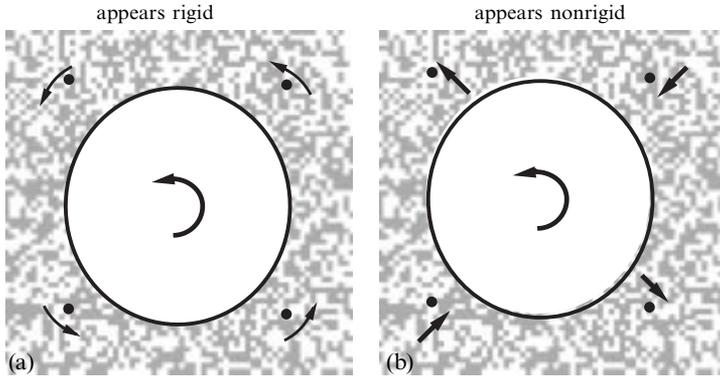
appears rigid   appears nonrigid

Figure 6. The effect of satellites persists over a static background. Even though a motion discontinuity is now formed between the satellites and the ellipse, the satellites continue to have an influence.

2.4 *Effect of satellites persists upon a sheet of translating dots*

Although placing the satellites on a *static* texture does not diminish the effect of the dots, this might be attributed to some special mechanism for static stimuli as opposed to moving stimuli. We wanted to create a display where the stimulus was embedded among additional *moving* features.

The displays were generated in the manner shown in figure 7. We placed the ellipse and the satellites amidst 50 dots moving vertically (figure 8 and clip 4). The satellites and
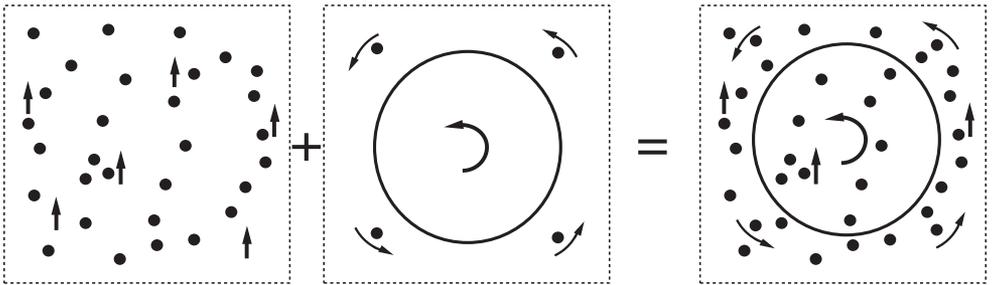

Figure 7. The method of constructing the displays in figure 8. The ellipse and satellites are placed amidst a sheet of translating dots.

appears rigid   appears nonrigid

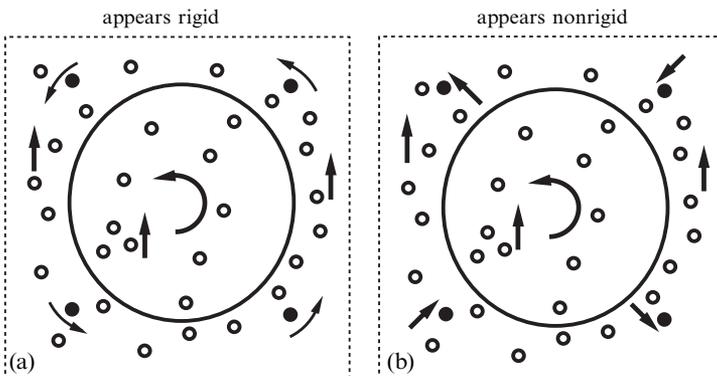**Figure 8.** The effect of satellites persists upon a sheet of translating dots. The 4 rotating (a) or deforming (b) dots have a dramatic influence on the perceived motion, while the 50 dots that move vertically do not influence the perceived motion. They are perceived as lying on a different surface. The vertically moving dots are shown here unfilled, but all dots were identical in the actual display.

the 50 additional dots were identical except for their motions. As with the static texture, the effect of the satellites persists. Observers continue to report that the perceived motion of the ellipse changes dramatically as the motion of the satellites is varied. The large effect of the 4 satellites should be contrasted with the lack of influence of the 50 translating dots. Observers report seeing the vertical dots as belonging to a single surface, while the ellipse and the 4 satellites appear to be on a separate surface.

2.5 *Effect of satellites persists at rather large distances*
In the previous display, many of the vertically moving dots were far from the ellipse contour. This raises the possibility that their lack of influence on the ellipse is simply a result of the distance. We wanted to see whether the effect of satellites is restricted only to dots that are very close to the contour.

   We increased the distance between the satellites and the ellipse contour (figure 9 and clip 5). Subjects reported that the satellite effect was reduced. However, some effect persisted at rather large distances. Even when the distance of the satellites was equal to the minor axis of the ellipse, subjects reported a marked difference between the perception of the ellipse in the two conditions—rotating sats and nonrigid sats.

2.6 *Competition between satellites at different distances*
To further study the decrease of influence as a function of distance, we constructed a stimulus in which satellites at different distances compete (figure 10 and clip 6). The inner satellites underwent rigid motion and the outer dots underwent nonrigid motion.
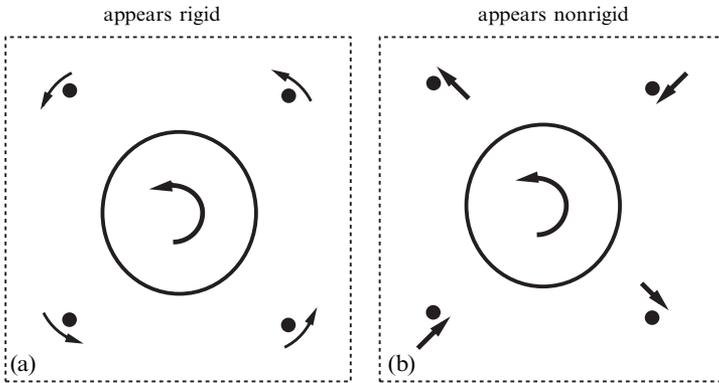
Figure 9. The effect of satellites persists at rather large distances. Even when the satellites are at a distance that is equal to the minor axis of the ellipse, the perceived motion varies when the motion of the satellites is changed. The influence on the motion, however, is weaker at larger distances.
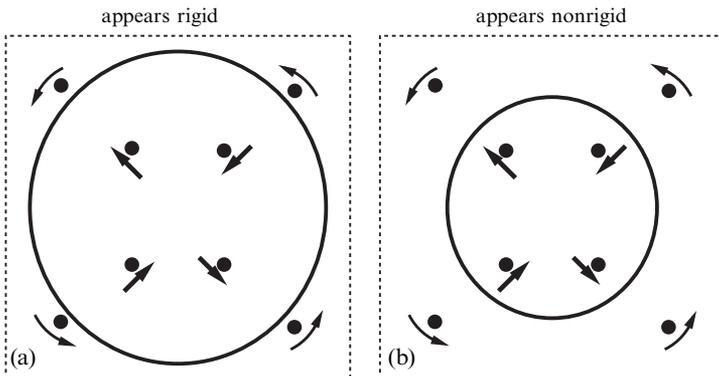
Figure 10. A display in which the relative distance between the ellipse and the satellites is varied. (a) With a large ellipse, the outer satellites win. (b) With a small ellipse the inner satellites win.

Subjects reported that the closer set of satellites won. Thus in figure 10a the ellipse was perceived as rotating rigidly while in figure 10b it was perceived as deforming nonrigidly, even though it was always doing the same thing. When the ellipse was roughly halfway in between, observers reported a bistable percept that they could often flip at will.

Note that in this display we held the satellite locations constant and varied the radius of the ellipse. Similar results were obtained when we varied the satellite locations and held the ellipse radius constant. Also, similar results were obtained when the inner dots underwent rigid motion and the outer dots underwent nonrigid motion. In both cases, the closer set of satellites won.

