The Ouchi Illusion as an Artifact of Biased Flow Estimation

Cornelia Fermüller Robert Pless Yiannis Aloimonos

Center for Automation Research, University of Maryland

Abstract

A pattern by Ouchi (Figure 1) has the surprising property that small motions can cause illusory relative motion between the inset and background regions. The effect can be attained with small retinal motions or a slight jiggling of the paper and is robust over large changes in the patterns, frequencies and boundary shapes. In this paper, we explain that the cause of the illusion lies in the statistical difficulty of integrating local one-dimensional motion signals into two-dimensional image velocity measurements. The estimation of image velocity generally is biased, and for the particular spatial gradient distributions of the Ouchi pattern the bias is highly pronounced, giving rise to a large difference in the velocity estimates in the two regions. The computational model introduced to describe the statistical estimation of image velocity also accounts for the findings of psychophysical studies with variations of the Ouchi pattern and for various findings on the perception of moving plaids. The insight gained from this computational study challenges the current models used to explain biological vision systems and to construct robotic vision systems. Considering the statistical difficulties in image velocity estimation in conjunction with the problem of discontinuity detection in motion fields suggests that theoretically the process of optical flow computations should not be carried out in isolation but in conjunction with the higher level processes of 3D motion estimation, segmentation and shape computation.

Key words: optical flow, statistics, bias, plaid, motion.

1 Introduction

The perception of motion by visual means plays an important role for many living organisms. Neuroethologists have argued that the most basic capabilities found in animals are based on motion, and the vision of many simple animals, such as medusae, worms and insects, is based entirely on motion. The visual stimulus which forms the basis for all motion interpretation is some form of

2D visual image motion which is derived from the light received over time on the retinae of the eyes. These image motion measurements are further processed by the brain to solve many visual tasks, the most fundamental ones of which include controlling eye movement [Daniilidis, 1996; Stark, 1968; Coombs and Brown, 1993], tracking [Fermüller and Aloimonos, 1992; Smith and Brady, 1995; Koller *et al.*, 1993], segmenting the scene [Nelson, 1991; Zeki *et al.*, 1993], estimating 3D motion and reconstructing the scene in view or some representations of the third space dimension [Daniilidis, 1992; Faugeras, 1992].

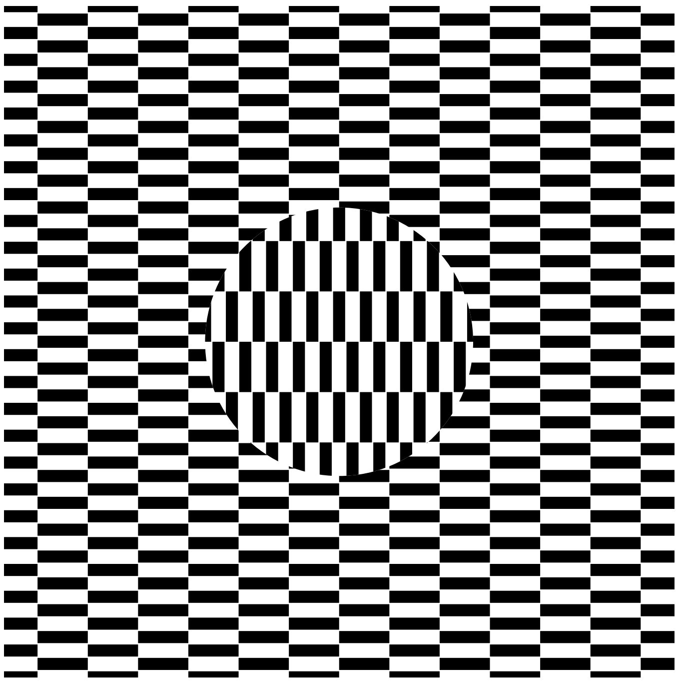


Fig. 1. A pattern similar to one by Ouchi [1977].

For humans and other primates it is considered that two-dimensional image measurements are computed which correspond to the velocity measurements of image patterns, called optical flow. The corresponding field of measurements, the optical flow field, represents an approximation of the projection of the field of motion vectors of the 3D scene points on the image. Computational considerations as well as biological measurements suggest that optical flow is derived in a two-stage process [Adelson and Movshon, 1982; Welch, 1989].

In a first stage, from local image measurements, the velocity component perpendicular to linear features is computed. The situation is illustrated in Figure 2. The velocity vector of a one-dimensional feature (such as a line or piece of contour) viewed through a small aperture is inherently ambiguous, as it is consistent with any vector falling on the constraint line [Wallach, 1935]. Only the velocity component perpendicular to the feature in the direction of the motion is well defined. In the computational literature this component is referred to as normal flow and the ambiguity is referred to as the aperture problem [Marr and Ullman, 1981].

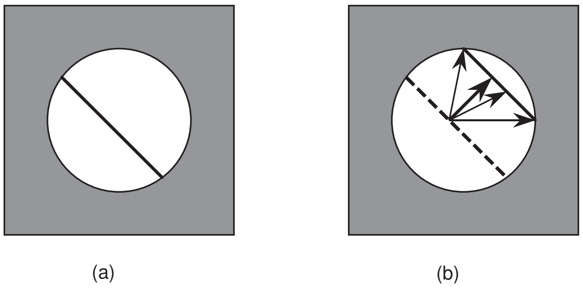


Fig. 2. Aperture problem: (a) Line feature observed through a small aperture at time t. (b) At time $t + \delta t$ the feature has moved to a new position. It is not possible to determine exactly where each point has moved to. From local measurements only the flow component perpendicular to the line feature can be computed.

In order to derive the complete optical flow in a second stage, normal flow measurements from features in different directions residing in a neighborhood are combined. The combination of flow vectors, however, constitutes an intricate computational problem. The 2D image measurements are determined by the 3D motion of the scene relative to the observer and by the depth of the scene in view. Thus, in order to compute the optical flow some models about the 3D scene and the 3D motion have to be employed [Horn, 1986; Horn and Schunk, 1981].

As is well known, the result is that computational problems arise at the locations of flow discontinuities, which are due to objects at different depths or differently moving scene elements. Within small image patches arising from coherently moving, smooth parts of the scene, the optical flow field is well approximated with a parametric model varying, for example, as a constant, linear or quadratic function of the image coordinates [Besl and Jain, 1988]. At the locations of discontinuities, however, it is not, and if image measurements across discontinuities are combined, very erroneous optical flow measurements may be derived [Horn and Schunk, 1981]. To avoid the smoothing over boundaries, knowledge of where the discontinuities are seems to be necessary, which is difficult to obtain from local image measurements.

Even within areas of smooth flow, the computation of optical flow poses a problem. The focus of this paper is to show that for statistical reasons it is very difficult to obtain accurate optical flow estimates. The ideas underlying the statistical explanation of optical flow estimation are as follows: Local one-dimensional flow components—normal flow measurements—are estimated with error. We assume that the estimates of these components are unbiased. However, when combining the one-dimensional measurements in a neighborhood an estimate of optical flow is obtained which is biased. The estimated value depends on the distribution of image gradients, the actual flow, and the error in the normal flow.

The statistical model is used to explain a number of psychophysical findings, which are concerned with the perception of motion in patterns with a sparse, limited set of spatial frequencies. The gradient distribution in the patterns is such that the bias is highly pronounced. In particular, we elaborate on the Ouchi illusion and related experiments [Hine *et al.*, 1995; Hine *et al.*, 1997]. The Ouchi illusion, as shown in Figure 1, consists of two black-and-white rectangular checkerboard patterns oriented in orthogonal directions—a background orientation surrounding an inner disc. Scanning eye movements over these patterns generate the striking perception of relative movement of the inner disc. Our explanation lies in the estimation of differently biased flow vectors in the two patterns which in turn give rise to different 3D motion estimates that cause one pattern to move relative to the other. Furthermore, we explain a number of observations found in the study of moving plaids. These two-dimensional patterns, called plaids, consist of two one-dimensional gratings (such as sine, cosine, or rectangular waves) with different orientations whose motion appears coherent or incoherent depending on various parameters such as contrast, speed, and spatial frequency. They were introduced originally by Adelson and Movshon in [1982] and have since then been studied extensively to assess models used in the explanation of human flow computation [Jasinschi *et al.*, 1992; Yo and Wilson, 1992]. The combination of measurements of patterns different than plaids has sought to determine when smooth contours are seen to move rigidly as opposed to non-rigidly [Nakayama and Silverman, 1988a; Nakayama and Silverman, 1988b].

Models used in the computational and biological literature to estimate the optical flow in a two-stage process can be placed in roughly two categories, those modeling computations in image space and those in spatiotemporal frequency space. The modeling conducted in this paper concentrates on the first category.

In image space models, the one-dimensional motion component of features is estimated by assuming the conservation of image intensity or some function of it. The gradient based approaches assume that image intensity does not change over a small time interval. Denoting the image intensity as I, its spatial (in x and y direction) and temporal derivatives as I_x, I_y, and I_t, respectively, and the velocity of image points in x and y direction as u and v, the following constraint is obtained:

$$I_x u + I_y v + I_t = 0 \tag{1}$$

This equation, called the optical flow constraint equation [Horn and Schunck, 1981], defines the component of flow in the direction of spatial gradient (I_x, I_y) —the normal flow. Other, more elaborate techniques consider functions of the image intensity or the local intensity distribution to be conserved. In order to derive from the normal flow measurements in a neighborhood the optical flow, a second constraint has to be invoked. Usually it is assumed that optical flow varies smoothly. This is achieved by either modeling the flow field explicitly as a polynomial in the image coordinates, or modeling the smoothness through some function in the derivatives of the flow values leading to a regularization formulation [Hildreth, 1984; Horn, 1986; Shulman and Hervé, 1989].

The estimation of flow then amounts to an optimization problem minimizing some function of deviation from the model; usually, a least squares minimization is used. The intersection of constraints (IOC) model often used in the psychological literature is a typical instance of a smoothness constraint. It assumes the optical flow to be constant within a neighborhood, an assumption that is justified within small regions, or if the motion in view originates in a translation due to a fronto-parallel plane.

In the modeling conducted here, we employ the optical flow constraint equation and assume constant flow within a neighborhood. As the psychological experiments analyzed in this paper are concerned with translations in the fronto-parallel plane, this model is approximate and simplifies the exposition. For combining normal flow vectors into optical flow, we use the least squares estimation model. We will show, however, that the bias found in the estimation of flow is not due to the particular models employed, rather it is inherent in the geometry of the constraints placed on combining one-dimensional motion components into optical flow.

The remainder of the paper is organized as follows: In Section 2, we discuss the psychophysical studies detailing the perception of the Ouchi illusion and related biases in the perceptions of plaid motion. In Section 3, we analyze an IOC type model to compute estimates of patch velocity directly from noisy measurements of image derivatives. We then discuss the bias and provide graphic illustrations of the estimated flow for the patterns of limited sets of gradient directions occurring in the psychophysical stimuli. In Section 4, this analysis is used to explain why local patch velocities do not combine to form a coherent perception of pattern motion in the Ouchi illusions and related patterns, and also to explain both coherence judgments and directional biases in the perception of plaid motion. Section 5 is devoted to a general discussion of statistical techniques proposed in the literature on estimation theory to deal with the noise model used in the analysis and the inherent problems in applying these techniques to the problem of optical flow estimation. As will be shown, correcting the bias would require knowledge of noise not attainable from a limited set of measurements of the particular psychophysical stimuli—thus demonstrating that the bias is not a peculiarity of the particular computational models we employ. Section 6 discusses the impact of these findings on current models of motion processing used in the computational and biological sciences.

