Statistical Biases in Optic Flow

Cornelia Fermüller, Robert Pless and Yiannis Aloimonos
Center for Automation Research, University of Maryland
College Park, MD 20742-3275

Abstract
The computation of optical flow from image derivatives is biased in regions of non uniform gradient distributions. A least-squares or total least squares approach to computing optic flow from image derivatives even in regions of consistent flow can lead to a systematic bias dependent upon the direction of the optic flow, the distribution of the gradient directions, and the distribution of the image noise. The bias a consistent underestimation of length and a directional error. Similar results hold for various methods of computing optical flow in the spatiotemporal frequency domain. The predicted bias in the optical flow is consistent with psychophysical evidence of human judgment of the velocity of moving plaids, and provides an explanation of the Ouchi illusion. Correction of the bias requires accurate estimates of the noise distribution; the failure of the human visual system to make these corrections illustrates both the difficulty of the task and the feasibility of using this distorted optic flow or undistorted normal flow in tasks requiring higher level processing.

1 Introduction
The perception of motion by visual means plays an important role for many living organisms. The first stage of many visual algorithms consists of computing optic flow – the projection of the 3D motion vectors of the scene point onto the image. Two broad classifications encompass approaches to the computation of the optic flow: methods that seek to match particular points, and methods that consider the spatio-temporal image intensity function and compute gradients or frequency responses to estimate the flow. This paper concentrates on the second method, explicitly delineating the effect of an error model on optic flow computed from gradient measurements. We will argue at the end of section 2 that similar results hold for computations in frequency space.

Gradient-based approaches assume that image intensity does not change over a small time interval. Denoting the image intensity by $E$, its spatial temporal derivatives by $E_x, E_y, E_t$, and the velocity of an image point in the $x$ and $y$ directions by $u$ and $v$, the following constraint is obtained:

$$E_x u + E_y v + E_t = 0$$

This equation, called the optical flow constraint equation [6], defines the component of flow in the direction of the spatial gradient $(E_x, E_y)$—the normal flow. The gradients at a single image point do not allow a solution for both components of the optic flow – deriving the optical flow requires combining the gradient measurements or normal flow vectors in a small region of the image. 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. Small regions of the image may contain projections from 3D scene points at different depths or undergoing different motions.

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 by parametric models which are constant, linear or quadratic in the image coordinates. At the locations of discontinuities, however, this parameterization fails, and combining image measurements across discontinuities may give very erroneous optical flow estimates. To avoid smoothing over boundaries, knowledge of where the discontinuities are seems to be necessary, but this is difficult to obtain from local image measurements. The problem has been attacked with various methods: searching for filters which conform to boundaries [13] or boundary preserving regularization [4, 10].

What is less known – and the subject of this paper – is that even within regions of constant flow, the computation of optical flow from noisy gradient measurements has a systematic bias dependent upon the gradient distribution of the image region.

The estimation and interpretation of optical flow from a statistical point of view has received attention previously in the computational literature [1, 2, 3, 11, 14, 16]. Most closely related to this paper are the studies of Nagel and Haag [7, 8], who investigate and at-
Image gradients are related to image velocity by the over-determined system of equations through a covariance matrix $\mathbf{E}$, with one remark. As the spatial and temporal derivatives at each point, but with possible dependencies between the locations, but it does not affect the analysis. Thus, the variances and covariances of the noise components are

\[
E(n_{x_1}, n_{y_1}, n_{t_1}) = (0, 0, 0)
\]

\[
E(n_{x_2}^2, n_{y_2}^2, n_{t_2}^2) = (\sigma_s^2, \sigma_s^2, \sigma_t^2)
\]

\[
E(n_{x_3}, n_{y_3}) = 0
\]

\[
E(n_{x_4}, n_{t_4}) = \sigma_{xt} = -\text{sgn}(E_{x_4}, E_{t_4}) \cdot \sigma_{st}
\]

\[
E(n_{y_4}, n_{t_4}) = \sigma_{yt} = -\text{sgn}(E_{y_4}, E_{t_4}) \cdot \sigma_{st}
\]