2.7 *Effect of satellites persists when they are displaced in depth*

When stereo disparity was used to place the satellites at a different depth plane than that of the ellipse, we expected the effect of the satellites to disappear (cf Shiffrar et al 1995). Surprisingly (figure 11 and clip 7), the satellites continued to exert a powerful influence on the percept of the ellipse. Observers reported seeing a significant depth difference between the ellipse and the satellites; at the same time, the percept of the ellipse switched from rigid to nonrigid depending on the motion of the satellites.
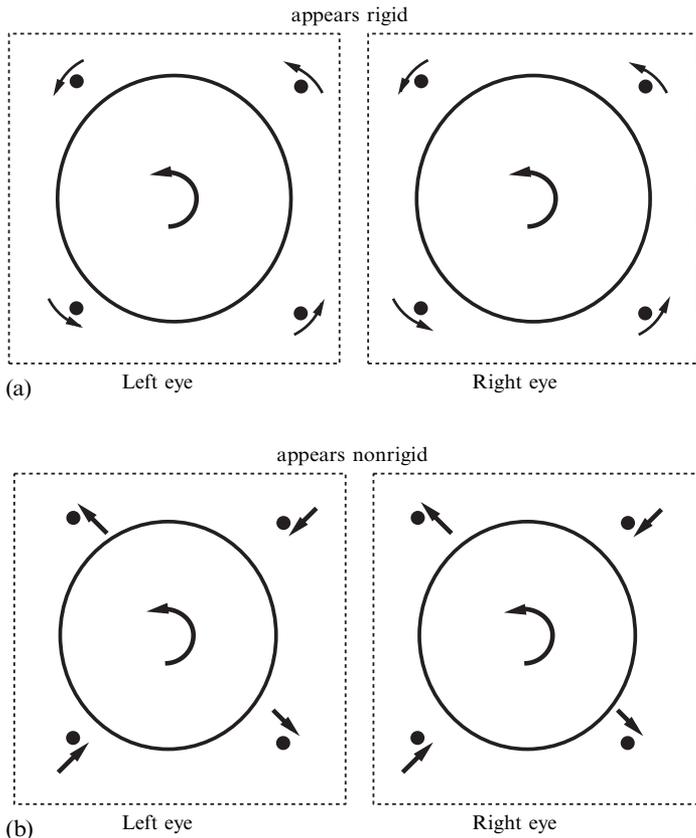

Figure 11. Stereo pairs showing an ellipse at a different depth plane than the satellites. Even when the satellites and the ellipse are placed at different stereo depth planes, the perceived motion of the ellipse continues to be strongly influenced by the motion of the satellites.

2.8 Competition between satellites at different depths

In analogy with the effect of 2-D distance, we wanted to see whether a competition stimulus would reveal the effect of stereo depth.

We created a display with two sets of dots whose image-plane distance from the ellipse contour was equal. However, when this display was viewed in stereo, the two sets of dots were seen to lie on different depth planes (figure 12). When we moved the ellipse in depth (clip 8) we again found a shift in the percept of the ellipse—it appeared to move with the dots that were closer to its depth plane. This effect, however, appears to be rather weak. Out of five subjects in the present experiment two subjects reported a strong effect of the satellites, two reported a weak effect, and one reported no effect.
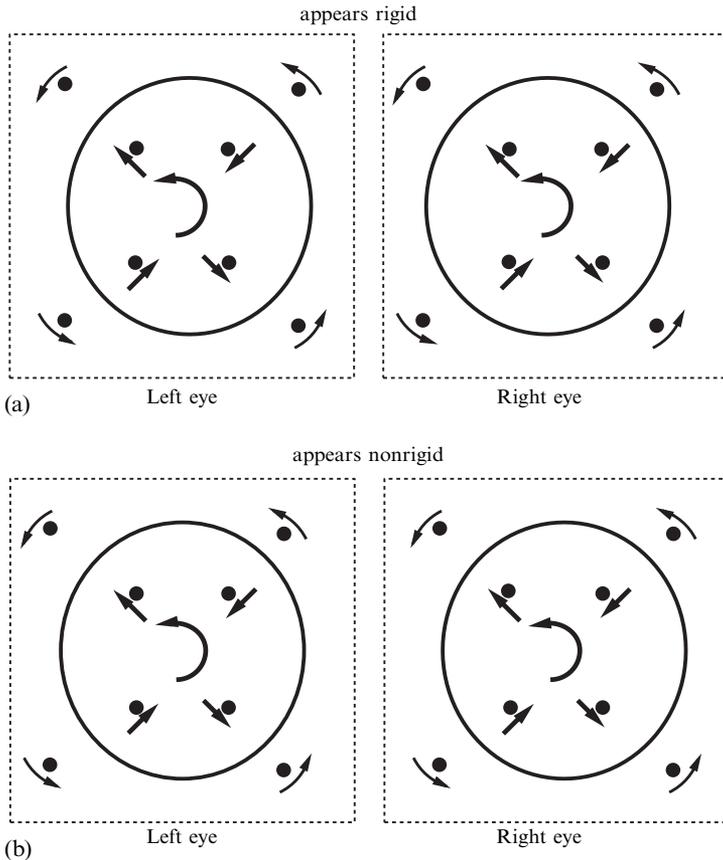
Figure 12. (a) A stereo pair with two sets of satellites at different depth planes; the ellipse is in the same depth plane as the outer dots. (b) A stereo pair with two sets of satellites at different depth planes; the ellipse is in the same depth plane as the inner dots. The ellipse appears to move with the set that is closer in depth.

2.9 Summary of phenomena

The perceived motion of a plain rotating ellipse is a smooth velocity field that may be rigid or nonrigid. Satellites off the ellipse may influence its motion, provided they move in a way consistent with the motion of the ellipse. Adding a second layer does not block the influence. The influence of satellites depends to some extent on their proximity in 2-D space, and depends slightly on proximity in depth. These are qualitative conclusions that constrain motion analysis models. In the next section, we implement representative models to get more quantitative constraints.

3 Models

In order to understand the constraints the phenomena place on models of motion analysis, we implemented three representative models. We chose these models because they can be applied directly to the stimuli shown to humans. Unlike verbal models that can only be applied to an experimenter's description of the stimulus, these models calculate a velocity field for an arbitrary image sequence. We compared this predicted velocity field to the percept reported by our observers.

The three models are each representative of a class of models. All classes share an assumption of smoothness but define smoothness slightly differently—local smoothness in a window, smoothness along contours, and global smoothness. Figure 13 illustrates the three classes of algorithms. The shaded area denotes the region over which smoothness is assumed. In least-squares algorithms, the motion is assumed to vary smoothly within a small 'window' of the image. In contour smoothness, the motion along the contour is assumed to vary smoothly, while in global smoothness the motion is assumed to vary smoothly over all the image. These smoothness assumptions lead to predictions of influence given the three models—each model predicts that only features within the shaded area can influence the perceived motion of the ellipse. We now give a more detailed explanation of the three algorithms implemented.
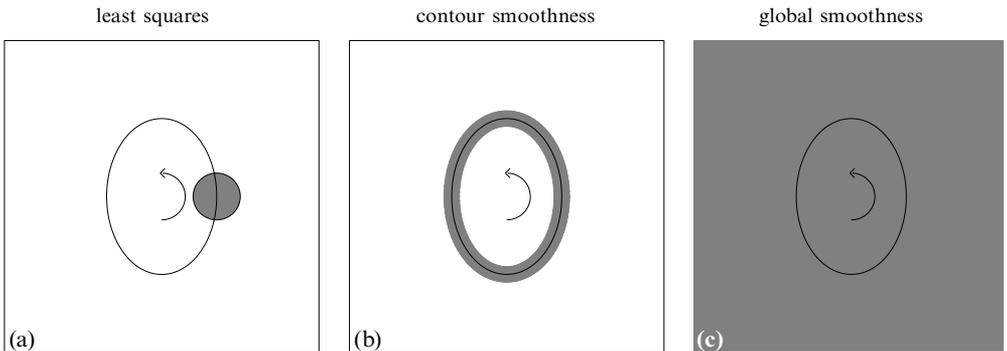
least squares   contour smoothness   global smoothness

(a)   (b)   (c)

Figure 13. The three classes of algorithms compared. The shaded area denotes the region over which smoothness is assumed. The algorithms predict that features within the shaded area will influence the perceived motion of the ellipse.

3.1 The algorithms

3.1.1 *Local least squares* (Lucas and Kanade 1981). Local-least-squares algorithms are based on the idea that even though a single image location does not contain sufficient information to solve for the local velocity (again, the 'aperture problem'), a local patch typically will have sufficient constraints. The Lucas and Kanade (1981) algorithm finds the velocity vector that best fits all constraints within a local region. The phrase 'best fit' should be understood in the least-squares sense. The algorithm finds the vector $\boldsymbol{v} = (v_x, v_y)$ that minimizes:

$$J(\boldsymbol{v}) = \sum_{x,y} w(x, y)[I_x(x, y)v_x + I_y(x, y)v_y + I_t(x, y)]^2, \tag{1}$$

where $w(x, y)$ is a windowing function (eg a Gaussian in space) that defines the local patch; and I_x, I_y, I_t are the spatiotemporal derivatives of the image sequence $I(x, y, t)$. Velocity is estimated at each point in the image based on a window centered at that point. We will henceforth refer to this algorithm as the local-least-squares algorithm.

Lucas and Kanade is an easily implemented area-based algorithm that gives similar results to more biologically oriented area-based algorithms such as Heeger's (1987) and Bülthoff et al's (1989). Nowlan and Sejnowski (1995) have presented an extension to

area-based algorithms whereby within each window one calculates (i) a velocity vector and (ii) a reliability estimate. The responses of the reliability units in their model are difficult to predict since they are based on a training procedure with many examples, but their velocity estimates are similar to those computed by Lucas and Kanade.