2 Psychophysical Experiments on Motion Perception

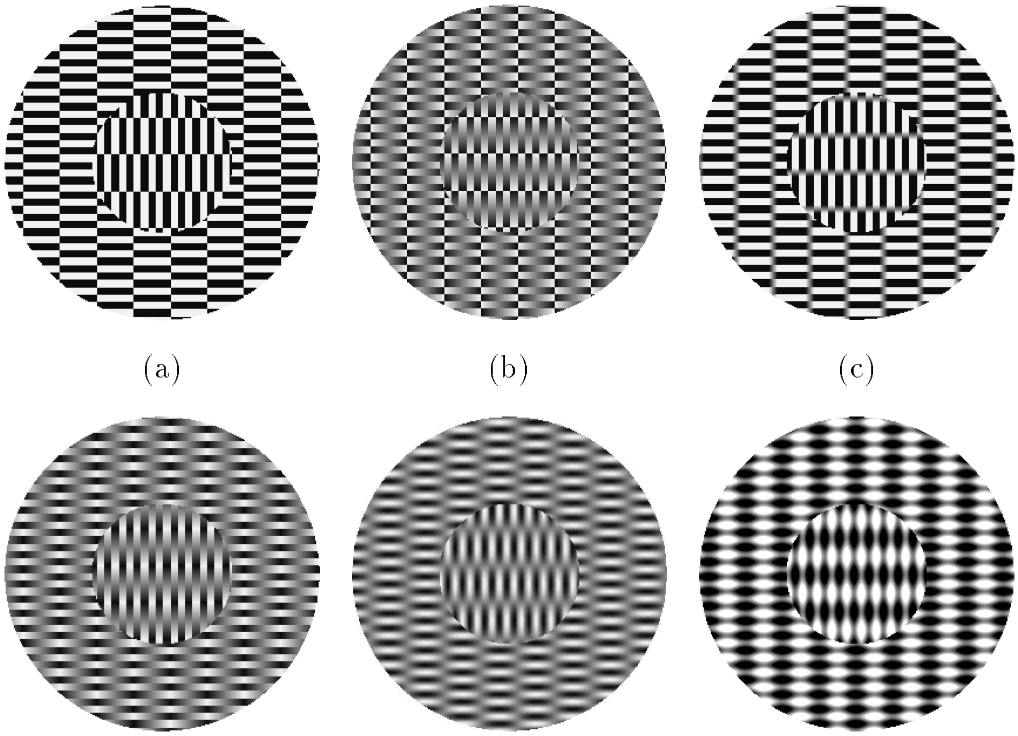
The striking illusion discovered in 1977 by the graphic artist H. Ouchi is evoked by a stationary picture which consists of a checkerboard pattern superimposed on another rectangular checkerboard oriented in orthogonal direction (Figure 1). Small retinal motions, or slight movements of the paper, evince a segmentation of the inset of the pattern, and motion of this inset relative to the surround. The illusion remains under a variety of viewing distances and angles. Some observers report an apparent depth discontinuity, with the center floating as it moves atop the background [Spillmann *et al.*, 1993]. Here, we summarize the findings from psychophysical experiments which have studied this illusion specifically, and then continue with results of plaid experiments which attempt to find general parameters of how local flow measurements are combined.

Khang and Essock [1997b; 1997a] performed experiments with a number of variations of the original pattern to evaluate the impact of various parameters, such as orientation and size of the pattern elements, luminance, and blurring, on the perceived strength of the illusion. In most of the figures they used a simplified version of the illusion with just a one-dimensional square wave grating present in the inset. We concentrate here on the first set of experiments in [Khang and Essock, 1997b] conducted with only two-dimensional patterns. In these experiments they replaced the periodic rectangular checkerboard patterns in the inset and surround by various other 2D periodic patterns, each composed of two 1D functions, one of a short frequency and one of an orthogonal longer frequency.

The particular patterns used, namely the original rectangular checkerboard, a sinusoidal, a trapezoidal, a triangular, a sawtooth and an added sinusoidal pattern are described and shown in Figure 3. Subjects were asked to view the patterns freely and rate the magnitude of the apparent motion; the results of their findings are displayed in Figure 4.

The second set of studies [Hine *et al.*, 1995; Hine *et al.*, 1997] used a simplified stimulus replacing the 2D patterns in the inset and surround with sinusoidally modulated contrast gratings of the same spatial frequency. The two gratings as shown in Figure 5a were tilted symmetrically about the vertical axis: $\theta/2$ degrees to the left and right, respectively. To give the illusion, this stimulus was presented moving vertically on a computer screen, which can be simulated through vertical up-and-down movements of the paper. The apparent motion of the inset is seen orthogonal to the grating orientation, oriented in the direction whose angle with the overall motion of the paper is less than $90°$.

The parameters, which they varied in their figures, were the spatial frequencies



(a) (b) (c)

(d) (e) (f)

Fig. 3. Variations of the Ouchi pattern used in [Khang and Essock, 1997b]. Patterns were formed by combining two one-dimensional periodic functions. (a) Rectangular checkerboard composed by multiplying a horizontal square-wave and a vertical square-wave function. (b) Sawtooth pattern composed of the product of a sawtooth-wave and a square-wave function. (c) Trapezoidal pattern composed of the product of a trapezoidal-wave and a square-wave function. (d) Triangular pattern composed of the product of a triangular wave and a square-wave function. (e) Sinusoidal pattern composed of the product of a horizontal sine wave and a vertical sine wave function. (f) Added sinusoidal pattern composed by adding a horizontal sine-wave and a vertical sine-wave function.

and the angle between the gratings. With short presentation times preventing the possibility of tracking motions, they found that the strength of the illusion of relative motion decreases with the angle between the gratings, and strong responses only for angles smaller than 90 degrees and frequencies between 6–12 cycles/degree as shown in Figure 5b.

Similar findings regarding the dependency on spatial frequency are reported by Khang and Essock [1997a]. In their experiments with one-dimensional gratings in the inset, they have found the illusion to be strongest when the rectangular grid in the surround has elements with size 20–30 min in width and 4–6 min in height, which translates into fundamental components of 6.11 cycles/degree in spatial frequency for the mean rectangle

As a possible explanation for the illusion, Hine et al. [1995; 1997] suggest an

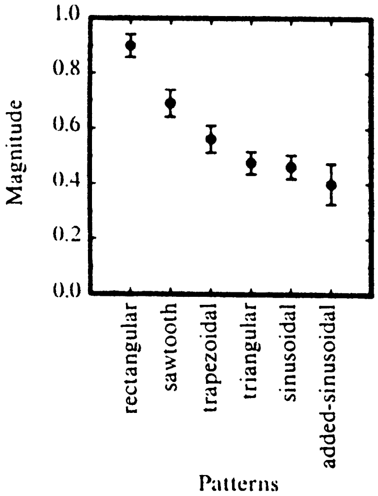
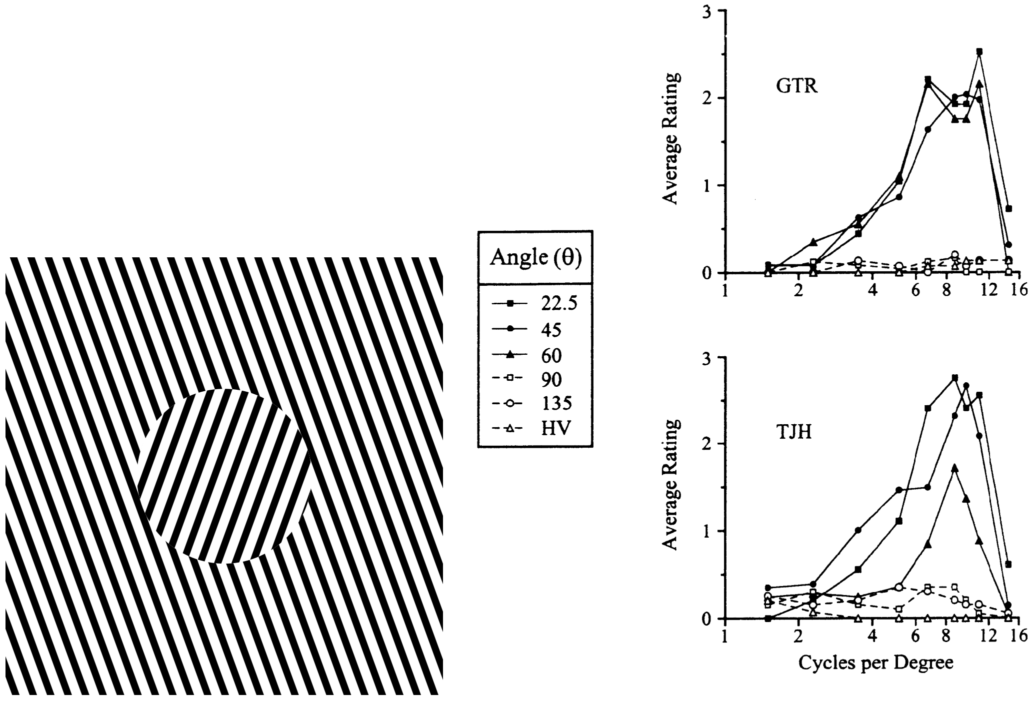


Fig. 4. Means and standard errors of the magnitude of the motion illusion as a function of the six different 2D patterns [Khang and Essock, 1997b].

anomaly of the visual system in integrating local velocity signals into a rigid percept—component motion vectors that differ in direction by more than 120° stimulate entirely different grating cells and motion channels and are not combined. Khang and Essock [1997a; 1997b], cite as a possible cause the interactions between spatially overlapping ON and OFF units—in particular, saturation and non-linear response profiles of visual channels responsive to brightness changes leads to an overall impression of motion. In particular, there is a) a non-linear response of channels responding to luminance change over time, and b) the visual system cannot report accurate local pattern intensity if contrast reversals occur abruptly over large parts of the image. These are distortions of the spatial and temporal image intensity derivatives—a formal analysis of the effects of these errors on the integration of local flow measurements is the focus of this paper.

The integration of local velocity signals has been studied extensively in the context of understanding the perception of moving plaids. Plaids are combinations of two wave gratings of different orientations each moving with a (typically identical) constant speed. For any such "moving plaid", there is always some planar velocity the whole pattern can undergo which would produce exactly the same retinal stimuli. However, for particular variations of the spatial frequencies of the component gratings, their relative orientations, contrasts or speeds, human perception is of two separate motions, with one grating "sliding" over the other. In particular cases, one can perceive a constant, coherent motion of the pattern, biased away from the unique velocity which would account for all retinal signals. Below the findings of a small number of studies are summarized.

Whether or not a particular plaid pattern is judged to be coherent depends upon a number of parameters, such as the contrast, the speed, the spatial



(a) (b)

Fig. 5. (a) Reduced stimulus in experiments used in [Hine *et al.*, 1995; Hine *et al.*, 1997]. The surround and inset gratings were tilted symmetrically about the vertical meridian, each by an angle $\theta/2$ from this meridian. (b) Strength of the relative illusion (evaluated by average ratings) as a function of the angle θ and spatial frequency of the stimulus, plotted for each observer. In HV the inset was vertical and surround horizontal. The acute angles ($\theta < 90°$) produced a greater illusory effect than obtuse angles.

frequency and motion directions of the components. Already in the first study by Adelson and Movshon [1982], it has been found that the perception of coherence decreases as the angle between the motion components perpendicular to the individual gratings increases, and as the difference between the spatial frequencies of the gratings increases. Kim and Wilson gave more quantitative results in a study measuring the effects of the orientations of the gratings (or the motion vectors) and their spatial frequencies [1993]. With a pair of sinusoidal gratings of frequencies 1 and 6 cycles/degree moving at 3 degrees/second, they reported coherent motion when the motion directions differed by 52° or less, and incoherent motion when the directions differed by 126° or more. Plaids with motion vectors 126° apart were consistently judged to be incoherent for plaids of 1 and 6 cycles/degree, while the same directional difference led to coherent motion when the components had frequencies of 1 and 3 cycles/degree.

The motion of a coherent plaid pattern can be theoretically computed using the intersection of constraints model (IOC) — the vector component obtained from each individual grating constrains the local velocity vector to lie upon a line in velocity space, the intersection of the lines defines the motion of the plaid [Adelson and Movshon, 1982]. Some plaid patterns are perceived as coherently moving, but with a velocity different than that predicted by the IOC model. This bias affects both the direction and magnitude of the perceived velocity. In the case of type 1 plaids, that is where the common motion is between the component motion of the two gratings, the velocity of the plaid is biased towards the grating of higher contrast [Stoner *et al.*, 1990; Kooi *et al.*, 1992]. For type 2 plaids, where the IOC velocity is not between the component directions [Ferrera and Wilson, 1987], the bias is towards the average of the component vectors [Ferrera and Wilson, 1991; Burke and Wenderoth, 1993]. Plaids comprised of gratings of different spatial frequencies are also biased in both direction and length [Smith and Edgar, 1991; Kooi *et al.*, 1992]; for example, plaids with orthogonal grating are perceived in direction closer to the gradients of higher spatial frequency than computed by the IOC model [Smith and Edgar, 1991]. In no case is the there an overestimate of the plaid velocity compared to the IOC prediction.