In the absence of errors in the spatial gradient measurements $E_i$, standard least squares methods give an unbiased estimator. The expected value $E(\mathbf{u})$, obtained from (4) 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 $\mathbf{u} = E(\mathbf{u})$ 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 covariance of the noise in the measurement matrices. According to the noise model, the first-order terms vanish, and the only non-zero terms that remain in the expansion at zero noise ($\mathbf{n} = 0$) are

\[
\mathbf{u} = \mathbf{u}_0 + \nabla \mathbf{u}_0 \mathbf{n}
\]

\[
+ \sum_i \frac{\partial^2}{\partial n_{x_i} \partial n_{y_i}} (\mathbf{M}^{-1} \mathbf{b}) \mathbf{n}_{x_i} \mathbf{n}_{y_i}
\]

\[
+ \sum_i \frac{\partial^2}{\partial n_{x_i} \partial n_{t_i}} (\mathbf{M}^{-1} \mathbf{b}) \mathbf{n}_{x_i} \mathbf{n}_{t_i}
\]

\[
+ \sum_i \frac{\partial^2}{\partial n_{y_i} \partial n_{t_i}} (\mathbf{M}^{-1} \mathbf{b}) \mathbf{n}_{y_i} \mathbf{n}_{t_i}
\]

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

Algebraic manipulation of the above derivative leads to an expression for $\mathbf{u}$ that 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 variance...
terms of the noise in the temporal and spatial measurements (which we refer to as covariant noise). The dominant factors are

\[
\hat{u} = u_0 - K_1 \left( \sum_i M^{-1} u_i \right) - \sum_i K_2^i M^{-1} \begin{bmatrix} \text{sgn}(\sigma_{xt,i}) \\ \text{sgn}(\sigma_{yt,i}) \end{bmatrix}
\]

(5)

with \( M = E_t^i E_s \), computed from the real spatial derivatives uncorrupted by noise, and constants \( K_1 = \sigma_t^2 \) and

\[
K_{2,i} = \left[ \sigma_s^2 + \sigma_t^2 + \sigma_s \sigma_t + 2 \frac{\sigma_{st}^2}{\sigma_s \sigma_t} \right] + \left( \frac{\sigma_{xt,i}}{\sigma_s \sigma_t} u + \frac{\sigma_{yt,i}}{\sigma_s \sigma_t} v \right) (\sigma_t^2 + 2 \sigma_s \sigma_t) \cdot \frac{\sigma_{st}}{\sigma_s \sigma_t}
\]

Both constants are independent of the gradient distribution, so the effect of the gradient distribution on the bias of the computed flow can be interpreted through its effect on the matrix \( M^{-1} \). In the case of a uniform distribution of image gradients in the region where flow is computed, \( M \) (and therefore \( M^{-1} \)) is a multiple 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 underestimation of the length. In a region where there is a unique gradient vector, \( M \) will be of rank 1; this is the aperture problem. In the general case, the bias can be understood by analyzing the eigenvectors of \( M \). As \( M \) 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 the gradient measurements. \( M^{-1} \) has the same eigenvectors as \( M \) and inverse eigenvalues. Thus, the eigenvector corresponding to the larger eigenvalue of \( M^{-1} \) has a direction dominated by the normal to the major orientation of the image gradients, and the product of \( M^{-1} \) with any vector is most strongly influenced by this orientation. This affects the variant term, leading to underestimation of the magnitude of the optical flow and a bias in its direction toward the major direction of the 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 the gradients, depending on the gradient distribution.

Explicit analysis of a pattern with a simple gradient distribution demonstrates the bias. Figure 1 shows a variant of a pattern created by Hajime Ouchi [9]; a pattern with inset and background regions containing different and non-uniform gradient distributions. The “bricks” used to make up the figure are 4 times longer than they are wide, leading to a gradient distribution in a small region with four times as many normal flow measurements in one direction as the other. For such a gradient distribution the bias can be understood rather easily. The eigenvectors of \( M \) are in the directions of the two gradient measurements, with the larger eigenvalue corresponding to the larger number of gradients. Figure 2 shows the relationship between the optic flow bias and the angle between the optic flow and the dominant gradient direction. These plots are based upon the exact second-order Taylor expansion.