3.1.2 *Smoothness along contours.* Hildreth (1983) presented a model that calculates the velocity field of least variation along a contour in the scene. The algorithm begins with extracted normal flow along a contour. That is, it assumes a set of points along the contour indexed by the variable i, at which the normal component $\boldsymbol{v}_i^N$ is known. However, the orthogonal component of the velocity field, $\boldsymbol{v}_i^\perp$, is unconstrained. The algorithm finds these orthogonal components by requiring that the resultant velocity field, $\boldsymbol{v}_i = \boldsymbol{v}_i^N + \boldsymbol{v}_i^\perp$ be maximally smooth. This is accomplished by minimizing a 'non-smoothness' measure:

$$J(\{\boldsymbol{v}_i^\perp\}) = \sum_i \|\boldsymbol{v}_i - \boldsymbol{v}_{i-1}\|^2 . \tag{2}$$

Hildreth also discussed using other non-smoothness measures and we have experimented with them as well. The results do not qualitatively change when we use higher-order derivatives. Hereinafter we call the result of minimizing equation (2) the contour-smoothness algorithm.

3.1.3 *Global smoothness* (Horn and Schunck 1981; Grzywacz and Yuille 1991). Smoothness-based algorithms find the velocity field that fits the local constraints at every image location and is maximally smooth. This is a special case of the regularization approach to computational vision (Poggio et al 1985), whereby smoothness constraints are used to solve ill-posed problems in early vision. Horn and Schunck (1981) presented an algorithm that minimized a sum of two costs, a 'data' term that penalizes for velocity fields that do not satisfy the local constraints and a 'smoothness' term that penalizes velocity fields that change rapidly. If we denote by $\boldsymbol{v}(x, y)$ the velocity at location x, y, the algorithm minimizes:

$$J(\{\boldsymbol{v}\}) = \sum_{x, y} [I_x(x, y)v_x + I_y(x, y)v_y + I_t(x, y)]^2 + \lambda \|\mathbf{D}\boldsymbol{v}(x, y)\|^2 . \tag{3}$$

The differential operator D measures the derivative of the velocity field at every location. Horn and Schunck used the first derivative, whereas Grzywacz and Yuille used an infinite sum of derivatives of different orders. We used here the Grzywacz and Yuille definition of D:

$$\mathbf{D}\boldsymbol{v} = \sum_{n=0}^{\infty} a_n \frac{\partial^n}{\partial \boldsymbol{r}} \boldsymbol{v} , \tag{4}$$

where $\boldsymbol{r} = (x, y)$ and $a_n = \sigma^{2n}/(n!2^n)$. We used $\sigma = 0.7$ where x, y ranged from -1 to 1. We have obtained qualitatively similar results with using only the first derivative ($a_1 = 1, a_i = 0$). Hereinafter we call the result of minimizing equation (3) the global-smoothness algorithm.

The inputs to the local-least-squares algorithm and to the global-smoothness algorithm are spatiotemporal derivatives of the image sequence. We therefore generated two frames of gray-level image sequences corresponding to the stimuli described in the phenomena section and we estimated derivatives using a simple finite-difference scheme (Horn 1986). This simple scheme is only an approximation to the true image derivatives but it works fine when the images are appropriately smooth. In order to avoid temporal and spatial aliasing, we used blurred images moving in small amounts between frames (eg figure 17). For the contour-smoothness algorithm, we used as input the normal velocities at 100 points along the contour of a rotating ellipse.

3.2 *Results*

A shorthand summary of our results is shown in table 1. Since subjects reported only verbally on the perceived motion, we did not conduct a quantitative comparison of the model velocity fields and the subjects report. Rather, an entry 'succeeds' in table 1 signifies a qualitative agreement between the reported percept and the output of the model. An entry 'fails' means that the model cannot even account qualitatively for the human percept; and an entry 'depends' refers to cases when the model's output varies with parameters as does the degree of fit with the human percept.

The fact that none of the models in table 1 can fully account for all percepts is not surprising. After all, models are meant to be simplified abstractions that will not account for all possible data. Thus our goal here is not merely to list failures, but rather to understand what is lacking in these models so that this can aid the construc-tion of new models. We now discuss the successes and failures of each model separately.

Table 1. A shorthand summary of the comparison between the output of three motion-analysis algorithms and the percept reported by observers: 'succeeds' means that the output of the model qualitatively agrees with the report of human observers, 'fails' means the output disagrees, and 'depends' means that the outputs of the models vary with parameters.

Phenomena	Least squares (Lucas and Kanade 1981)	Contour smoothness (Hildreth 1983)	Global smoothness (Grzywacz and Yuille 1991)
(1) Narrow vs fat	fails	succeeds	succeeds
(2) Satellites	depends	fails	succeeds
(3) Sats and background	fails	fails	fails
(4) Sats and moving	fails	fails	fails
(5) Sats at distance	fails	fails	depends
(6) Sats proximity	depends	fails	depends
(7) Sats at depth	NA	NA	NA
(8) Sats stereo proximity	NA	NA	NA

3.2.1 *Local-least-squares.* This model fails to account even for the basic narrow-versus-fat phenomena. This is illustrated in figure 14. For the 'narrow' ellipse that subjects report seeing as roughly rigid (ie the velocity field shown in figure 2a), the estimated velocity fields at the top and bottom of the ellipse are consistent with the percept. However, at the two sides of the ellipse, the estimate is terribly wrong. The algorithm predicts a fast local velocity in the opposite direction to the rotational velocity, and the resulting predicting velocity field is highly nonrigid.

Figure 14b illustrates the source of the failure of the local-least-squares algorithm for the narrow ellipse. The algorithm is searching for the best *translation* that is consistent with the local data, but the ellipse is rotating, not translating. Figure 14b shows two frames from a rotating-ellipse sequence superimposed. For the location at the side of the ellipse, the only location in the subsequent frame that matches lies in the direction opposite to the veridical rotation. This failure is a result of assuming local translation in the algorithm. The fact that humans perceive a narrow ellipse as rotating rigidly suggests that the human visual system does not make the same assumption of local translation.

The local-least-squares algorithm with a fixed window size also fails to explain the effect of satellites at a distance. In this algorithm, the estimated velocity at a point depends only on the information available within a local patch around that point. Thus any features outside this local patch should have no influence on the perceived motion. In particular, satellites that are far away from the contour should have no
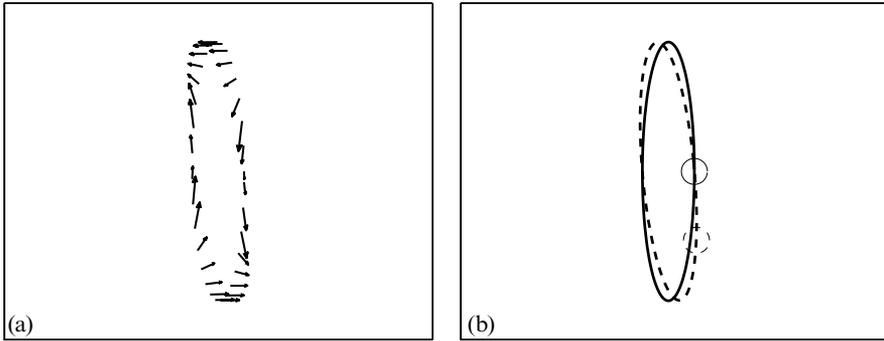
Figure 14. (a) Local-least-squares algorithms such as that of Lucas and Kanade (1981) fail to predict human percept of plain ellipses. For a narrow ellipse, subjects report perceiving motion similar to the rotational flow in figure 2. The local-least-squares algorithm, however, predicts a very different velocity field. (b) The source of the failure for the local-least-squares algorithm: the algorithm attempts to find the best-fitting local translation and in a rotating stimulus this may lead to errors.

influence on the motion of the contour. The fact that humans perceive a different motion for the ellipse, depending on the motion of the satellites, suggests that the human visual system does not make such a strong locality assumption.

3.2.2 *Contour smoothness.* Figures 15a and 15b show the success of the contour-smoothness algorithm in accounting for the narrow-versus-fat effect (this was already shown in Hildreth 1983). Although the predicted velocity field for the narrow ellipse is not completely rigid, it is far more similar to the rotational flow than the predicted velocity field for the fat ellipse. Indeed, if we calculate Hildreth's non-smoothness measure [equation (2)] for the rotational and normal flow in the case of a fat ellipse, we find that the normal flow has higher smoothness given by this measure. In contrast, when we do the same calculation for a narrow ellipse, we find that the normal flow has lower smoothness. The fact that humans perceive a rigid rotation for narrow ellipses and not for fat ones suggests [along with other stimuli discussed in Hildreth (1983) and Mulligan (1992)] that the human visual system makes some assumption of smoothness, not unlike Hildreth's definition of smoothness.