Monte-Carlo experiments have attempted to determine the expected value and variance of velocity calculated with the IOC method, for the case where the one dimensional motion is measured with some Gaussian distributed error [Ferrera and Wilson, 1991; Nakayama and Silverman, 1988a]. Both experiments proceeded by generating a speed measurement for each component direction, corrupting this measurement with Gaussian noise, and then computing the IOC prediction from this pair of constraints. The distribution of estimates created in this method is not biased away from the IOC motion [Ferrera and Wilson, 1991], and the variance of these estimates is correlated with accuracy of directional perception [Nakayama and Silverman, 1988a]. The next section extends the analysis of the IOC model to accept more than two noisy local motion measurements, and finds a bias that is dependent on the distributions of the local orientations.

3 Analysis of Optical Flow Estimation

We analyze the estimation of optical flow from local measurements of changes in the image intensity using least squares minimization of the optical flow constraint equation. We assume that the flow is constant within the region of gradient measurements. As input we consider a set of estimated spatial and temporal gradient measurements $(\hat{I_{x_i}}, \hat{I_{y_i}}, \hat{I_{t_i}})$ which are compounded of the actual values $(I_{x_i}, I_{y_i}, I_{t_i})$ and noise $(n_{x_i}, n_{y_i}, n_{t_i})$.

$$\hat{I_{x_i}} = I_{x_i} + n_{x_i} \tag{2}$$
$$\hat{I_{y_i}} = I_{y_i} + n_{y_i} \tag{3}$$
$$\hat{I_{t_i}} = I_{t_i} + n_{t_i} \tag{4}$$

with

$$\hat{\mathbf{I}}_s = \begin{bmatrix} \vdots & \vdots \\ \hat{I_{x_i}} & \hat{I_{y_i}} \\ \vdots & \vdots \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{I}}_t = \begin{bmatrix} \vdots \\ \hat{I_{t_i}} \\ \vdots \end{bmatrix} \tag{5}$$

The optical flow constraint equation relates the locally image intensity derivatives to the image velocity. Assuming that the optical flow $\mathbf{u} = (u, v)$ is constant within the region considered, it thus is described by the following over-determined system of equations:

$$\hat{\mathbf{I}}_s \mathbf{u} + \hat{\mathbf{I}}_t = 0 \tag{6}$$

Solving (6) by a standard least squares estimation for the flow yields the estimate

$$\hat{\mathbf{u}} = -\left(\hat{\mathbf{I}}_s^T \hat{\mathbf{I}}_s\right)^{-1} \hat{\mathbf{I}}_s^T \hat{\mathbf{I}}_t \tag{7}$$

We consider the effects of the following noise model. The measurement of each image derivative is corrupted by an additive error, these errors are zero mean Gaussian random variables, independent at different image locations, but with possible dependencies between the spatial and temporal derivatives at one location. The second moments of such noise are simply described through a covariance matrix, with one remark. As the model should provide measurements which are symmetric with respect to reflections along the coordinate axes, we assume the noise component due to correlation between the spatial and temporal derivatives dependent in sign on the sign of the derivatives. If one of the derivatives is positive and the other is negative, such as in the first quadrant, we assume positive correlation, otherwise we require sign change. This kind of noise would result if the derivative operations are carried out by a symmetric set of unidirectional derivative operators which are activated selectively depending on the sign of the gradients, and thus collectively performing either forward or backward differentiation.

To obtain a more compressed notation, we also assume the noise in the two spatial components is equal. This might be an oversimplification for real systems, but it does not affect the forthcoming analysis. This means that the variances and covariances of the noise components are given as

$$
\begin{aligned}
E(n_{x_i}) = E(n_{y_i}) = E(n_{t_i}) &= 0 \\
E(n_{t_i}^2) = \sigma_t^2, E(n_{x_i}^2) = E(n_{y_i}^2) &= \sigma_s^2 \\
E(n_{x_i} n_{y_i}) &= 0 \\
E(n_{x_i} n_{t_i}) = \sigma_{xt} &= -\mathrm{sgn}(I_{x_i} I_{t_i}) \cdot \sigma_{st} \\
E(n_{y_i} n_{t_i}) = \sigma_{yt} &= -\mathrm{sgn}(I_{y_i} I_{t_i}) \cdot \sigma_{st}
\end{aligned}
$$

In the absence of error in the spatial gradient measurements $\hat{\mathbf{I}}_s$, standard least squares methods give an unbiased estimator. The expected value $E(\mathbf{u})$, obtained from (7), corresponds to the true optical flow $\mathbf{u}_0$.

However, errors in this measurement matrix can lead to a bias such that the expected value of the estimated flow, $E(\mathbf{u})$, is no longer the true optical flow. The form of this bias is apparent in the second order Taylor expansion of the expected value of the least squares solution as a function of the variance and covariances of the noise in the measurement matrices. According to the noise model, the first order terms vanish. Thus, the only non zero terms that remain in the expansion at zero noise ($\mathbf{n} = 0$) are:

$$
\begin{aligned}
E(\mathbf{u}) = \mathbf{u}_0 + \sum_i \frac{\partial^2}{\partial {n_{(x,y,t)_i}}^2} (\hat{\mathbf{M}}^{-1} \hat{\mathbf{b}}) \rfloor_{\mathbf{n}=0} \frac{\sigma^2_{n(x,y,t)_i}}{2} + \\
\sum_i \frac{\partial^2}{\partial n_{x_i} \partial n_{t_i}} (\hat{\mathbf{M}}^{-1} \hat{\mathbf{b}}) \rfloor_{\mathbf{n}=0} \sigma_{n_{x_i} n_{t_i}} + \sum_i \frac{\partial^2}{\partial n_{y_i} \partial n_{t_i}} (\hat{\mathbf{M}}^{-1} \hat{\mathbf{b}}) \rfloor_{\mathbf{n}=0} \sigma_{n_{y_i} n_{t_i}}
\end{aligned}
\quad (8)
$$

where $\hat{\mathbf{M}} = \hat{\mathbf{I}}_s^T \hat{\mathbf{I}}_s$ and $\hat{\mathbf{b}} = \hat{\mathbf{I}}_s^T \hat{\mathbf{I}}_t$.

Algebraic manipulation of the above derivative leads to an expression of $E\mathbf{u}$ which can be written as a sum of three components: the true optical flow $\mathbf{u}_0$, a component which is due to the variance in the spatial derivative noise only (which we refer to as variant noise), and a component which originates from the covariance terms of the noise in the temporal and spatial measurements (which we refer to as covariant noise). The exact expression is given in the appendix; its dominant factors are

$$
E(\mathbf{u}) = \mathbf{u}_0 - K_1 \left(\sum_i \mathbf{M}^{-1} \mathbf{u}_0 \right) - \sum_i K_{2_i} \mathbf{M}^{-1} \begin{bmatrix} \mathrm{sgn}(\sigma_{xt_i}) \\ \mathrm{sgn}(\sigma_{yt_i}) \end{bmatrix} \quad (9)
$$

with $K_1 = \sigma_s^2$ and $K_{2_i} = \frac{\sigma_{st}}{\sigma_s \sigma_t} \cdot (\sigma_s^2 + \sigma_t^2 + \sigma_s \sigma_t + 2 \frac{\sigma_{st}^2}{\sigma_s \sigma_t} + (\frac{\sigma_{xt_i}}{\sigma_s \sigma_t} u + \frac{\sigma_{yt_i}}{\sigma_s \sigma_t} v)(\sigma_t^2 + 2\sigma_s \sigma_t))$ and $\mathbf{M} = \mathbf{I}_s^T \mathbf{I}_s$, the matrix of uncorrupted spatial gradient values.

The effect of the gradient distribution on the bias of the computed flow can be interpreted through its effect on the matrix $\mathbf{M}^{-1}$. In the case of a uniform distribution of image gradients in the region where flow is computed, $\mathbf{M}$, (and therefore $\mathbf{M}^{-1}$) are multiples of the identity matrix, leading to a bias solely in the length of the computed optical flow. Both the variant term and the covariant term lead to an underestimation in the length. In a region where there is a unique gradient vector, $\mathbf{M}$ will be of rank 1, this is the aperture problem. In the general case, the bias can be understood by analyzing the eigenvectors of $\mathbf{M}$. As $\mathbf{M}$ is a real, symmetric matrix, its two eigenvectors are orthogonal to each other with the direction of the eigenvector corresponding to the larger eigenvalue dominated by the major direction of gradient measurements. $\mathbf{M}^{-1}$ has the same eigenvectors as $\mathbf{M}$ and inverse eigenvalues. Thus, the eigenvector corresponding to the larger eigenvalue of $\mathbf{M}^{-1}$ has a direction dominated by the normal to the major orientation of image gradients, and the product of $\mathbf{M}^{-1}$ with any vector is most strongly influenced by this orientation. This effects the variant term to lead to an underestimation in the length of the optical flow and a bias in direction toward the major direction of gradients. The covariant term in most cases also leads to an underestimation in the length and its influence on the direction can be either way, toward or away from the major direction of gradients, depending on the particular gradient distribution.

The following figures illustrate the bias. Figure 6 displays the expected values of the noise terms for a gradient distribution as it occurs in one of the regions of the Ouchi illusion shown in Figure 1 with blocks four times longer than they are wide. In particular, image gradients are in two orthogonal directions with four times as many measurements in the one direction as in the other. The actual optical flow is along the positive y axis and of length one and the plots show the change in the bias as the relative angle between the perpendicular gradients and the true flow direction varies. The angle θ is measured between the positive x axis and the direction of more gradients; the other gradient direction forms an angle $\theta + \pi/2$ with the x axis (see Figure 6a). Figures 6(b,c) show the error in length and angle due to the variant term and Figures 6(d,e) show the same errors for the covariant terms. The plots are based on the exact second order Taylor expansion as given in the appendix.

For such gradient distribution the bias can be understood rather easily. The eigenvectors of $\mathbf{M}^{-1}$ are in the directions of the two gradient measurements with the larger eigenvalue corresponding to the fewer gradients. As $\mathbf{u}_0 = (0, 1)$, the variant term in (9) leads to a bias in length as shown by the curve in Figure 6b, which takes its minimum at 0 and maximum at $\pi/2$ (that is, when $\mathbf{u}_0$ is aligned with the major gradient direction). The error in angle is greatest for $\pi/4$ (that is, when $\mathbf{u}_0$ is exactly between the two eigenvectors of $\mathbf{M}^{-1}$) and it is 0 for 0 and $\pi/2$ (Figure 6c). Overall, this means the bias due to the variant term is largest when the major gradient direction is normal to the flow and is nearly eliminated when it is aligned with the flow, that is, in the Ouchi pattern, when the long edge of the block is perpendicular to the motion. It is always negative in length and towards the major gradient direction.