As \( u_0 = (0,1) \), the variant term in (5) leads to the bias in length shown by the curve in figure 2b, which has its minimum at 0 and maximum at \( \pi / 2 \) (that is, when \( u_0 \) is aligned with the major gradient direction). The error in angle is greatest for \( \pi / 4 \) (that is, when \( u_0 \) is exactly between the two eigenvectors of \( M^{-1} \)) and it is zero for 0 and \( \pi / 2 \) (Figure 2c). 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. The bias is always negative in length and toward the major gradient direction.

The covariant term is constant for \( \theta \in [0, \pi / 2] \), and is a bias that is negative in length and usually towards the direction with fewer gradients. Figure 2(d,e) combine both bias terms to show the expected length and direction of the optic flow computation for different \( \theta \).

This provides an explanation for the effect seen in the Ouchi Illusion, the different perceived motion in the inset and surround regions. Because the central region and the surrounding region have different gradient dist-
of this plaid pattern can be interpreted as independent orientation, or coherent, motion of the entire pattern.

The optical flow is along the positive y axis and of length 1. (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 length of computed optic flow and expected angular error. Derivative measurements are corrupted by noise with distributions: $\sigma_s = \sigma_t = 0.15$ and $\sigma_{st} = 0.1 \cdot \sigma_s^2$.

Figure 2: (a) 16 measurements are in the direction making angle $\theta$ with 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 1. (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 length of computed optic flow and expected angular error. Derivative measurements are corrupted by noise with distributions: $\sigma_s = \sigma_t = 0.15$ and $\sigma_{st} = 0.1 \cdot \sigma_s^2$.

Surprisingly, humans can perceive a coherent pattern motion which does not correspond to the veridical pattern motion – for plaid patterns with gratings of different frequencies, the perceived coherent pattern motion is biased towards the direction perpendicular to the orientation of the grating with higher frequency [12]. This effect increases with frequency difference, as predicted in our model since a higher frequency pattern gives a larger set of easily measurable image gradients. If the difference in component frequencies is too large, the relative motion directions too different, or the pattern contrast too low, the components are perceived to move independently — these conditions serve to maximize the flow bias. The difference in estimates from very biased gradient based flow techniques and blob tracking techniques may be used by higher level processes in the visual cortex to segment the scene motion. Such an analysis predicts human coherence judgements for plaid patterns and the “false segmentation” perceived in the reduced Ouchi stimulus (introduced in [5]).

This analysis was founded upon the assumption that the intensity function, $E$, is constant for corresponding image points in successive images: $\frac{de}{dt} = 0$. To have a differentiable function, typically $E$ is convolved with a linear operator, $G$, giving $G(x, y, t) = \int \int G(a, b, c)E(x + a, y + b, t + c)da db dc$, if we assume that the linear operator is independent of position. All derivative based image methods either implicitly or explicitly do such pre-processing. To compare frequency and image gradient based techniques for this prob-
lem, we consider the Fourier transform of $\frac{df}{dt}$; which is convolution by a set of exponential filters. By Parseval's Theorem, the quadratic norm is preserved under Fourier Transforms: $\|f\|^2 = \int f(x, y, t)^2 dx dy dt$ and $F$ is an operator computing the Fourier transform of a function, then $\|F f\|^2 = \|f\|^2$. Thus, a least squares formulation in image space is equivalent to a least squares formulation in frequency space.

3 Correcting the Bias?

In the statistics literature the model we used to describe the estimation of flow is referred to as the classical “Errors-In-Variable” (EIV) model. It is usually expressed in the notation $Ax = b$ with $A = A_0 + \delta A$ and $b = b_0 + \delta b$ where $A_0$ and $b_0$ are the true but unobservable variables (in our case the actual spatial and temporal derivatives $E_{x_1}, E_{y_1}, E_t$ at points $t$), $\delta A$ and $\delta b$ are the measurement errors, $A$ and $b$ are the corresponding observable variables, and $x$ represents the unknown parameters to be estimated (in our case $u$ and $v$).