Figure 15c shows the failure of the contour-smoothness algorithm in accounting for the effect of satellites. The predicted velocity field for an ellipse with rotating satellites
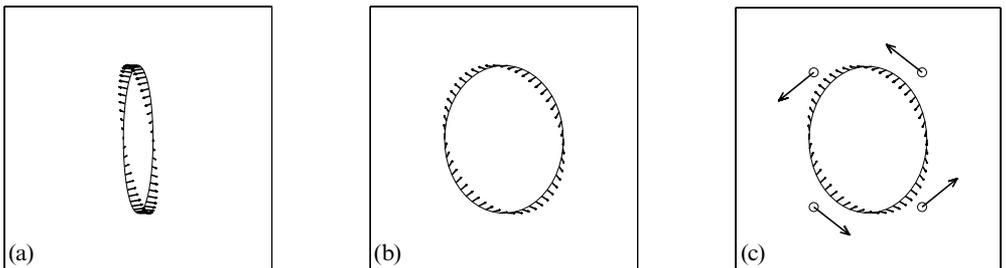

Figure 15. (a) The velocity field predicted by the contour-smoothness algorithm (Hildreth 1983) for a narrow rotating ellipse. The velocity field is more similar to a rotational flow than to normal flow. (b) The velocity field predicted by the contour-smoothness algorithm for a fat rotating ellipse. The velocity field is more similar to normal flow than to rotational flow. Thus, this model qualitatively accounts for the narrow-versus-fat phenomena. (c) The velocity field predicted by the contour-smoothness algorithm for a fat rotating ellipse with 4 rotating satellites. Unlike the human percept, there is no influence of the motion of the dots on the motion of the ellipse.

is identical to the predicted velocity for the plain ellipse (figure 15a). This is of course to be expected from the definition of the contour-smoothness algorithm in which motion information is only propagated along the contour. The satellites that are off the contour therefore have no influence on the perceived motion of the contour.

One way to explain the influence of satellites in the framework of contour smoothness is to assume some sloppiness in the extraction of the contour, and thus satellites near the contour may be mistakenly thought to lie on the contour. However, the fact that satellites continue to exert an influence at large distances from the contour and in the presence of texture makes this explanation unlikely. Thus the ellipse phenomena support the conclusion reached by Grzywacz and Yuille (1991) when they analyzed the Nakayama and Silverman (1988b) translating ogive—propagation of motion signals only along contours is insufficient to account for human perception. Motion constraints must also propagate across space.

3.2.3 *Global smoothness.* Figures 16a and 16b show the success of the global-smoothness algorithm in accounting for the narrow-versus-fat effect. (The algorithm outputs a velocity vector at every point in the image, but for clarity we show these vectors only along the ellipse contour.) In line with the contour-smoothness results, the smoothness assumption here causes the fat-ellipse velocity field to be more nonrigid than that of the narrow one. Figure 16c shows an important success of the global-smoothness approach in comparison with the contour-smoothness approach. When satellites are added to the display, the predicted motion of the ellipse varies according to the motion of the satellites. Thus in figure 16c the rotating sats cause the motion field of the fat ellipse to be almost perfect rigid rotation (compare to figure 16b where the fat ellipse by itself has a very nonrigid motion field predicted by the same algorithm). This result depends somewhat on λ in equation (3)—for large values of λ the dots do not capture the ellipse.
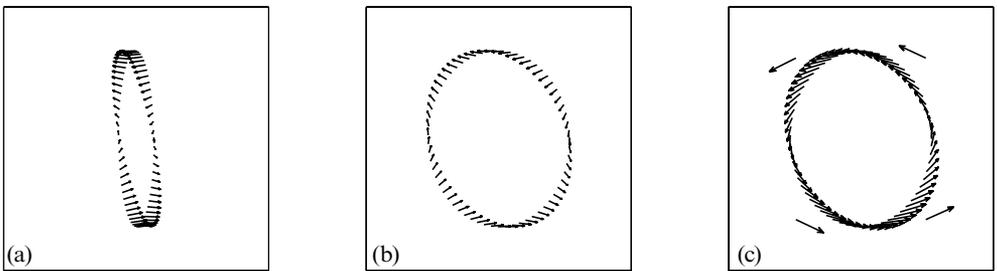

Figure 16. Global-smoothness approaches succeed in predicting perceived velocity in stimuli containing a single motion. (a) Estimated flow for a narrow rotating ellipse which is perceived as rigid; the flow is nearly rigid. (b) Estimated flow for a rotating fat ellipse which is perceived as nonrigid; the estimated flow is nonrigid. (c) Estimated flow for a rotating fat ellipse flanked by rotating dots; the ellipse is perceived as rigid, and the estimated flow is indeed rotational.

This highlights the major distinction between the global-smoothness approach and the contour-smoothness approach: in global-smoothness motion measurements are propagated across 2-D space and hence features off the contour may influence the motion of the contour. The satellite effects on ellipses suggest that propagation of measurements across space is a part of human motion analysis.

Global-smoothness approaches, however, fail whenever the scene contains more than one motion (eg in cases of occlusion and transparency). Figure 17a shows the output of the global-smoothness algorithm for the ellipse with satellites amidst a sheet of vertically moving dots (figure 8). The algorithm attempts to describe the scene with a single smoothly varying motion field, and thus the motion of the ellipse and the rotating
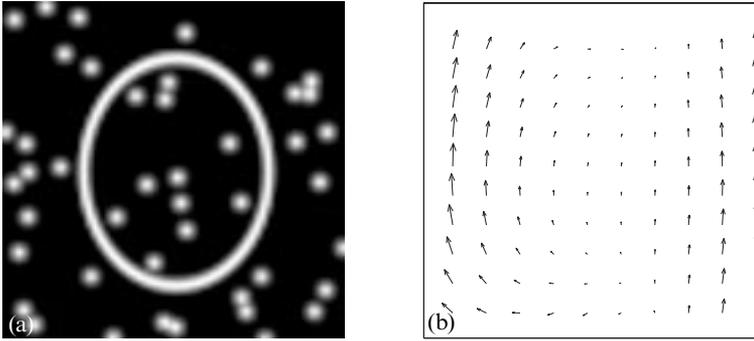
Figure 17. Global-smoothness approaches fail to predict perceived velocity when the scene contains multiple motions. (a) A single frame from a sequence of a rotating ellipse with rotating satellites amidst a sea of vertically moving dots (figure 8); subjects perceive the ellipse to be rigid. (b) The velocity field calculated with the use of the global-smoothness algorithm; the algorithm combines the motion of the ellipse and all dots so that the resulting velocity field does not resemble rotation. Similarly, the algorithm fails when the satellites are placed on a static background.

satellites is estimated very incorrectly. A similar feature is observed when the satellites and the ellipse are placed on a textured background: the static motion signals cause the ellipse to deform nonrigidly. These estimates by the global-smoothness algorithm should be contrasted with the percept reported by human observers for the same stimuli—the motion of the satellites influences the percept of the ellipse, but the motion of the vertically moving dots (or the static texture) does not.

Is this failure of the global-smoothness algorithm a result of the particular parameters we chose? When we change the parameters of the algorithm (eg use different differential operators D or different constants λ) the results remain qualitatively similar. This limitation can not be fixed with simple tweaking of parameters—something is missing from the representation. The global-smoothness assumption represents the scene in terms of a single, smoothly moving object. The failure of such algorithms in accounting for human perception suggests that humans use a more elaborate representation.

3.3 Summary of existing models
We again emphasize that our goal in implementing the three algorithms surveyed above was not merely to show that they sometimes fail. All three models capture important aspects of motion analysis and we wanted to understand the conditions under which they succeed or fail to account for human perception of our phenomena. All three models assume a form of smoothness—implicitly in the case of local-least-squares and explicitly in the case of contour-smoothness and global-smoothness algorithms. The gelatinous percept of fat ellipses suggests that some sort of smoothness assumption is indeed made by the human visual system. At the same time, the phenomena rule out contour smoothness (as in Hildreth 1983) or global smoothness (as in Grzywacz and Yuille 1991). We need a representation that will allow constraints to propagate across space (as in Grzywacz and Yuille) but at the same time allow for multiple layers of motion.

4 Layered models for motion analysis
In computer vision, it has long been recognized that global-smoothness approaches will give erroneous estimates in the presence of occlusion or transparency. One approach to fixing the global-smoothness assumption involves the use of discontinuities or 'line processes' (Geman and Geman 1984; Terzopoulos 1986). Rather than assuming that the image motion can be well explained with a single smooth velocity field, these approaches to motion analysis (eg Horn 1986; Hutchinson et al 1988) allow discontinuities

to form in the velocity field. Thus, in general, the motion of two neighboring locations is assumed to be similar, but if the local motions are highly dissimilar this assumption is abandoned and a boundary is posited between the two locations.