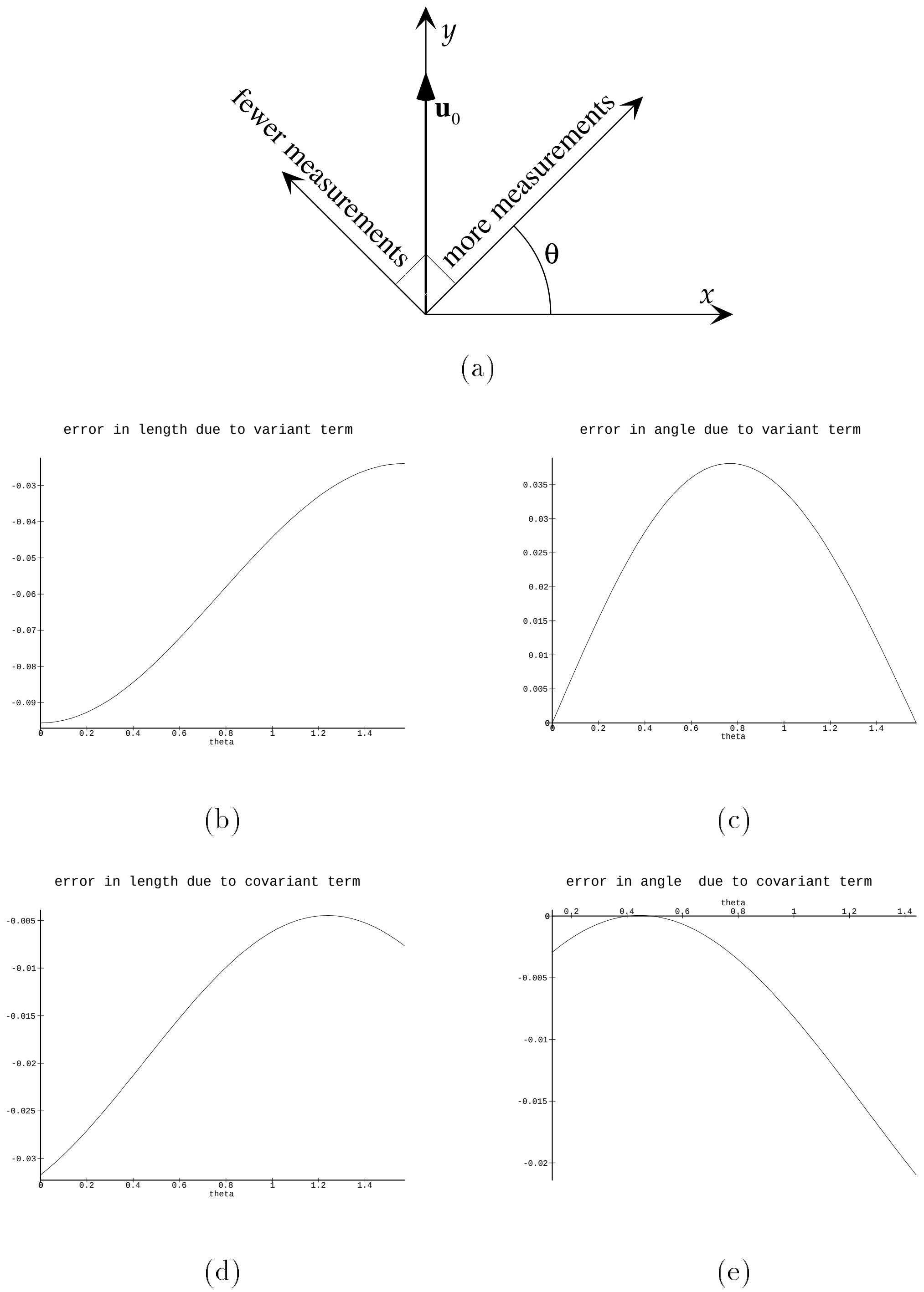
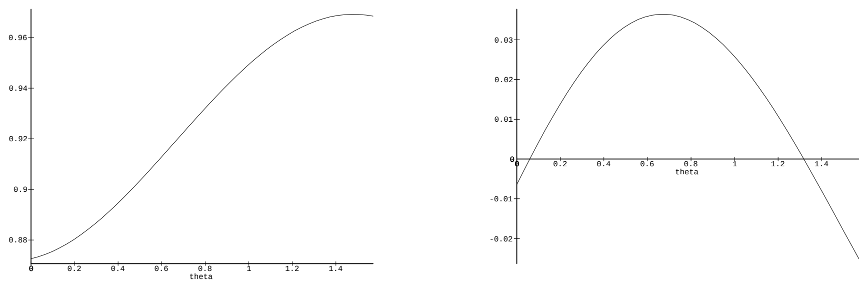


Fig. 6. (a) 16 measurements are in a direction at angle θ from the x axis and 4 measurements are in the direction $\theta + \pi/2$. The optical flow is along the positive y axis and of length one. (b) Expected error in length of variant term. (c) Expected error in angle due to variant term measured in radians between the expected flow and the actual flow. (d, e) Expected error in length and angle for covariant term. The error has values $\sigma_s = \sigma_t = 0.15$ and $\sigma_{st} = 0.1 \cdot \sigma_s^2$.

Regarding the covariant term, as K_{2_i} in (9) is constant within the range of 0 to $\pi/2$ and the vectors $(\text{sgn}(\sigma_{xt_i}), \text{sgn}(\sigma_{yt_i}))$ are along the first and second meridian, the covariant term can be written as $K\mathbf{M}^{-1}\mathbf{a}$, with K a positive constant and $\mathbf{a} = 4(1,1) + (-1,1) = (3,5)$. This leads to error functions, as shown in Figures 6(d,e) which appear to be shifted to the left of the θ axis with regard to the variant bias. This bias is always negative in length and mostly toward the minor gradient direction. The bias for angles θ between $\pi/2$ and π is obtained from the above plots by reflecting the curves upon $\pi/2$ and changing the sign for the error in angle (such that the variant bias is always toward the major and the covariant term mostly toward the minor direction).

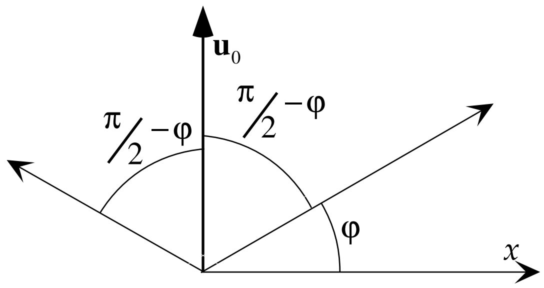
To see the combined effect of the error terms, Figures 7(a,b) show the expected length of the estimated flow and the error in angle for the same configuration as in Figure 6. As we expect the covariant to be much smaller than the variant term, the graph is mostly determined by the latter.



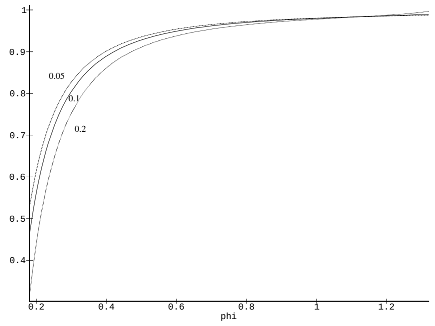
(a) (b)

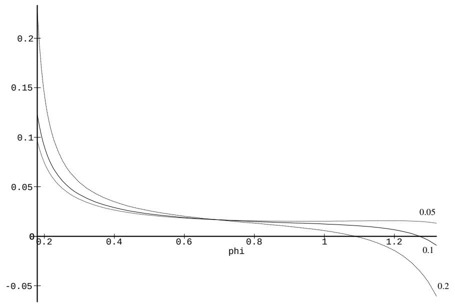
Fig. 7. Expected length of optical flow and expected error in angle for the gradient distribution and error terms of Figure 6.

Figure 8 illustrates the measurements of component motions in the reduced Ouchi stimulus or symmetric plaids. The two gradient directions are symmetric with regard to the y axis (the direction of motion) with the angle φ measured between the direction of more gradients and the x axis (Figure 8a). The correct pattern motion is of length one, and the bias varies depending on the gradient distribution in each receptive field. For a receptive field which has more measurements in rightward component direction, simulating a receptive field receiving more input from the outset of the pattern than the inset, Figure 8 shows the bias relative to the optic flow direction. The same relative number of measurements would come from a plaid pattern made from components of different frequency. The full explanation, in the next section, involves this bias in addition to a segmentation process which also can be explained statistically.



(a)



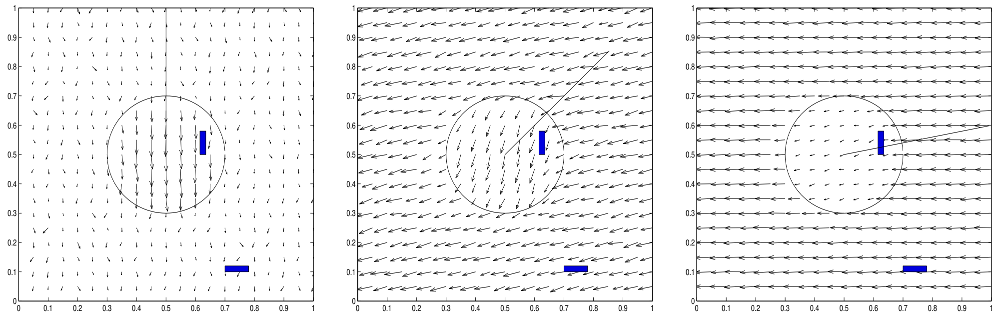


(b) (c)

Fig. 8. (a) Gradient motions in symmetric diagonal directions; For a receptive field with four times as many measurements in the rightward component direction, (b, c) give the expected length of optical flow and the expected error in angle due to both variant and co-variant terms. The noise has values $\sigma_s = \sigma_t = 0.1$ and $\frac{\sigma_{st}}{\sigma_s^2} = 0.05$, 0.1 and 0.2.

4 Explanation

The previous analysis underlies the nature of the Ouchi illusion. The relative angles between the real motion and the predominant gradient direction differ in the inset and the surround, so the regional velocity estimates are biased in different ways. When, instead of freely viewing the pattern of Figure 1, the page is moved in different directions, we observe that the illusory motion of the inset is mostly a sliding motion orthogonal to the longer edges of the rectangle and in the direction whose angle with the motion of the paper is less than 90°. Using Figure 7, it can be verified that the projection of the vector resulting from the difference of the bias vector in the inset and the bias vector in the surrounding area is, for almost all angles in this direction. For example, when the motion is along the first meridian (to the right and up), the bias in the inset is found in the graph at angle $\theta = \pi/4$ and in the outset at $\theta = 3\pi/4$. The two bias vectors are of about the same length, each in direction towards the gradients of the longer edge, and the resulting projection of difference vectors is to the right. If the motion of the paper is to the right, the difference in bias vectors is mostly due to length resulting in a perceived motion to the right, and if the motion of the paper is upwards, the difference vector is downwards. Its projection on the major gradient direction of the inset is close to zero and thus hardly any illusory motion is perceived. Figure 9 shows, for a set of true motions, the biases in the perceived motion. The three bias fields were created with a variety of noise magnitudes, receptive field sizes, and covariance between the noise in the gradient measurements to show the robustness of the effect.

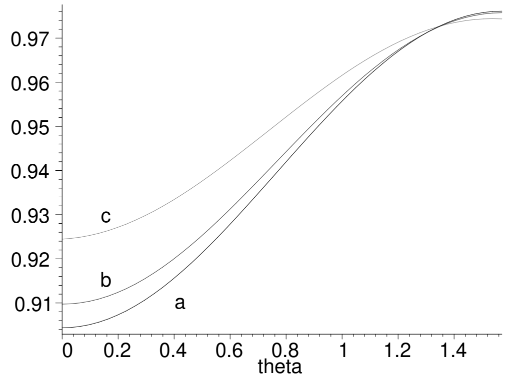


(a) (b) (c)

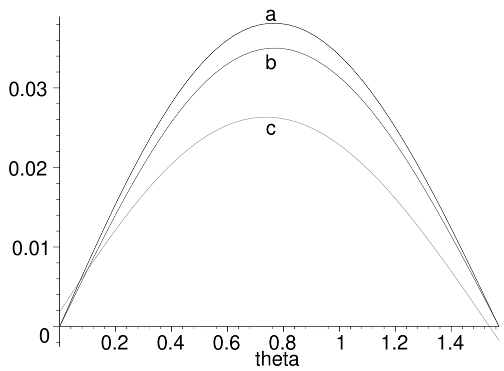
Fig. 9. The regional motion error vector field. The vectors shown are the difference between the true motion and the calculated motion. To derive the sliding motion, compute the difference of the error in the inset minus the error in the surrounding and project the resulting vector on the dominant gradient direction in the inset. One block is shown to show the relative orientation for the inset and the outset of the illusion, the width of the block gives the relative number of vertical and horizontal gradient measures. The line from the center is the direction of the true motion. The noise is Gaussian and the spatial gradient magnitude is one. In (a) and (b), $\sigma_s = \sigma_t = 0.1$ and there is no covariance; in (c) $\sigma_s = \sigma_t = 0.2$ and $\sigma_{st} = 0.2 \cdot {\sigma_s}^2$.

We assume that in addition to computing flow the system also performs segmentation, which is why a clear relative motion of the inset is seen. When experiencing the Ouchi illusion under free viewing conditions, the triggering motion is due to eye movements which can be approximated through random, fronto-parallel translations. As the difference in the bias vectors of the inset and surround has a significant projection on the dominant gradient direction of the inset for a large range of angles (that is, directions of eye movements) the illusion is easily experienced.

In the figures of Khang and Essock [1997b], patterns were used which have more than just two spatial gradient directions. From the rectangular to the sawtooth, the trapezoidal, the triangular, and the sinusoidal to the added sinusoidal, there occurs an increase in the range of gradients. With the spreading of directions, the amount of bias in the estimated flow decreases, as shown in Figure 10, which explains the decrease in the perceived illusory motion found in the experiments.