It is well known from the literature that estimation by least squares (LS) generally provides an inconsistent and biased estimate of the true parameter $x$. The LS estimator gives an unbiased solution only for the regression model, that is, when $\delta A$ is considered to be zero and the measurements $\delta b$ are independent, zero mean and equally 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, the theoretical possibilities of correcting the bias is computationally difficult for realistic visual systems.

Any statistical method of compensating 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 this were available, the bias in the least squares estimate 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 must be employed.

However, the major problem lies in the acquisition of the statistics of the noise. The noise parameters are not intrinsic to the system, but depend on the viewing situation and the scene in view; in general these statistics can only be considered to be patch-wise constant. The noise parameters have to be estimated from the flow estimates within a spatiotemporal neighborhood by using the model which relates the image derivatives and noise to the flow estimates. With the limited amount of data within one patch where it is reasonable to consider the statistics constant, it is very difficult to obtain good estimates. Furthermore, the variance in the motion estimates turns out to be large with respect to the bias. For example, in simulations (see Figure 4), 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, would in many cases lead to a worsening of the solution. 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 temporal integration of measurements very difficult as the system has only a short time to obtain the noise parameters.

In recent years the nonlinear “Total Least Squares” estimator has received a lot of attention and has also been applied to the problem of flow estimation [15, 16]. 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, it is again necessary to obtain its covariance matrix. Without whitening, total least squares also gives biased solutions. In particular, if the noise in the spatial derivatives is greater than the temporal derivative noise the bias has the same form as the least squares estimation discussed here. In addition, the variance in the total least squares solution is much larger than in

\[ \text{Figure 4: Expected error in length (solid lines) and standard deviation (dotted lines) obtained by a Monte Carlo simulation using Gaussian noise with three different standard deviations: } \sigma_s = \sigma_t = 0.2, 0.1 \text{ and } 0.05. \]

The optical flow is $(0, 1)$, the magnitude of the spatial gradients is 1, and the gradients are distributed with 15 vectors at angle $\theta$ from the $x$ axis and 5 vectors at angle $\pi/2 + \theta$. 

1063-6919/99 $10.00$ (c) 1999 IEEE
ordinary least squares. Total least squares also performs very poorly if outliers are present, and these are difficult to detect from a few measurements.

4 Conclusion and Implications

The preceding analysis shows the difficulty of obtaining accurate optical flow estimates from local image measurements. This analysis considers the effect of noise when using least squares to compute the optic flow from image derivative measurements. It shows that in addition to the problems present at flow discontinuities, there are significant biases present in the estimation of flow in regions of constant or smoothly varying flow. This bias is dependent on the local image gradient distribution and affects both the direction and the magnitude of the computed optic flow. The bias cannot be corrected without accurate knowledge of the statistics of the noise distribution, information which is difficult to obtain from local image measurements of dynamic scenes. For reasonable estimates of the noise, the model presented explains a set of human perceptual errors and illusions.

If one chooses as a goal to find the best possible estimate of optic flow, there are several possibilities. Using data from larger image regions allows more accurate statistical noise sampling, but requires knowledge of the location of flow discontinuities. An iterative or feedback process can approximate the image velocity initially to find bounds on flow values or qualitative descriptions of local flow fields. Using this to create a partial three-dimensional shape model permits estimation of flow boundaries allowing subsequent flow estimation to use data from larger image regions and further improve the knowledge of the scene structure. This simultaneous estimation of the scene structure and motion holds the best promise for accurate measurements.

However, even with the best computations, it cannot be guaranteed that optical flow will always be accurately estimated – this must be taken into account when performing visual tasks. Many tasks do not require scene reconstructions or dense optical flow fields. Normal flow measurements, or optic flow bounds which can be accurately defined can allow the generation of less powerful shape representations sufficient for many tasks – representations describing the qualitative shape of scene patches or depth ordering of scene elements. Robust, qualitative descriptions may be best able to avoid statistical biases from measurement noise.

References