To illustrate the smoothness-plus-boundaries approach consider figure 18. Figure 18a shows a single frame from the ellipse plus static background display (figure 6), and figure 18b shows idealized horizontal local velocity along a single scan line in the image. At locations corresponding to the texture, the local velocity is zero. At locations corresponding to the ellipse contour, and at locations corresponding to the satellites there is a nonzero horizontal velocity, while inside the ellipse the velocity cannot be locally determined.
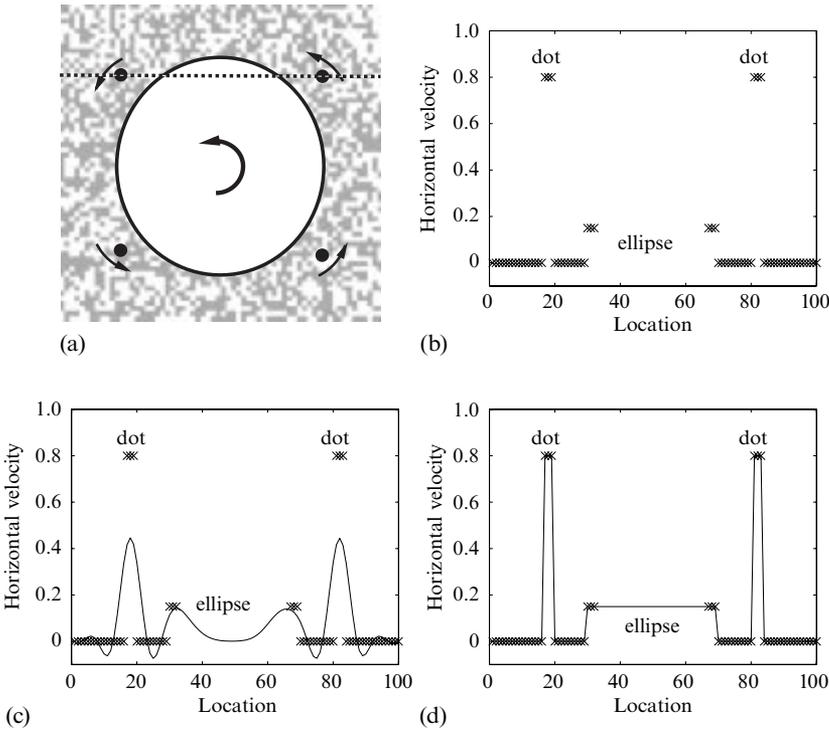

Figure 18. An illustration of the smoothness-plus-boundaries approach. (a) A cross-section from the image sequence of an ellipse and satellites rotating over a textured background. (b) Idealized horizontal motion estimated locally. In practice the motion will be more noisy. (c) The result of smoothing the local flow. (d) The result of smoothing with discontinuities.

Figure 18c shows the output of a global smoother on the local data; by trying to fit all the data with a single, smooth function, such an approach oversmoothes. Static texture is pulled along with the nearby motions, as if the entire image were on a single rubber sheet. Figure 18d shows the output of a smoothness-plus-boundaries approach. The same smoother was used as in figure 18c but no smoothing was done across discontinuities. The advantage of the discontinuity approach is that when the boundaries are estimated correctly, it avoids the oversmoothing problems associated with global smoothness. Thus, in figure 18d, the highly dissimilar velocities in the ellipse and the static background cause a boundary to form and therefore the motion of the ellipse is not influenced by the static texture. This should be contrasted with the results of global smoothness in which constraints from the static texture are propagated to the ellipse contour.

Despite the successes of smoothness-plus-boundaries in computer vision, these approaches cannot fully account for the perception of our phenomena. An important

prediction of this approach is that once a boundary is formed, measurements on one side of the boundary have no influence on the perceived motion on the other side of the boundary. Referring again to figure 18, the highly dissimilar motions of the satellites and the background would cause a boundary to form around the dots. Thus there would be no predicted effect of the satellites on the motion of the ellipse once the display is placed on static texture. As we discussed in section 2, human subjects report a major influence of the satellites even when the satellites are placed on a static background. This suggests that human vision does not use the smoothness-plus-boundaries assumption.

The inability of the smoothness-plus-boundaries approach to propagate information across boundaries has been discussed in the computer-vision literature as well. This has led to the development of alternative approaches known as 'layered models' (Adelson 1991; Darrell and Pentland 1991; Madrasmi et al 1993; Wang and Adelson 1994).

The concept of layered models is based on the representation used by Metelli and others (Metelli 1974) in analyzing transparency displays with 'scission'. In this representation, a scene is assumed to consist of multiple, overlapping layers. Transparency is a case where two or more layers overlap each other, each contributing some fraction of the observed luminance. This sort of representation is also widely used in computer graphics (Duff 1985).

Adelson (1991) advocated the use of layered models for the analysis of moving scenes. Figure 19 illustrates how a layered decomposition can be used in motion analysis. The scene is described by means of overlapping surfaces or layers, each of which has an
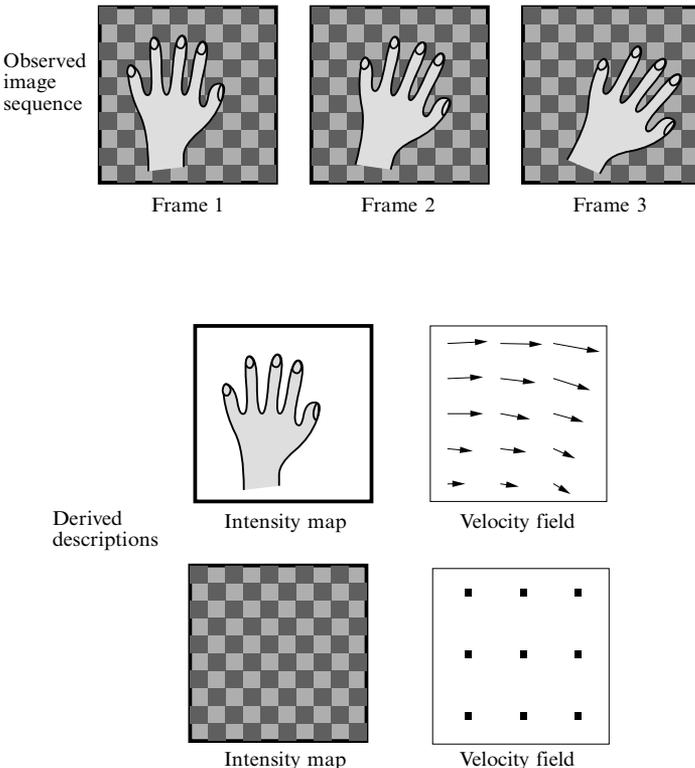

Figure 19. Layered decomposition of image sequences (adapted from Wang and Adelson 1994). In a layered description, an image sequence is decomposed into a small number of occluding layers or surfaces, and each layer has a corresponding motion field. An algorithm that assumes smoothness of the velocity field of each layer is consistent with many of our phenomena.

intensity map and a velocity field. In a full layered representation each layer would also have a transparency map (not shown).

Much progress in motion analysis has been made with the use of such models (Irani and Peleg 1992; Jepson and Black 1993; Ayer and Sawhney 1995; Weiss 1997). Although differing in implementation, these algorithms share the notion of extracting *multiple* velocity fields for each scene rather than a single velocity field as in global smoothness. The scene is assumed to consist of a small number of layers and each layer has an associated velocity field. The algorithms find the layered description that best explains the motion data.

The development of algorithms to extract layered decompositions from image sequences is still a field of active research. The various approaches differ in the following dimensions:

- How to decide how many layers should be used.
- How to describe and estimate the motion field of each layer.
- How to decide which locations in the image should belong to each layer.

As a representative of layered decomposition models, we have calculated the response of the smoothness-in-layers model described by Weiss (1997) to the stimuli described here.

The smoothness-in-layers algorithm finds velocity fields and assignment of pixels to layers from a given image sequence. It does so by minimizing a cost function that favors a small number of layers with smooth motions within a layer. More formally, let K denote the number of velocity fields and $\boldsymbol{v}^k(x, y)$ denote the velocity of layer k at location x, y. The algorithm solves for K, $\boldsymbol{v}^k(x, y)$ as well as soft assignment fields $g_k(x, y)$. These soft assignment fields range from 0 to 1—a number that designates the degree to which the pixel (x, y) is assigned to layer k. We use soft assignment fields because they can represent uncertainty about assignments. For example, if the algorithm is very sure that the location belongs to layer 1 then $g(x, y) = (0.99, 0.01)$, and less certain locations will have $g(x, y)$ closer to $(0.5, 0.5)$.

The velocity fields and assignments are found by minimizing:

$$J(\{\boldsymbol{v}^k\}) = \sum_k \sum_{x, y} \lambda_1 g_k(x, y)[I_x(x, y)v_x^k + I_y(x, y)v_y^k + I_t(x, y)]^2 + \lambda_2 \|\mathbf{D}\boldsymbol{v}^k(x, y)\|^2$$
$$+ g_k(x, y) \log g_k(x, y), \tag{5}$$

subject to the constraint that $\sum_k g_k(x, y) = 1$ and with $\mathbf{D}\boldsymbol{v}$ defined as in equation (4).