(a)



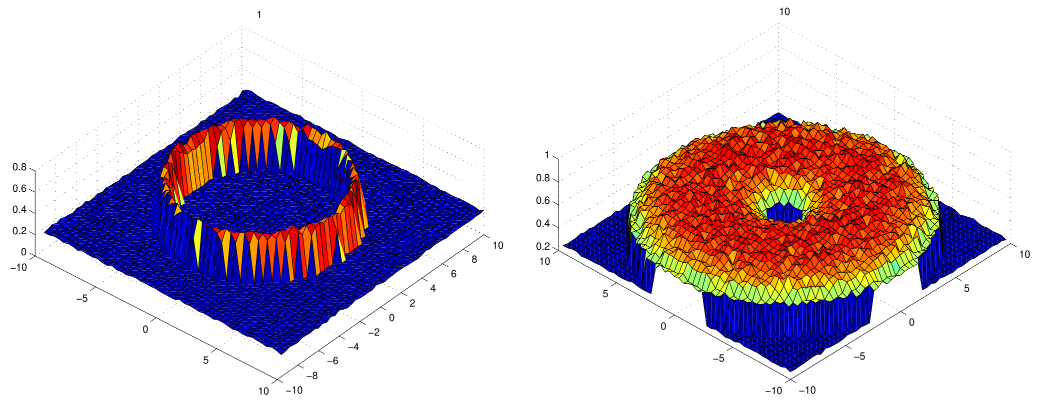
(b)

Fig. 10. Expected length of optical flow and expected error in angle for the following three gradient distributions. (a) Rectangular checkerboard pattern with 16 vectors in major and 4 vectors in minor gradient directions. (b) 12 vectors in major direction, two vectors each at 10° and 20° to the left and right of major direction, 2 vectors in minor direction, (90° from the major) and 2 vectors at 10° to the left and right of the minor direction. (c) Approximation to distribution of gradients of the function $f(x, y) = \sin(x) \cdot \sin(5y)$. The optical flow is $(0, 1)$ and $\sigma_s = \sigma_t = 0.15$. With an increase in the spread of gradient directions, a decrease in the amount of bias occurs.

In the reduced Ouchi illusion [Hine *et al.*, 1995; Hine *et al.*, 1997] the inset and surround regions are each sine-wave gratings of the same frequency, oriented in different directions (Figure 5); solving for the pattern motion requires combining measurements from both the inset and the surround. We assume that there are receptive fields which are large enough to cover both the inset and surround and derive flow from gradient measurements in both orientations. The relative amount of each orientation used in computing a particular value of the flow affects the amount of bias in the final estimate. If, for example, the flow in each region is derived from input which has four times as many gradients from its own region than from the other region, the bias corresponds to the one shown in Figure 8. Different receptive fields will have different relative inputs from each region; we show a 4-1 ratio to allow comparison to the case when gradient directions are perpendicular. Each region gives flow estimates biased by the same amount in length and (for all angles of interest) in direction toward the gradient direction in that region. The projection of the resulting difference vector on the gradient direction of the inset is in the direction whose angle with the overall motion is less than 90°. From Figure 8 it can be seen that the bias in direction decreases with increasing φ. Furthermore, the angle between the difference vector and the gradient direction of the inset increases, causing a decrease in the size of the projection. Thus, an increase in φ (or equivalently, the angle between the gratings) will cause a decrease in the strength of the illusion.

In addition, segmentation processes must be performed which are responsible for the detection and localization of motion boundaries as well as the grouping of single velocity estimates within the motion boundaries into coherent flow regions. A possible measure used in the segmentation is the residual of the least squares solution. The residual is an indicator of how well the computed flow within a neighborhood satisfies the constraints (that is, the optical flow constraint equation). In the absence of noise, the least squares solution of any set of component motions deriving from a single pattern motion will be zero, but in the presence of noise, the residual due to a pattern may be indistinguishable from the residual of a solution when measurements are combined from entirely different pattern motions. This is the case if the values of the residual across different regions are compared and most regions involve only a single orientation.

Figure 11 shows the residual of the least squares solution for the configuration of Figure 8 for different receptive field sizes—the residual takes on a much smaller value in regions where the receptive field only gets input from one component direction. A small receptive field gives a more clear demarcation of where no optical flow fits well with the measured constraints, it is more difficult to localize the boundary with a broader receptive field, which provides an explanation for the dependency of the strength of the illusory perception on the spatial frequency as found in [Hine *et al.*, 1995; Hine *et al.*, 1997] and [Khang and Essock, 1997a].



(a) (b)

Fig. 11. Residual error of the least squares solution, when combining measurements in a small receptive field (a) and a much larger receptive field (b).

The experimental results find the inset to be segmented consistently for grating frequencies between 6 and 12 cpd [Hine *et al.*, 1995; Hine *et al.*, 1997]. The degradation of the illusory perception for higher frequencies is easily explained by the sensitivity exhibited by the human visual system. For grating speeds, as used in these experiments, ranging from 0.2 to 2 degrees per second (the component speed changes with the relative angle of the inset and surround), the human visual system is sensitive to spatial frequencies between approximately 0.5 and 14 cycles per degree [Kelly, 1979]. The degradation for lower frequencies can be explained by the variation of receptive field size of cells in the visual pathway responding to gradients of different frequencies. As shown in Figure 11, the size of the receptive field changes the accuracy with which a boundary can be detected. Whereas high frequencies and thus small receptive fields give rise to a sudden spike in the residual magnitude indicating a likely motion boundary, low frequencies perceived with large receptive fields lead to extended regions of increased magnitude and thus the position of the boundary is much less clear and segmentation is not induced.

The dependency of the residual value on the gradient direction also is in agreement with the experimental data [Hine *et al.*, 1995; Hine *et al.*, 1997]. The residual within regions of more than one gradient direction is largely unaffected by the particular gradient distribution, but the residual in regions of single gradients decreases with the angle between the actual motion and the gradient, and this effects the ratio of residuals across different regions. Figure 12 presents the ratio of the background residual to the residual of a receptive field on the boundary; all points are computed with a Monte Carlo simulation adding Gaussian noise to the derivative measurements, points marked with "X" and "Y" show component motion angles tested in the reduced Ouchi stimulus (in [Hine *et al.*, 1995]). This ratio, too, gives a clear separation (on the y-axis) of the stimuli that were judged to be coherent (those whose orientation difference was less than or equal to 45°), and those that led to a segmentation.

The full explanation for the reduced Ouchi stimulus, which is described in [Hine *et al.*, 1995; Hine *et al.*, 1997] as "a sliding motion of the inset grating with respect to the surround grating", entails a combination of biased smoothing and segmentation on the basis of the residual. As will be further elaborated on in Section 6, we envision ([Fermüller *et al.*, 1997b]) that smoothing and segmentation (together with the processes of motion and shape estimation) are carried out in a feedback fashion, with the localization of motion boundaries succeeding the smoothing of motion vectors and the grouping into coherent regions following localization. Therefore, the segmentation does not cancel out the results of the smoothing. For higher spatial and temporal frequencies the smoothing becomes less important. When moving the pattern of Figure 5 very fast and keeping fixation on the surround the perception appears to be less that of a sliding of the inset but more that of two different motions, one in the inset and another one in the surround, each in the gradient direction of its grating. This also can be explained by the model. As for high frequencies the receptive field sizes of the cells employed are small, smoothing does not take a significant role and the velocity vector in each area is derived from a single orientation, and is thus in the direction of the normal flow.

Regarding the experiments on the judgment of coherence and non-coherence in moving plaids, we assume that the system, in addition to computing optical flow from the normal flow vectors of the individual gratings, can also derive information about the changes of position of the intersection points of the

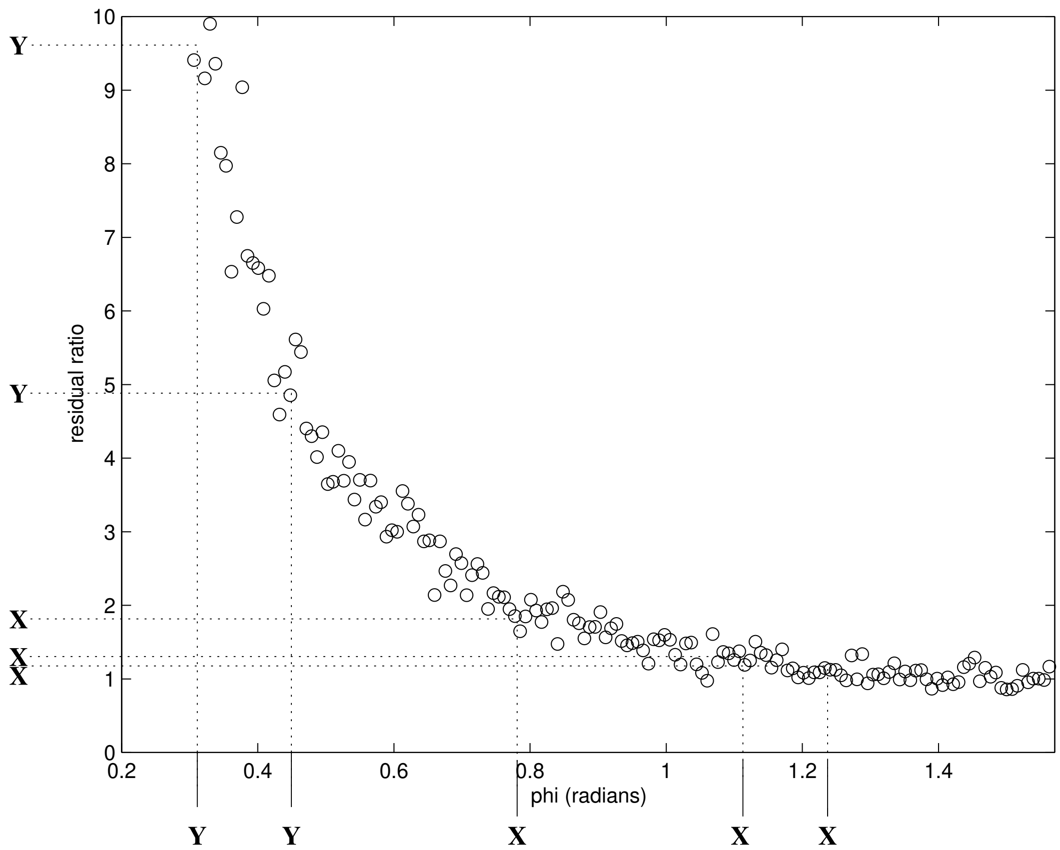
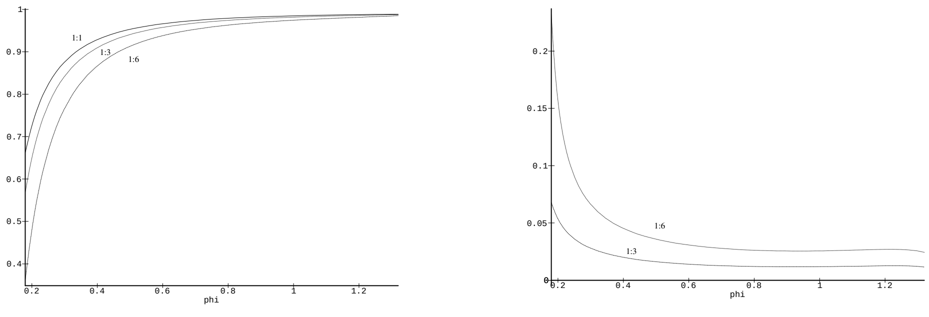


Fig. 12. Ratio of background residual to boundary residual for the configuration of Figure 8a and noise $\sigma_s = \sigma_t = 0.1$ and $\sigma_{st} = 0$. "Y" denotes angles that led to the illusion and "X" angles which did not.

two gratings and can compare the estimated flow to the estimated motion of the intersection points. As can be seen from Figure 8, with the actual motion between the two gradient motion components, a bias is obtained which decreases with φ. The effects of different frequencies can be modeled as different numbers of measurements in each direction. Figure 13 illustrates the dependency of the bias on the ratio of gradient measurements in the two directions for a configuration similar to the one in Figure 8, and it demonstrates that the bias increases as the ratio of the major to minor gradient directions increases. The larger the bias, the larger the discrepancy between the motion of the intersection points and the optical flow, which explains why the perception of coherence decreases with the angle between the motion components of the gratings and the difference between their spatial frequencies, as found in [Adelson and Movshon, 1982; Kim and Wilson, 1993].