It is interesting to compare equation (5) with equation (3). In both equations, there is a 'data' term that penalizes deviation from the constraint line and a 'smoothness' penalty that rewards smooth velocity fields. In equation (5), however, the 'data' term is gated by the assignment field so that a velocity field only needs to lie on the constraint lines at locations that are assigned to that layer. Moreover, the 'smoothness' penalty is applied to each velocity field separately. Thus the algorithm rewards smoothness of the velocity field of each layer, rather than rewarding global smoothness of a single velocity field. As we discuss below, the third term in equation (5) gives an implicit preference for a small number of unique layers.

Figure 20 illustrates the distinction between three approaches to imposing smoothness on a velocity field. Figure 20a shows a graph of hypothetical velocity measurements as a function of position that would arise when one moving surface partially occludes a second moving surface. Figure 20b shows the output when a global-smoothness algorithm is applied to this data. The constraints from each surface are blended with those of the other surface. Figure 20c shows the output when a smoothness-plus-discontinuities algorithm is used. The boundaries disable the propagation between different segments of the same surface. Finally, figure 20d shows the output of a smoothness-in-layers algorithm. Two smooth velocity functions are found, one for each surface.
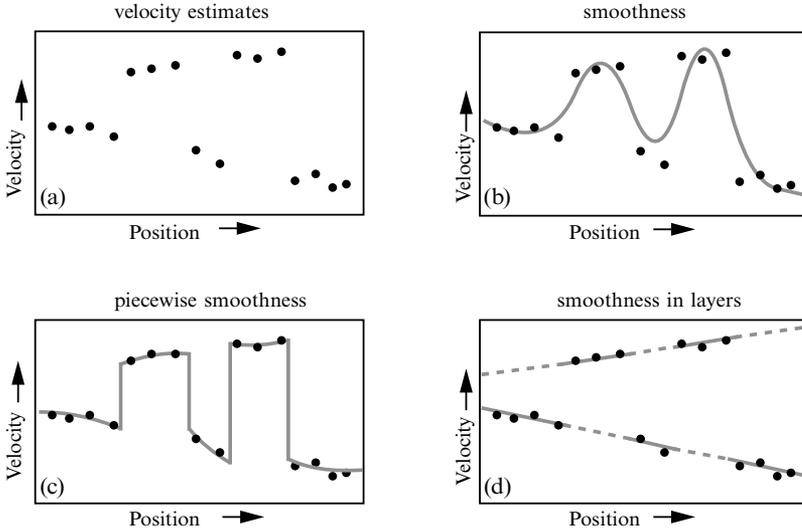
Figure 20. An illustration of the smoothness-in-layers assumption in 1-D (adapted from Wang and Adelson 1994). (a) Hypothetical velocity estimates as a function of position; such data would typically arise from two surfaces in depth. (b) Global-smoothness assumption applied to this data; the measurements from the two surfaces are mixed together rather than segmented. (c) Piecewise smoothness; information is not propagated across discontinuities. The resulting estimate is rather noisy. (d) Smoothness in layers. Two smooth velocity functions are found, one for each surface.

One can show that the cost function J in equation (5) can be derived from a probabilistic model (Weiss and Adelson 1994; Weiss 1998a). It corresponds to the negative logarithmic posterior probability of the velocity fields under a particular choice of prior probabilities. Furthermore, the soft assignments $g_k(x, y)$ correspond to the probability that a pixel belongs to layer k under the mixture model. This correspondence to statistical estimation allows us to use the expectation–maximization algorithm (Dempster et al 1977) to perform the minimization.

The correspondence to statistical estimation also allows us to find the number of layers K that is most probable given the image data (Weiss 1998b). To choose whether one layer or two layers are needed, we compare the value of J in equation (5) with $K = 2$ for the two choices. The one-layer description is encoded as two layers with identical velocity fields. Finding the number of layers in this fashion leads to an implicit preference for a small number of velocity fields (Weiss 1998b)—the third term in equation (5) is minimized when all the soft assignments are (0.5, 0.5) and that happens when the two layers have identical velocity fields. Verbally, the algorithm can be described as trying to find a *minimal* number of layers with *smooth* motions. That is, the algorithm prefers a decomposition consisting of one layer with a smooth velocity field over two layers with equally smooth velocity fields. On the other hand, the algorithm prefers two layers with smooth velocity fields over one layer with a highly nonsmooth velocity field.

The algorithm has two free parameters—λ_1, λ_2—and both are held fixed in the results shown here. We used the same input sequences that were used as input in the previous section—again we used blurred images moving slightly between frames to avoid aliasing.

To summarize, the 'smoothness-in-layers' algorithm receives as input an image sequence; it outputs (i) the number of layers, (ii) a velocity field for each layer, and (iii) a soft assignment of pixels to layers.

4.1 Application to ellipse stimuli
We now describe the results of running the smoothness-in-layers algorithm on the ellipse stimuli. For the narrow-versus-fat effect the algorithm behaves identically to the

global-smoothness algorithms surveyed above. Figure 21 shows the output on a fat ellipse. The algorithm finds that a single, smooth layer is sufficient both for fat and for narrow ellipses. Since there is only one layer here, the assignment function $g(x, y)$ is constant as a function of space and is not shown. In this single-layer case, the algorithm reduces to the algorithm in Grzywacz and Yuille (1991). Note that the algorithm predicts that the background (which is untextured) is moving along with the ellipse.
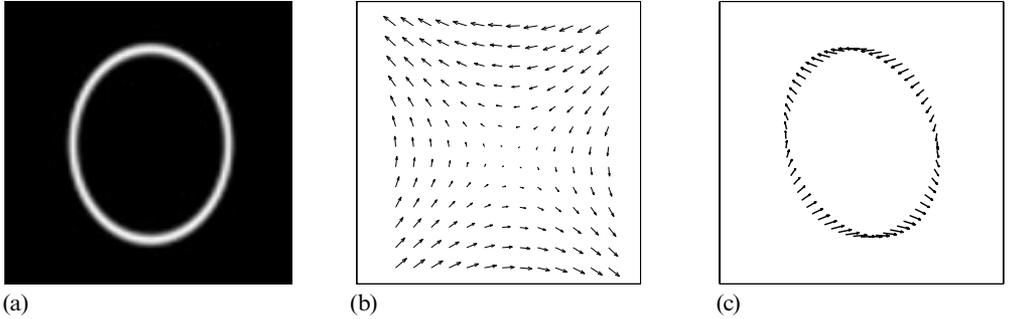

(a)   (b)   (c)

Figure 21. (a) A single frame from a sequence in which an ellipse rotates rigidly in the image plane. (b) The output of the smoothness-in-layers algorithm; a single layer is found with nonrigid deformation. (c) The velocity field in (b) plotted only at locations along the ellipse contour. Note that the result is identical to the global-smoothness algorithm (figure 16). Indeed when a single layer is found, the smoothness-in-layers algorithm reduces to the Grzywacz and Yuille (1991) algorithm.

Figure 22 shows the results on a fat ellipse with rotating satellites. Again, a single layer was found and the assignment function is not shown. Unlike the contour-smoothness models, the algorithm succeeds in predicting the human percept of rigid rotation for the ellipse.

Why does the algorithm predict a rigid rotation for the ellipse with satellite? Note that if the ellipse were seen as deforming nonrigidly then two layers would be needed: a nonrigid layer for the ellipse and a rigid one (rotating) for the sats. On the other hand, a single layer is sufficient when the ellipse is seen as rigid, and that is what the algorithm indeed predicts.

Figure 23 shows the output of the smoothness-in-layers algorithm on a sequence in which the ellipse rotates rigidly in the image plane with 4 satellites amidst a field of vertically moving dots. The assignments are shown as gray-level images—white pixels correspond to high probability of belonging to that layer and black pixels correspond to low probability. Note that for the rotating layer, the pixels corresponding to the
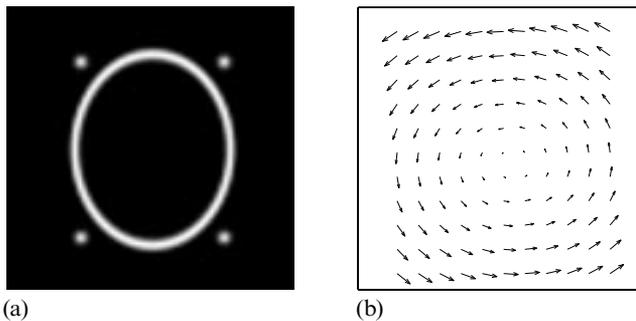

(a)   (b)