For plaid patterns that are perceived as coherent, we can predict the bias in the perceived direction. Recalling that the plaid velocity is biased in direction toward the eigenvector corresponding to the largest eigenvalue of $\mathbf{M}$, we can directly map changes in the plaid pattern to the expected bias. If the contrast of one component sine-wave grating increases, the major eigenvector moves



(a) (b)

Fig. 13. Expected length of optical flow and expected error in angle for gradient distributions with measurements in the ratio ($6:1$, $3:1$ and $1:1$) at angles φ and $\pi - \varphi$ from the x axis. The actual flow is along the y axis and of length 1. For the symmetric distribution (in the ratio $1:1$), no error in angle occurs.

towards the direction of motion of that component. For components of equal contrast and frequency, the major eigenvector is close to the vector average of the component motion vectors. In type II plaids, where both component motion vectors are on the same side of the IOC motion, this gives a bias towards this vector average direction. Different frequencies amount to different numbers of measurements in each direction; this also changes the direction of the major eigenvector. In the case of orthonormal grating, as in [Smith and Edgar, 1991], the larger eigenvector is in the gradient direction of higher spatial frequency, and thus the estimated flow of the plaid should be closer in direction to the motion of the higher spatial frequency grating than predicted by the IOC model.

5 Correcting the Bias?

In the statistical literature the model we used to describe the estimation of flow is referred to as the classic, "Errors-In-Variable" (EIV) model. It is usually expressed in the notation $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A} = \mathbf{A}_0 + \boldsymbol{\delta}\mathbf{A}$ and $\mathbf{b} = \mathbf{b}_0 + \boldsymbol{\delta}\mathbf{b}$ where $\mathbf{A}_0$ and $\mathbf{b}_0$ are the true but unobservable variables (in our case the actual spatial and temporal derivatives $I_{x_i}, I_{y_i}, I_{t_i}$ at points i), $\boldsymbol{\delta}\mathbf{A}$ and $\boldsymbol{\delta}\mathbf{b}$ are the measurement errors, $\mathbf{A}$ and $\mathbf{b}$ are the corresponding observable variables and $\mathbf{x}$ the unknown parameters to be estimated (in our case u and v).

It is well known from the literature that estimation with least squares (LS) generally provides an inconsistent and biased estimate of the true parameter $\mathbf{x}$. The LS estimator gives an unbiased solution only for the regression model, that is, when $\boldsymbol{\delta}\mathbf{A}$ is considered to be zero and measurements $\boldsymbol{\delta}\mathbf{b}$ are independent, zero mean and equal distributed. The literature on estimation theory also provides a wealth of information on techniques dealing with the EIV model and how to compensate for the bias. However, it is very difficult to apply these techniques to flow estimation. In many situations, theoretically it should be possible to improve upon the estimation, but the particular stimuli discussed here pose problems for any statistical procedure.

Before going on, let us clarify one point. The assumption of constant flow is strictly true only for a scene consisting of a fronto-parallel plane moving with translation parallel to the image plane. In general, the motion field on the eye of a moving system depends on the 3D motion and the scene in view and is much more complex. Thus, to cope with general motions and scenes, the processing of flow has to be carried out in several stages. In a first stage, normal flow measurements should only be combined very locally to generate an estimate of optical flow in a small patch of the image. In following stages, flow measurements of neighboring patches can be compared to find larger regions of common 3D motion or to delineate motion boundaries. These considerations exclude models which assume that the motion component in each direction is computed over very large areas and then the single components are combined into a common 2D motion estimate with the simple IOC rule. Such models would not lead to biased optical flow estimates; however, they are of very limited applicability.

Any statistical technique to compensate for the bias requires knowledge of the statistics of the noise. For the noise model considered in the previous sections, this means knowledge of the covariance matrix of the noise vector $(\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_t)$. If such is available, the bias in the least squares estimation could be removed. If the model of constant flow is valid, this can be achieved with the "Corrected Least Squares" estimator. If a more complicated model of general smooth flow within an image patch is necessary, iterative techniques have to be employed.

However, the major problem lies in the acquisition of the statistics of the noise. The noise parameters are not constant, but they change spatially and temporally in complex ways. They depend on many parameters, such as the lighting conditions, the physical properties of the objects being viewed, and the orientation of the viewer in 3D space. Thus, usually there is only a limited amount of data to obtain the noise parameters, and this makes it very difficult to obtain good estimates. This is true even for the simple model used here, because the variance in the motion estimates turns out to be large with respect to the bias. For example, in simulations (see Figure 14), it has been found that for a noise level of 10% (that is, $\sigma_s = \sigma_t = 10\%$ of the value of the spatial gradient and the length of the flow) the standard deviation is twice as large as the bias. Thus, correction, even with an accurate estimate of the bias, in many cases would lead to a worsening of the solution.

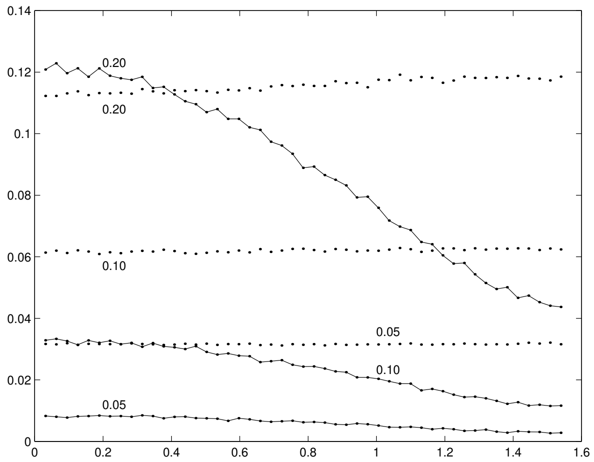


Fig. 14. Expected error in value of length (solid lines) and standard deviation (dotted lines) obtained by a Monte Carlo simulation using Gaussian noise for three different standard deviations: $\sigma_s = \sigma_t = 0.2, 0.1$ and 0.05. The optical flow is $(0, 1)$, the magnitude of the spatial gradients is one, and gradients are distributed with 15 vectors in the direction at angle θ from the x axis and 5 vectors at angle $\pi/2 + \theta$.

In the particular situation of the Ouchi illusion, the 3D motion (either due to random eye movement or jiggling motion of the paper) changes rapidly. This makes the temporal integration of measurements very difficult as the system has only a short time-span to obtain the noise parameters.

In recent years the nonlinear estimator of "Total Least Squares" has received a lot of attention and it has also been applied to the problem of flow estimation [Wang *et al.*, 1992; Weber and Malik, 1995]. This estimator has been shown to provide an asymptotically unbiased solution for the EIV model in the case of white noise, that is, if the noise values are independent, and identically distributed. To whiten the noise, however, again it is necessary to obtain the covariance matrix up to a scale factor. Without whitening, total least squares also gives biased solutions. In addition, total least squares is known to perform very poorly if outliers are present, and these are difficult to detect from few measurements.

The estimation and interpretation of optical flow from a statistical point of view has received attention before in the computational literature [Daniilidis, 1992; Daniilidis and Spetsakis, 1997; Heeger, 1988; Simoncelli *et al.*, 1991; Szeliski, 1990; Weber and Malik, 1995]. It has been pointed out in [Nagel, 1995; Nagel and Haag, 1998] that optical flow estimated using gradient methods is biased. In these studies the bias is interpreted only with regard to the underestimation in the length of the flow; but as shown here the direction of the flow is affected by the bias as well. Also, what is not emphasized there is that it is hard in practice to correct for the bias.

From a computational point of view, the problem of flow estimation is very difficult. In order to obtain very accurate flow estimates, a sufficiently large number of normal flow measurements is necessary. This means that data has to be spatially and temporally integrated via further computational models. As such computational models are based on assumptions about the 3D motion and the scene in view, they are not generally valid for systems moving in varied environments. The integration is possible only in image patches where the data is approximated well by the model. Thus, for the system to use a certain model, it first has to test its validity. For example, in order to employ a model of smooth flow within a spatiotemporal neighborhood, the system has to check for discontinuities in a spatial neighborhood, verify that the flow doesn't change abruptly between frames and evaluate how well the flow is approximated by the particular model used. Clearly, these computations cannot be carried out on the basis of one-dimensional image velocity measurements alone, but require further spatiotemporal 3D information.

6 Computational Models of Motion Processing

The current view which dominates the modeling in the computational sciences as well as research in anatomy, physiology and psychophysics, is that the computation of optical flow is accomplished prior to any other computations involving image motion measurements. First, the optical flow is computed, then it is used to perform other processes, such as 3D motion estimation, segmentation and shape estimation.

Studies in neurophysiology have mapped much of the organization and functional properties of neurons involved in the early stages of motion processing. Evidence suggests that in the primate cerebral cortex the estimation of locally computable one-dimensional motion occurs in the primary visual cortex (V1). There, motion-sensitive neurons with small receptive fields tuned to a specific size and orientation have been found. These cells respond maximally to the direction of motion perpendicular to the cells' preferred orientation [de Valois *et al.*, 1982; Hubel and Wiesel, 1968].

V1 Neurons project onto the middle temporal (MT) area where most neurons appear to respond preferentially to motion [Maunsell and Essen, 1983; Zeki, 1974]. Neurons in this area have considerably larger receptive fields and, in general, the precision of selectivity for directions of motion is less than in V1. MT-neurons feed further motion processing modules, namely FST and MST, where neurons with much larger receptive fields have been found which respond to particular 3D motion configuration [Duffy and Wurtz, 1991a; Duffy and Wurtz, 1991b; Orban, 1992; Ungerleider and DeSimone, 1986]. The most commonly found interpretation is that optical flow is computed in MT, that is, the neurons there integrate the one-dimensional edge motion signals derived in V1 to compute two-dimensional pattern velocity [Heeger, 1987; Movshon, 1990; Movshon *et al.*, 1986; Rodman and Albright, 1989], and this information is sent to higher-level areas to solve navigational tasks.

However, current biological findings don't provide sufficient conclusive evidence as to whether optical flow is computed first, or two-dimensional pattern motion is derived in a combined fashion with 3D motion estimation, segmentation and shape estimation.

Computational considerations suggest that the second approach is advantageous over the first one. On the one hand, the estimation of optical flow and the detection of discontinuities seem to be intricately coupled problems, each difficult to solve by itself. Without knowing the locations of discontinuities, it is hard to estimate flow there, but in order to detect the discontinuities information about optical flow within neighborhoods is needed. On the other hand, as shown in the preceding analysis, it is for statistical reasons very difficult to obtain accurate optical flow estimates—even within areas of smoothly changing flow. Theoretically, to achieve good flow estimation it is necessary to integrate motion information from large neighborhoods. This, however, requires detailed models of the flow field which can only be obtained from additional 3D information. Specifically, this means that knowledge about the discontinuities, the shape of the scene in view, and the 3D motion is necessary.

The only remedy to this chicken-and-egg situation lies in performing the computations of the different processes simultaneously. One can envision an architecture which carries out the computations in a feedback loop. First, approximate image velocity is estimated by combining normal flow measurements. The representation of these estimates does not necessarily have to be a quantitative one, but could be in the form of qualitative descriptions of local flow field patches or bounds on the flow values. The flow computed this way is used to obtain partial shape estimates and perform discontinuity detection, and at the same time an estimate of 3D motion is derived. Then, the computed 3D information is fed back to utilize it together with the image measurements to perform better flow estimation, discontinuity localization, and improve upon 3D motion and structure estimation.

However, even with the best computations, it cannot be guaranteed that optical flow will be estimated accurately all the time, and this has to be taken into account when conducting visual navigation processes. Most computational models assume generically computed flow which is used to obtain accurate 3D motion and metric shape estimation. Consideration of the computational difficulties calls for more purposive computations. Depending on the particular computation of 3D information, different representations of flow may be derived. For example, instead of attempting accurate egomotion estimation from optical flow, the approximate directions of translation and rotation can easily be obtained from patterns of the sign of normal flow [Fermüller and Aloimonos, 1995a]. Instead of reconstructing the scene in view, it is computationally more feasible to derive less powerful shape representations sufficient for particular tasks, for example, representations, which only describe the qualitative shape of scene patches, or only allow an ordering of scene elements with respect to their depth value [Fermüller and Aloimonos, 1995b; Fermüller *et al.*, 1997a]. Also, instead of attempting segmentation straightforwardly from image measurements, segmentation may be conducted only in conjunction with other tasks.

7 Conclusion

We have shown in this paper the problems of estimating two-dimensional image velocity from local one-dimensional motion measurements from a statistical point of view. As noise affects local motion measurements, that is, normal flow vectors, in both length and direction, the estimation of optical flow is biased. Theoretically, the design of any unbiased estimator would require knowledge of the statistics of the noise which often is hard to obtain. To model the computation of optical flow we used linear least squares estimation and showed that this model explains a number of psychophysical experiments; in particular, the Ouchi illusion, experiments with variations of this pattern, and studies on the perception of plaid motion.