Figure 22. (a) A single frame from a sequence in which an ellipse rotates rigidly in the image plane with 4 rotating dots. (b) The output of the smoothness-in-layers algorithm. A single layer is found, with rotational motion.
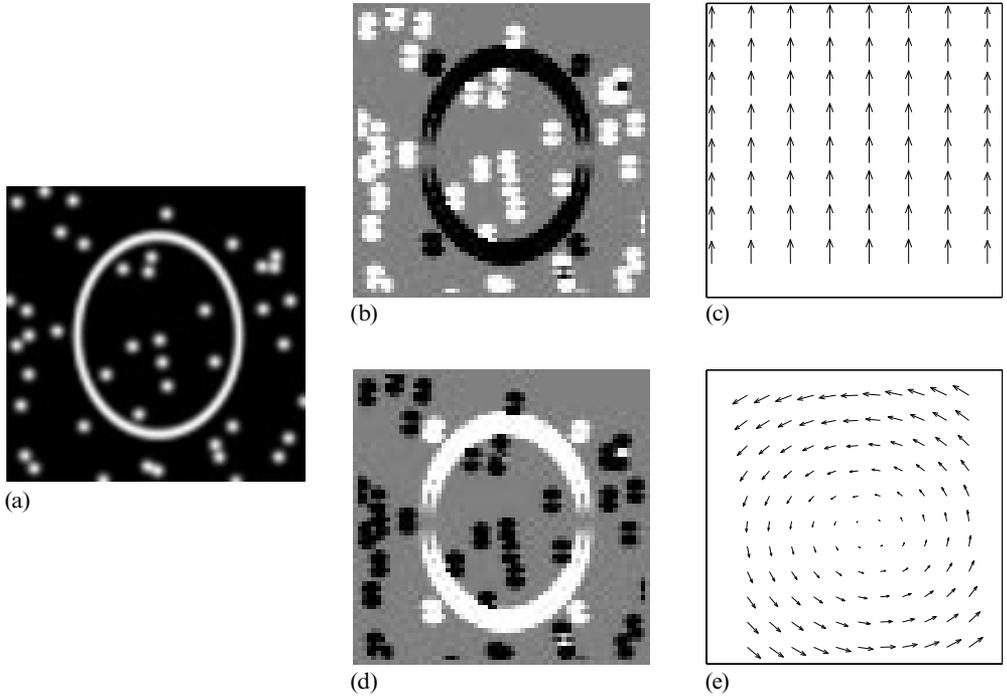
Figure 23. The results of the smoothness-in-layers algorithm when the satellites are embedded in a field of translating dots. Two layers are found, one with rotational motion that includes the ellipse and the 4 dots, and the other with vertical motion that includes the rest of the dots. (a) A single frame from the sequence. (b) Ownership map for the first layer—white pixels correspond to high probability of belonging to this layer. (c) Velocity field for the first layer. (d) Ownership map for the second layer. (e) Velocity field for the second layer.

ellipse and the satellites are white, while the pixels corresponding to the vertically moving dots are black. This means that the algorithm segments the ellipse and satellites from the rest of the display. The ellipse is therefore predicted to rotate rigidly, consistent with the percept reported by human subjects.

To understand the performance of the algorithm, note that to accurately describe the scene with a single layer one would need a very nonsmooth motion field. The two-layer description, on the other hand, is made up of two smooth layers and is therefore preferred by the algorithm.

Figure 23 should be contrasted with figure 17 and with figure 15c. The contour-smoothness algorithm and the global-smoothness algorithm both assume smoothness. Thus the smoothness-in-layers algorithm shares this assumption with them. However, it predicts very different velocity fields for this stimulus. Simply put: contour smoothness predicts no features off the contour will influence the contour; global smoothness predicts all features off the contour will influence the contour. Smoothness in layers, however, predicts only some features will influence the contour. Specifically, only features that are calculated to belong to the same layer will influence the contour. In this stimulus, the smoothness-in-layers algorithm predicts a rigid rotation, consistent with the human percept.

We emphasize that the smoothness-in-layers model implemented here should be viewed as a representative of the class of layered decomposition models. We expect similar results with other layered models. Nevertheless, the success of the smoothness-in-layers algorithm for accounting for the ellipse percepts suggests that human vision may include a notion of smoothness as suggested by Hildreth and others, but smoothness in layers rather than smoothness along a contour or global smoothness.

4.2 *Limitations of simple smoothness-in-layers model*

The simple smoothness-in-layers model described here cannot account for all of our phenomena. Specifically it fails to predict the two competition phenomena (effects 6 and 8) that show a variation in the percept of the ellipse as the relative proximity of the ellipse to each dot is varied. To understand this failure, note that in figure 10 one needs two layers to describe the scene—one for the nonrigid deformation and one for the rigid rotation. Whether the ellipse is assigned to the rigid or nonrigid layer does not change the number of layers needed or their corresponding velocity fields. To account for this percept, the algorithm needs to include a notion of 'proximity'—the algorithm should prefer a decomposition whereby nearby points are assigned to the same layer. As we discuss elsewhere (Weiss and Adelson 1996b; Weiss 1998a), the smoothness-in-layers algorithm can be augmented to include a notion of proximity as well as other nonmotion cues to grouping.

The notion of proximity required to account for our phenomena goes beyond the notion of proximity that is implicitly assumed in most smoothness-based motion algorithms. In all the algorithms surveyed here, the influence of satellites on the motion of the ellipse will decrease as a function of distance—the further the satellites the less the influence. Thus all the models (including the smoothness-in-layers one) would predict that when the satellites are sufficiently far away they would cease to influence the ellipse. A comparison of effects 5 and 6, however, shows that this notion of proximity is not sufficient to explain our phenomena. Both sets of satellites in effect 6 are sufficiently close to the contour to exert an influence by themselves. When both are shown at the same time, however, only the closer set of dots seems to have an influence.

A second limitation of the simple smoothness-in-layers model is that it computes a 2-D motion field. It shares this limitation with all the models surveyed here. Since these models have no representation of the third dimension, they cannot explain the percept of a fat ellipse as a coin rotating in depth. How can this percept be explained?

Ullman (1979) has argued that the perception of 3-D motion is performed in two distinct stages. The first stage computes a 2-D motion field while the second stage interprets this motion field in 3-D—either as the projected motion of a rigid 3-D object or as a nonrigid object.

According to this view of motion perception, the rotating-coin percept is possible only if the 2-D motion field calculated in the first stage is consistent with the projection of the coin's motion. Indeed, it can be shown that a nonrigid flow similar to that computed by smoothness-based algorithms for the fat ellipse is consistent with the projection of a rigid 3-D rotation outside the image plane (Todorović 1993). However, the flow computed by smoothness-based algorithms for narrow ellipses is not consistent with rigid rotation outside the image plane. Thus algorithms such as smoothness-in-layers can account for the fact that fat ellipses are much more likely to be seen as coins rotating in depth than are narrow ellipses. They cannot, however, account for subjects seeing the coin percept rather than the deforming percept—both of these percepts give rise to similar 2-D motion fields.

5 Conclusion

As first reported by Musatti over 70 years ago, an ellipse rotating rigidly in the image plane is often not perceived veridically. We have investigated the perceived motion of such ellipses under varying conditions and have shown that the addition of satellites off the contour has a large influence on the percept of the ellipse. The effect is quite robust and holds when the satellites are at a large distance from the ellipse contour and even when they are placed at different stereo-depth planes. Furthermore, the effect holds when a motion boundary is introduced between the satellites and the ellipse and when the display is placed amidst a field of moving dots to create the percept of transparency.

We have also looked at the effect of satellites on some other ambiguous motion displays, including the translating ogive (Nakayama and Silverman 1988a, 1988b) and a single, diagonal line moving in an aperture (Wallach 1935). We compared the percept of the ogive when the satellites moved rigidly versus the case when the satellites executed normal motion. Similarly, we compared the perceived motion of the line when the satellites moved horizontally versus vertically. In both cases we find the same results — a robust effect of satellites that persists at large distances and across stereo-depth planes.

By calculating the responses of three published computational models to the same stimuli we have shown how these phenomena constrain the types of models needed to account for the human percepts. Although all three models (local least squares, contour smoothness, and global smoothness) capture an essential aspect of motion analysis, none of them fully accounts for the percepts. Local-least-squares and contour-smoothness models fail to account for effects of features at a distance from the ellipse. The global-smoothness model fails to account for the effect of satellites when the display is set on a static background or amidst a sea of moving dots.

We have also shown calculated responses of a recently proposed type of motion models. These layered decomposition models make use of an intermediate level description whereby the scene is described with a small number of layers, each with a corresponding velocity field. We have found that the smoothness-in-layers algorithm that tries to find a minimal set of layers with smooth velocity fields can account for a wider range of phenomena. Taken together, our results suggest the value of an intermediate-level layered description in accounting for human motion perception.

References

Adelson E H, 1991 "Layered representation for image coding" Technical Report 181, The MIT Media Laboratory, Cambridge, MA

Adelson E, Movshon J, 1982 "Phenomenal coherence of moving visual patterns" *Nature (London)* **300** 523 – 525

Ayer S, Sawhney H S, 1995 "Layered representation of motion video using robust maximum likelihood estimation of mixture models and MDL encoding", in *Proceedings of the International Conference on Computer Vision* (Cambridge, MA: IEEE Computer Society) pages 777 – 784

Braunstein M, Andersen G, 1984 "A counterexample to the rigidity assumption in the visual perception of structure from motion" *Perception* **13** 213 – 217

Bülthoff H, Little J, Poggio T, 1989 "A parallel algorithm for real-time computation of optical flow" *Nature (London)* **337** 549 – 553

Darrell T, Pentland A, 1991 "Robust estimation of a multi-layered motion representation", in *Proceedings of the IEEE Workshop on Visual Motion* (Princeton, NJ: IEEE) pp 173 – 178

Dempster A P, Laird N M, Rubin D B, 1977 "Maximum likelihood from incomplete data via the EM algorithm" *Journal of the Royal Statistical Society, Series B* **39** 1 – 38

Duff T, 1985 "Compositing 3-d rendered images" *Computer Graphics* **19**(3) 41 – 44 [Proceedings of SIGGRAPH '85, San Francisco, CA, 22 – 26 July 1985]

Geman S, Geman D, 1984 "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721 – 741

Grzywacz N, Yuille A, 1991 "Theories for the visual perception of local velocity and coherent motion", in *Computational Models of Visual Processing* Eds M S Landy, J A Movshon (Cambridge, MA: MIT Press) pp 231 – 252

Heeger D J, 1987 "Model for the extraction of image flow" *Journal of the Optical Society of America A* **4** 1455 – 1471

Hildreth E C, 1983 *The Measurement of Visual Motion* (Cambridge, MA: MIT Press)

Horn B K P, 1986 *Robot Vision* (Cambridge, MA: MIT Press)

Horn B K P, Schunck B G, 1981 "Determining optical flow" *Artificial Intelligence* **17** 185 – 203

Hutchinson J, Koch C, Luo J, Mead C, 1988 "Computing motion using analog and binary resistive networks" *IEEE Computer Magazine* **21** 52 – 64

Irani M, Peleg S, 1992 "Image sequence enhancement using multiple motions analysis", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Los Alamitos, CA: IEEE Computer Society) pp 216 – 221

Jepson A, Black M J, 1993 "Mixture models for optical flow computation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York: IEEE) pp 760 – 761

Lucas B, Kanade T, 1981 "An iterative image registration technique with an application to stereo vision", in *Proceedings of the 7th International Joint Conference on Artificial Intelligence* (Vancouver, BC: Morgan – Kaufmann) pp 674 – 679

Madrasmi S, Kersten D, Pong T, 1993 "Multi-layer surface segmentation using energy minimization", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York: IEEE) pp 774 – 775

Marr D, Ullman S, 1981 "Directional selectivity and its use in early visual processing" *Proceedings of the Royal Society of London, Series B* **211** 151 – 180

Metelli F, 1974 "The perception of transparency" *Scientific American* **230**(4) 91 – 98

Mulligan J, 1992 "Anisotropy in an ambiguous kinetic depth effect" *Journal of the Optical Society of America A* **9** 521 – 529

Musatti C, 1924 "Sui fenomeni stereocinetici" *Archivio Italiano di Psicologia* **3** 105 – 120

Nakayama K, Silverman G H, 1988a "The aperture problem—I: Perception of nonrigidity and motion direction in translating sinusoidal lines" *Vision Research* **28** 739 – 746

Nakayama K, Silverman G H, 1988b "The aperture problem—II: Spatial integration of velocity information along contours" *Vision Research* **28** 747 – 753

Nowlan S J, Sejnowski T J, 1995 "A selection model for motion processing in area MT of primates" *Journal of Neuroscience* **15** 1195 – 1214

Poggio T, Torre V, Koch C, 1985 "Computational vision and regularization theory" *Nature (London)* **317** 314 – 319

Shiffrar M, Xiaojun L, Lorenceau J, 1995 "Motion integration across differing image features" *Vision Research* **35** 2137 – 2146

Terzopoulos D, 1986 "Regularization of inverse visual problems involving discontinuities" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8** 413 – 424

Todorović D, 1993 "Analysis of two- and three-dimensional rigid and nonrigid motions in the stereokinetic effect" *Journal of the Optical Society of America A* **10** 804 – 826

Ullman S, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)

Vallortigara G, Bressan P, Bertamini M, 1988 "Perceptual alternations in stereokinesis" *Perception* **17** 4 – 31

Wallach H, 1935 "Über visuell wahrgenommene Bewegungsrichtung" *Psychologische Forschung* **20** 325 – 380 [translated into English by S Wuerger, R Shapley, and N Rubin "On the visually perceived direction of motion" *Perception* **25** 1317 – 1367]

Wallach H, Weisz A, Adams P A, 1956 "Circles and derived figures in rotation" *American Journal of Psychology* **69** 48 – 59

Wang J Y A, Adelson E H, 1994 "Representing moving images with layers" *IEEE Transactions on Image Processing Special Issue: Image Sequence Compression* **3** 625 – 638

Weiss Y, 1997 "Smoothness in layers: Motion segmentation using nonparametric mixture estimation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York: IEEE) pp 520 – 527

Weiss Y, 1998a *Bayesian Motion Estimation and Segmentation* PhD thesis, Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA

Weiss Y, 1998b "Phase transitions and perceptual organization of video sequences", in *Advances in Neural Information Processing Systems* volume 10, Eds M Jordan, M Kearns, S Solla, pp 850 – 856

Weiss Y, Adelson E H, 1994 "Perceptually organized EM: a framework for motion segmentation that combines information about form and motion" Technical Report 315, The MIT Media Laboratory, Perceptual Computing Section, Cambridge, MA

Weiss Y, Adelson E, 1995a "Adventures with gelatinous ellipses" *Perception* **24** Supplement, 31b

Weiss Y, Adelson E H, 1995b "Integration and segmentation of nonrigid motion" *Investigative Ophthalmology & Visual Science* **36**(4) S228

Weiss Y, Adelson E, 1996a "Interactions of multiple surface cues in motion grouping" *Perception* **25** Supplement, 6b

Weiss Y, Adelson E H, 1996b "A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (New York: IEEE) pp 321 – 326

APPENDIX

Parameters used in the experiments

The ellipses oscillated about their centers—the angle of the ellipse relative to the vertical axis changed as a sawtooth function with an amplitude of $25°$. The rotation speed was $72° \ s^{-1}$ so that each period of the sawtooth took approximately 1.38 s. Each clip contained two conditions that were shown in alternation. Thus for clip 1, the two conditions were 'narrow' and 'fat' and the clip contained 4.5 s of 'narrow' followed by 4.5 s of 'fat' followed by 4.5 s of 'narrow' etc. Each clip was shown until the subject was comfortable describing his or her percept (typically after one or two showings of each condition). The stimuli were generated in a 400×400 pixel window that was shown on a monitor so that 400 pixels extended over 30 cm. We allowed subjects to adjust their viewing distance (in pilot experiments we found that the percepts are not changed over a range of viewing distances from 0.2 to 2.0 m). We now give details for each condition.

1. *Narrow versus fat.* The narrow ellipse had major axis of 304 pixels and minor axis of 61 pixels (aspect ratio 0.2). The fat ellipse had major axis of 304 pixels and minor axis of 277 (aspect ratio 0.91).

2. *The influence of satellites.* The ellipse had major axis of 304 pixels and minor axis of 271 pixels (aspect ratio 0.89). The four satellites were white circles with a radius of 4 pixels. The distance from the edge of the satellites to the ellipse contour was 7 pixels.

3. *Effect of satellites persists over static background.* Conditions were identical to the previous clip but the ellipse and the dots were both placed on a random-texture background – each pixel in the background was set randomly to be black or gray (quarter the luminance of the ellipse and the satellites).

4. *Effect of satellites persists upon a sheet of translating dots.* This was identical to clip 2 but 50 translating dots were added. Each translating dot was identical to a satellite (a circle with a radius of 4 pixels). The dots translated vertically with a speed of 72 pixels s^{-1}. The direction of motion oscillated with the same period as the rotation of the ellipse.

5. *Effect of satellites persists at rather large distances.* Here the distance between the satellites and the ellipse contour was 74 pixels. The ellipse major axis was 176 pixels and the minor axis 156 pixels.

6. *Competition between satellites at different distances.* Dots belonging to the inner set were at a distance of 95 pixels from the center. Dots belonging to the outer set were at a distance of 190 pixels from the center. The (unfilled) ellipse contour was varied from a distance of 110 pixels to a distance of 175 pixels.

7. *Effect of satellites persists when they are displaced in depth.* The ellipses shown to each eye were 200 pixels by 160 pixels (aspect ratio 0.8). The satellites had a disparity of 7 pixels relative to the ellipse. The satellites were 5 pixels in size and at a distance to the ellipse contour of 8 pixels.

8. *Competition between satellites at different depths.* The ellipses and satellites were the same shape as in the previous clip. The outer dots had a disparity of 7 pixels and the inner dots had zero disparity. The ellipse was flipped from having a disparity of 7 pixels to having zero disparity.