Although it long has been known that the estimation of optical flow is a very difficult problem—and if formulated in the classic way, an ill-posed one—this paper for the first time points out the inherent computational problems. The insight gained from this study calls for a reevaluation of the role of flow estimation in 3D motion processing.

Appendix Expected Value of Least Squares Flow Solution

In this section we explicitly derive the second order Taylor expansion for the expected value of the least squares flow solution as a function of the variance and the covariance of the noise.

The expected value $E(\mathbf{u})$ amounts to

$$E(\mathbf{u}) = E\left(-\hat{\mathbf{I}}_s^T\hat{\mathbf{I}}_s\right)^{-1}\left(\hat{\mathbf{I}}_s^T\hat{\mathbf{I}}_t\right)$$

As the noise is considered independent at different points and $E(n_{x_i}) = E(n_i) = E(n_{t_i}) = E(n_{x_i}n_{y_i}) = 0$, the expansion at points $\mathbf{n} = 0$ can be written as

$$E(\mathbf{u}) \approx \mathbf{u}_0 + \sum_i \left(\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{x_i}^2}\right]_{\mathbf{n}=0} \frac{E\left(n_{x_i}^2\right)}{2} + \left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{y_i}^2}\right]_{\mathbf{n}=0} \frac{E\left(n_{y_i}^2\right)}{2} + \left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{t_i}^2}\right]_{\mathbf{n}=0} \frac{E\left(n_{t_i}^2\right)}{2} \right.$$
$$\left. + \left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{x_i}\partial n_{t_i}}\right]_{\mathbf{n}=0} E\left(n_{x_i}n_{t_i}\right) + \left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{y_i}\partial n_{t_i}}\right]_{\mathbf{n}=0} E\left(n_{y_i}n_{t_i}\right)\right)$$

For notational simplicity, we define

$$\hat{\mathbf{M}} = \hat{\mathbf{I}}_s^T\hat{\mathbf{I}}_s, \quad \hat{\mathbf{b}} = \hat{\mathbf{I}}_s^T\hat{\mathbf{I}}_t,$$
$$\mathbf{M} = \mathbf{I}_s^T\mathbf{I}_s, \quad \mathbf{b} = \mathbf{I}_s^T\mathbf{I}_t$$

and rewrite the expected flow as:

$$E(\mathbf{u}) = E\left(-\hat{\mathbf{M}}\hat{\mathbf{b}}\right)$$

To express the correlation between the spatial and temporal derivative noise we introduce the symbols ρ_{x_i} and ρ_{y_i} with

$$\rho_{x_i} = \frac{\sigma_{xt_i}}{\sigma_s\sigma_t} \quad \text{and} \quad \rho_{y_i} = \frac{\sigma_{yt_i}}{\sigma_s\sigma_t}$$

and write the noise terms in a linear approximation as

$$n_{x_i} = \rho_{x_i}n_{t_i} + K_{1_i} \qquad n_{y_i} = \rho_{y_i}n_{t_i} + K_{2_i}$$
$$\text{and } n_{t_i} = \rho_{x_i}n_{x_i} + K_{3_i} \; n_{t_i} = \rho_{y_i}n_{y_i} + K_{4_i}$$

where K_{1_i} and K_{2_i} are noise variables independent of n_{t_i}, and K_{3_i} and K_{4_i} noise variables independent of n_{x_i} and n_{y_i}, respectively, that will not be used explicitly.

To compute the partial derivatives, the explicit terms of the matrix $\hat{\mathbf{M}}$ and $\hat{\mathbf{b}}$ are:

$$\hat{\mathbf{M}} = \begin{bmatrix} \sum_i (I_{x_i} + n_{x_i})^2 & \sum_i (I_{x_i} + n_{x_i})(I_{y_i} + n_{y_i}) \\ \sum_i (I_{x_i} + n_{x_i})(I_{y_i} + n_{y_i}) & \sum_i (I_{y_i} + n_{y_i})^2 \end{bmatrix} \quad (.1)$$

$$\hat{\mathbf{b}} = \begin{bmatrix} \sum_i (I_{x_i} + n_{x_i})(I_{t_i} + n_{t_i}) \\ \sum_i (I_{y_i} + n_{y_i})(I_{t_i} + n_{t_i}) \end{bmatrix}. \quad (.2)$$

Using the fact that for an arbitrary matrix $\mathbf{Q}$,

$$\frac{-\partial \mathbf{Q}^{-1}}{\partial x} = \mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial x}\mathbf{Q}^{-1}, \tag{.3}$$

we find the first order derivatives to be

$$\begin{aligned}\frac{\partial \hat{\mathbf{u}}}{\partial n_{x_i}} = & \underbrace{\hat{\mathbf{M}}^{-1}\begin{bmatrix} 2(I_{x_i}+n_{x_i}) & (I_{y_i}+n_{y_i}) \\ (I_{y_i}+n_{y_i}) & 0 \end{bmatrix}\hat{\mathbf{M}}^{-1}\hat{\mathbf{b}} - \hat{\mathbf{M}}\begin{bmatrix} (I_{t_i}+n_{t_i}) \\ 0 \end{bmatrix}}_{\text{variant terms}} \\ & - \rho_{x_i}\underbrace{\hat{\mathbf{M}}\begin{bmatrix} I_{x_i}+n_{x_i} \\ I_{y_i}+n_{y_i} \end{bmatrix}}_{\text{covariant terms}}\end{aligned} \tag{.4}$$

and a similar, symmetric expression for $\frac{\partial \hat{\mathbf{u}}}{\partial n_{y_i}}$

$$\begin{aligned}\frac{\partial \hat{\mathbf{u}}}{\partial n_{t_i}} = & \underbrace{-\hat{\mathbf{M}}^{-1}\begin{bmatrix} I_{x_i}+n_{x_i} \\ I_{y_i}+n_{y_i} \end{bmatrix}}_{\text{variant terms}} \\ & \underbrace{+\hat{\mathbf{M}}^{-1}\begin{bmatrix} 2\rho_{x_i}\left(I_{x_i}+n_{x_i}\right) & \rho_{x_i}\left(I_{y_i}+n_{y_i}\right)+\rho_{y_i}\left(I_{x_i}+n_{x_i}\right) \\ \rho_{x_i}\left(I_{y_i}+n_{y_i}\right)+\rho_{y_i}\left(I_{x_i}+n_{x_i}\right) & 2\rho_{y_i}\left(I_{y_i}+n_{y_i}\right) \end{bmatrix}\hat{\mathbf{M}}^{-1}\hat{\mathbf{b}}}_{\text{covariant terms}} \\ & \underbrace{-\hat{\mathbf{M}}^{-1}\begin{bmatrix} \rho_{x_i}\left(I_{t_i}+n_{t_i}\right) \\ \rho_{y_i}\left(I_{t_i}+n_{t_i}\right) \end{bmatrix}}_{\text{covariant terms}}\end{aligned}$$

The bias in the expansion is actually due only to the second order terms. As they become very long, we will write them out piecemeal.

$$
\begin{aligned}
\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{x_i}^2}\right|_{\mathbf{n}=0} = &- 2\mathbf{M}^{-1}\begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix}\mathbf{M}^{-1}\begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ \mathbf{M}^{-1}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ 2\mathbf{M}^{-1}\begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix}\mathbf{M}^{-1}\begin{bmatrix} \rho_{x_i} I_{x_i} + I_{t_i} \\ \rho_{x_i} I_{y_i} \end{bmatrix} \\
&- \mathbf{M}^{-1}\begin{bmatrix} 2\rho_{x_i} \\ 0 \end{bmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{y_i}^2}\right|_{\mathbf{n}=0} = &- 2\mathbf{M}^{-1}\begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix}\mathbf{M}^{-1}\begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ \mathbf{M}^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ 2\mathbf{M}^{-1}\begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix}\mathbf{M}^{-1}\begin{bmatrix} \rho_{y_i} I_{x_i} \\ \rho_{y_i} I_{y_i} + I_{t_i} \end{bmatrix} \\
&- \mathbf{M}^{-1}\begin{bmatrix} 0 \\ 2\rho_{y_i} \end{bmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{t_i}^2}\right|_{\mathbf{n}=0} = &- 2\mathbf{M}^{-1}\begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \\
&\mathbf{M}^{-1}\begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ \mathbf{M}^{-1}\begin{bmatrix} 2{\rho_{x_i}}^2 & 2\rho_{x_i}\rho_{y_i} \\ 2\rho_{x_i}\rho_{y_i} & 2{\rho_{y_i}}^2 \end{bmatrix}\mathbf{M}^{-1}\mathbf{b} \\
&+ 2\mathbf{M}^{-1}\begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix}\mathbf{M}^{-1}\begin{bmatrix} I_{x_i} + \rho_{x_i} I_{t_i} \\ I_{y_i} + \rho_{y_i} I_{t_i} \end{bmatrix} \\
&- \mathbf{M}^{-1}\begin{bmatrix} 2\rho_{x_i} \\ 2\rho_{y_i} \end{bmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{t_i} \partial n_{x_i}}\right|_{\mathbf{n}=0} = & -\mathbf{M}^{-1} \begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& -\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} & \rho_{y_i} \\ \rho_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} \rho_{x_i} I_{x_i} + I_{t_i} \\ \rho_{x_i} I_{y_i} \end{bmatrix} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} I_{x_i} + \rho_{x_i} I_{t_i} \\ I_{y_i} + \rho_{y_i} I_{t_i} \end{bmatrix} \\
& -\mathbf{M}^{-1} \begin{bmatrix} 1 + {\rho_{x_i}}^2 \\ \rho_{x_i}\rho_{y_i} \end{bmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial^2 \hat{\mathbf{u}}}{\partial n_{t_i} \partial n_{y_i}}\right|_{\mathbf{n}=0} = & -\mathbf{M}^{-1} \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& -\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 0 & \rho_{x_i} \\ \rho_{x_i} & 2\rho_{y_i} \end{bmatrix} \mathbf{M}^{-1}\mathbf{b} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} \rho_{y_i} I_{x_i} \\ \rho_{y_i} I_{y_i} + I_{t_i} \end{bmatrix} \\
& +\mathbf{M}^{-1} \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} I_{x_i} + \rho_{x_i} I_{t_i} \\ I_{y_i} + \rho_{y_i} I_{t_i} \end{bmatrix} \\
& -\mathbf{M}^{-1} \begin{bmatrix} \rho_{x_i}\rho_{y_i} \\ 1 + {\rho_{y_i}}^2 \end{bmatrix}.
\end{aligned}
$$

If we make the assumption of symmetry in the error distribution of the measurements of the x and y derivatives of image intensity, the expansion can be simplified to:

$$
\begin{aligned}
\hat{\mathbf{u}} = \mathbf{u}_0 + \sum_i \Bigg\{ & +\mathbf{M}^{-1} \left(\begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} + \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \right) \mathbf{u}_0 \\
& - \underline{\underline{\mathbf{M}^{-1}\mathbf{u}_0}} + \mathbf{M}^{-1} \left(\begin{bmatrix} 2I_{x_i} & I_{y_i} \\ I_{y_i} & 0 \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} I_{t_i} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & I_{x_i} \\ I_{x_i} & 2I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 0 \\ I_{t_i} \end{bmatrix} \right) \\
& + \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} I_{x_i} \\ I_{y_i} \end{bmatrix} - \underline{\underline{\mathbf{M}^{-1} \begin{bmatrix} \rho_{x_i} \\ \rho_{y_i} \end{bmatrix}}} \Bigg\} {\sigma_s}^2 \\
+ \sum_i \Bigg\{ & \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{u}_0 \\
& - \underline{\underline{\mathbf{M}^{-1} \begin{bmatrix} {\rho_{x_i}}^2 & \rho_{x_i}\rho_{y_i} \\ \rho_{x_i}\rho_{y_i} & {\rho_{y_i}}^2 \end{bmatrix} \mathbf{u}_0}} \\
& + \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} I_{x_i} + \rho_{x_i} I_{t_i} \\ I_{y_i} + \rho_{y_i} I_{t_i} \end{bmatrix} - \underline{\underline{\mathbf{M}^{-1} \begin{bmatrix} \rho_{x_i} \\ \rho_{y_i} \end{bmatrix}}} \Bigg\} \sigma_t^2 \\
+ \sum_i \Bigg\{ & 2\mathbf{M}^{-1} \begin{bmatrix} 2I_{x_i}\rho_{x_i} & I_{x_i}\rho_{x_i} + I_{y_i}\rho_{x_i} \\ I_{x_i}\rho_{y_i} + I_{y_i}\rho_{y_i} & 2I_{y_i}\rho_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{u}_0 \\
& - \underline{\underline{2\mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} & \rho_{x_i} + \rho_{y_i} \\ \rho_{x_i} + \rho_{y_i} & {\rho_{y_i}}^2 \end{bmatrix} \mathbf{u}_0}} \\
& + \mathbf{M}^{-1} \begin{bmatrix} 2\rho_{x_i} I_{x_i} & \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} \\ \rho_{x_i} I_{y_i} + \rho_{y_i} I_{x_i} & 2\rho_{y_i} I_{y_i} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} 2I_{t_i}\rho_{x_i} + I_{x_i}\left(1 + \rho_{x_i}^2 + \rho_{y_i}^2\right) \\ 2I_{t_i}\rho_{y_i} + I_{y_i}\left(1 + \rho_{x_i}^2 + \rho_{y_i}^2\right) \end{bmatrix} \\
& - \underline{\underline{\mathbf{M}^{-1} \begin{bmatrix} \rho_{x_i}\left(1 + {\rho_{x_i}}^2 + {\rho_{y_i}}^2\right) \\ \rho_{y_i}\left(1 + {\rho_{x_i}}^2 + {\rho_{y_i}}^2\right) \end{bmatrix}}} \Bigg\} \sigma_s \sigma_t
\end{aligned}
$$

where we mark by underlining the terms that diminish proportionally to $\frac{1}{n}$, (where n is the number of measurement being combined in a region). The sum of all n such terms will give a consistent, statistically constant response. The rest of the terms diminish proportionally to $\frac{1}{n^2}$. Informal experiments shows the sum of these terms to become negligible for $n > 5$, a number clearly smaller than the number of terms likely to be combined in any real system.

References

[Adelson and Movshon, 1982] E. H. Adelson and J. A. Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300:523–525, December 1982.

[Besl and Jain, 1988] Paul J Besl and Ramesh C Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2):167–192, 1988.

[Burke and Wenderoth, 1993] Darren Burke and Peter Wenderoth. The effect of interactions betwen one-dimensional component gratings on two-dimensional motion perception. *Vision Research*, 33(3):343–350, 1993.

[Coombs and Brown, 1993] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *International Journal of Computer Vision*, 11:147–164, 1993.

[Daniilidis and Spetsakis, 1997] K. Daniilidis and M. E. Spetsakis. Understanding noise sensitivity in structure from motion. In *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 4. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[Daniilidis, 1992] K. Daniilidis. *On the Error Sensitivity in the Recovery of Object Descriptions.* PhD thesis, Department of Informatics, University of Karlsruhe, Germany, 1992. In German.

[Daniilidis, 1996] K. Daniilidis. Fixation simplifies 3d motion estimation. *Computer Vision and Image Understanding*, 68:158–169, 1996.

[de Valois *et al.*, 1982] R. L. de Valois, E. W. Yund, and N. Hepler. The orientation and direction sensitivity of cells in macaque visual cortex. *Visual Research*, 22:531–544, 1982.

[Duffy and Wurtz, 1991a] C. J. Duffy and R. H. Wurtz. Sensitivity of MST neurons to optical flow stimuli I: A continuum of response selectivity to large field stimuli. *Journal of Neurophysiology*, 65:1329–1345, 1991.

[Duffy and Wurtz, 1991b] C. J. Duffy and R. H. Wurtz. Sensitivity of MST neurons to optical flow stimuli II: Mechanisms of response selectivity revealed by small field stimuli. *Journal of Neurophysiology*, 65:1346–1359, 1991.

[Faugeras, 1992] O. D. Faugeras. *Three-Dimensional Computer Vision.* MIT Press, Cambridge, MA, 1992.

[Fermüller and Aloimonos, 1992] C. Fermüller and Y. Aloimonos. Tracking facilitates 3-D motion estimation. *Biological Cybernetics*, 67:259–268, 1992.

[Fermüller and Aloimonos, 1995a] C. Fermüller and Y. Aloimonos. Direct perception of three-dimensional motion from patterns of visual motion. *Science*, 270:1973–1976, 1995.

[Fermüller and Aloimonos, 1995b] C. Fermüller and Y. Aloimonos. Representations for active vision. In *Proc. International Joint Conference on Artificial Intelligence*, August 1995.

[Fermüller *et al.*, 1997a] C. Fermüller, L. Cheong, and Y. Aloimonos. Visual space distortion. *Biological Cybernetics*, 77:323–337, 1997.

[Fermüller *et al.*, 1997b] C. Fermüller, R. Pless, and Y. Aloimonos. Families of stationary patterns producing illusory movement: Insights into the visual system. *Proc. Royal Society, London B*, 264:795–806, 1997.

[Ferrera and Wilson, 1987] Vincent P. Ferrera and Hugh R. Wilson. Direction specific masking and the analysis of motion in two dimenions. *Vision Research*, 27:1783–1796, 1987.

[Ferrera and Wilson, 1991] Vincent P. Ferrera and Hugh R. Wilson. Perceived speed of moving two-dimensional patterns. *Vision Research*, 31(5):877–893, 1991.

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

[Heeger, 1988] D. Heeger. Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, 1:279–302, 1988.

[Hildreth, 1984] E. Hildreth. Computations underlying the measurement of visual motion. *Artificial Intelligence*, 23:309–354, 1984.

[Hine *et al.*, 1995] T. J. Hine, M. Cook, and G. T. Rogers. An illusion of relative motion dependent upon spatial frequencies and orientation. *Vision Research*, 33(22):3093–3102, 1995.

[Hine *et al.*, 1997] T. J. Hine, M. Cook, and G. T. Rogers. The Ouchi illusion: An anomaly in the perception of rigid motion for limited spatial frequencies and angles. *Perception and Psychophysics*, 59(3):448–455, 1997.

[Horn and Schunck, 1981] B. K. P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[Horn and Schunk, 1981] B. K. P. Horn and B G Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[Horn, 1986] B. K. P. Horn. *Robot Vision*. McGraw Hill, New York, 1986.

[Hubel and Wiesel, 1968] D. Hubel and T. Wiesel. Receptive fields and functional architecture of the monkey striate cortex. *Journal of Physiology*, 195:215–243, 1968.

[Jasinschi *et al.*, 1992] R. S. Jasinschi, A. Rosenfeld, and K. Sumi. Perceptual motion transparency: The role of geometrical information. *Journal of the Optical Society of America A*, 9:1865–1879, 1992.

[Kelly, 1979] D H Kelly. Motion and vision. II. stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 609(10):1340–1349, October 1979.

[Khang and Essock, 1997a] B.-G. Khang and E. A. Essock. Apparent relative motion from a checkerboard surround. *Perception*, 26(7):831–846, 1997.

[Khang and Essock, 1997b] B.-G. Khang and E. A. Essock. A motion illusion from two-dimensional periodic patterns. *Perception*, 26(5):585–597, 1997.

[Kim and Wilson, 1993] J. Kim and H. R. Wilson. Dependence of plaid motion coherence on component grating directions. *Vision Research*, 33(17):2479–2489, 1993.

[Koller *et al.*, 1993] D. Koller, K. Daniilidis, and H.-H. Nagel. Model-based object tracking in monocular image sequences of road-traffic scenes. *International Journal of Computer Vision*, 10:257–281, 1993.

[Kooi *et al.*, 1992] Frank L Kooi, Karen K De Valois, David H Grosof, and Russell L De Valois. Properties of the recombination of one-dimensional motion signals into a pattern motion signal. *Perception and Psychophysics*, 52(4):415–424, 1992.

[Marr and Ullman, 1981] D. Marr and S Ullman. Directional selectivity and its use in early visual processing. *Proc. Royal Society, London B*, 211:151–180, 1981.

[Maunsell and Essen, 1983] J. H. R. Maunsell and D. C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey I. Selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology*, 49:1127–1147, 1983.

[Movshon *et al.*, 1986] J. A. Movshon, E. H. Adelson, Gizzi, and W. T. Newsome. The analysis of moving visual patterns. *Experimental Research*, 11(supplemental):117–151, 1986.

[Movshon, 1990] A. Movshon. Visual processing of moving images. In H. Barlow, C. Blakemore, and M. Weston-Smith, editors, *Images and Understanding*, pages 122–137. Cambridge University Press, 1990.

[Nagel and Haag, 1998] H.-H. Nagel and M. Haag. Bias-corrected optical flow estimation for road vehicle tracking. In *Proc. International Conference on Computer Vision*, pages 1006–1011, January 1998.

[Nagel, 1995] H.-H. Nagel. Optical flow estimation and the interaction between measurement errors at adjacent pixel positions. *International Journal of Computer Vision*, 15:271–288, 1995.

[Nakayama and Silverman, 1988a] 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.

[Nakayama and Silverman, 1988b] K. Nakayama and G. H. Silverman. The aperture problem - II. Spatial integration of velocity information along contours. *Vision Research*, 28:747–753, 1988.

[Nelson, 1991] R. C. Nelson. Qualitative detection of motion by a moving observer. *International Journal of Computer Vision*, 7:33–46, 1991.

[Orban, 1992] G. A. Orban. The analysis of motion signals and the nature of processing in the primate visual system. In G. A. Orban and H.-H. Nagel, editors, *Artificial and Biological Vision Systems*, ESPRIT Basic Research Series, pages 24–57. Springer-Verlag, 1992.

[Ouchi, 1977] H. Ouchi. *Japanese and Geometrical Art.* Dover, 1977.

[Rodman and Albright, 1989] H. R. Rodman and T. D. Albright. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, 75:53–64, 1989.

[Shulman and Hervé, 1989] D. Shulman and J.-Y. Hervé. Regularization of discontinuous flow fields. In *Proc. IEEE Workshop on Visual Motion*, pages 81–86, 1989.

[Simoncelli *et al.*, 1991] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 310–315, 1991.

[Smith and Brady, 1995] S. M. Smith and J. M. Brady. ASSET-2: Real time motion segmentation and shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 1995.

[Smith and Edgar, 1991] A. T. Smith and G. K. Edgar. Perceived speed and direction of complex gratings and plaids. *Journal of the Optical Society of America*, 8(7):1161–1171, July 1991.

[Spillmann *et al.*, 1993] L Spillmann, U Tulunay-Keesey, and J Olson. Apparent floating motion in normal and stabilized vision. *Investigative Ophthalmalogy and Visual Science, Supplement*, 34:1031, 1993.

[Stark, 1968] L. Stark. Neurobiological control systems. In *Studies in Bioengineering.* Plenum Press, New York, 1968.

[Stoner *et al.*, 1990] G R Stoner, T D Albright, and V S Ramachandran. Transparency and coherence in human motion perception. *Nature*, 344:153–155, March 1990.

[Szeliski, 1990] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–301, 1990.

[Ungerleider and DeSimone, 1986] L. G. Ungerleider and R. DeSimone. Cortical connections of visual area MT in the macaque. *The Journal of Comparative Neurology*, 248:190–222, 1986.

[Wallach, 1935] H Wallach. Uber visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung*, 20:325–380, 1935.

[Wang *et al.*, 1992] S. Wang, Y. Markandey, and A. Reid. Total least squares fitting spatiotemporal derivatives to smooth optical flow fields. In *Proc. of the SPIE: Signal and Data Processing of Small Targets*, volume 1698, pages 42–55. SPIE, 1992.

[Weber and Malik, 1995] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. *International Journal of Computer Vision*, 14:67–81, 1995.

[Welch, 1989] Leslie Welch. The perception of moving plaids reveals two motion-processing stages. *Nature*, 337:735–737, Feb 1989.

[Yo and Wilson, 1992] C. Yo and H. R. Wilson. Moving 2D patterns capture the perceived direction of both lower and higher spatial frequencies. *Vision Research*, 32:1263–1270, 1992.

[Zeki *et al.*, 1993] S. M. Zeki, J. D. G. Watson, and R. S. J. Frackowiak. Going beyond the information given: The relation of illusory visual motion to brain activity. *Proc. Royal Society, London B*, 252:215–222, 1993.

[Zeki, 1974] S. M. Zeki. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, 236:546–573, 1974. London.