Velocity likelihoods in biological and machine vision

Yair Weiss
Computer Science Division
485 Soda Hall
UC Berkeley
Berkeley, CA 94720-1776
yweiss@cs.berkeley.edu

David J. Fleet
Xerox Palo Alto Research Center
Palo Alto, CA 94304
fleet@parc.xerox.com

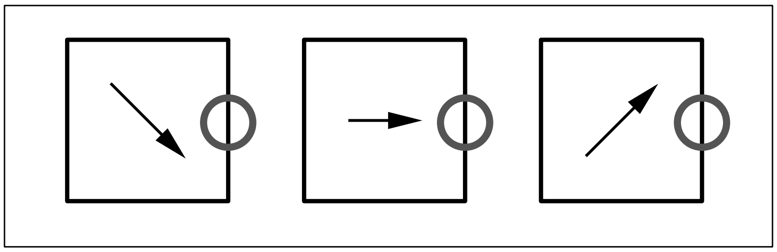
Abstract

Recent approaches to estimating two-dimensional image motion and to modeling the perception of image motion have achieved success with Bayesian formulations. With a Bayesian approach, the goal is to compute the posterior probability distribution of velocity, which is proportional to a likelihood function and a prior. The likelihood function describes the probability of observing the image data given the image velocity; surprisingly, there is still disagreement about the right likelihood function to use.

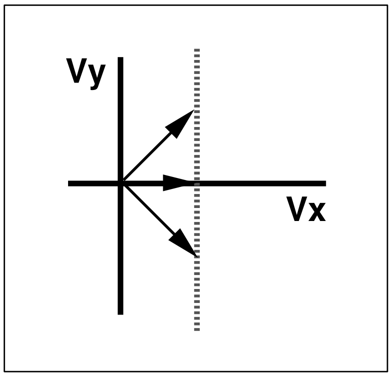
Here we derive a likelihood function by starting from a generative model. We assume that the scene translates, conserving image brightness, while the image is equal to the projected scene plus noise. We discuss the connection between the resulting likelihood function and existing models of motion analysis. We show that the likelihood can be calculated by a population of units whose response properties are similar to "motion energy" units. This suggests that a population of velocity tuned cells in visual cortex may be representing the likelihood function.

1 Introduction

What computations occur in early motion analysis in biological and machine vision? One common hypothesis is that visual motion analysis procedes by first computing local 2d velocities, and then by combining these local estimates to compute the global motion of an object. A well-known problem with this approach is that local motion information is often ambiguous, a situation often referred to as the "aperture problem" [13, 6, 2, 8]. Consider the scene depicted in Figure 1. A local analyzer that sees only the vertical edge of a square can only determine the horizontal component of the motion. Whether the square translates horizontally to the right, diagonally up and to the right, or diagonally down and to the right, the motion of the vertical



a

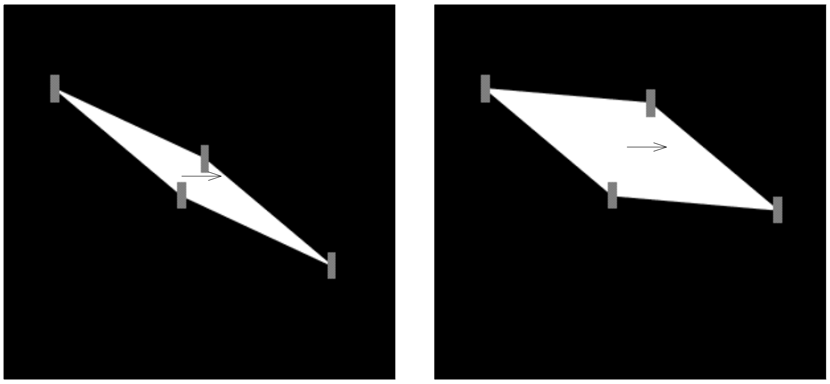


b

Figure 1: **a.** The "aperture problem" refers to the inability to determine the two dimensional motion of a signal containing a single orientation. For example, a local analyzer that sees only the vertical edge of a square can only determine the horizontal component of the motion. Whether the square translates horizontally to the right, diagonally up and to the right, or diagonally down and to the right, the motion of the vertical edge will be the same. **b.** The family of motions consistent with the motion of the edge can be depicted as a line in "velocity space", where any velocity is represented as a vector from the origin whose length is proportional to speed and whose angle corresponds to direction of motion. Graphically, the aperture problem is equivalent to saying that the family of motions consistent with the edge maps to a straight line in velocity space, rather than a single point.

edge will appear the same within the aperture. The family of velocities consistent with the motion of the edge can be depicted as a line in "velocity space", where any velocity is represented as a vector from the origin whose length is proportional to speed and whose angle corresponds to the direction of motion. Geometrically, the aperture problem is equivalent to saying that the family of motions consistent with the information at an edge maps to a straight line in velocity space, rather than a single point.

Because of ambiguities due to the aperture problem, as well as noise in the image observations, it would make sense that a system would represent the *uncertainty* in the local estimate as well as the best estimate. This would enable subsequent processing to combine the local estimates while taking their uncertainties into account. Accordingly, it has been argued that the goal of early motion analysis should be the extraction of local *likelihoods* (probability distributions over velocity), rather than a single estimate [10]. In a Bayesian approach to motion analysis, these local like-



a b

Figure 2: **a.** A "narrow" rhombus whose endpoints are occluded appears to move diagonally (consistent with VA). **b.** A "fat" rhombus whose endpoints are occluded appears to move horizontally

lihoods would be combined with the observer's prior assumptions about the world, to estimate the motion of objects.

In this paper, we assume that early motion analysis does indeed extract velocity likelihoods, and we address a number of questions raised by this assumption:

- What is the form of the likelihood? Can it be derived from first principles?
- What is the relationship between the local likelihood calculation and other models of early motion analysis?
- Can these likelihoods be represented by known physiology?

2 Motivation - Motion analysis as Bayesian inference

In the Bayesian approach to motion analysis, the goal is to calculate the posterior probability of a velocity given the image data. This posterior is related to the likelihoods and prior probabilities by Bayes' rule. Denoting by $I(x,\, t)$ the spatiotemporal brightness observation (measurement) at location x and time t, and by v the 2d image motion of the object, then

$$P(v \,|\, I(x,t))) \;=\; \alpha\, P(v)\, P(I(x,t) \,|\, v) \;, \tag{1}$$

where α is a normalization constant that is independent of v.

Bayes' rule represents a normative prescription for combining uncertain information. Assuming that the image observations at different positions and times are conditionally indpendent, given v, it is straightforward to show that this simplifies into:

$$P(v \,|\, I(x,t)) \;=\; \alpha\, P(v) \prod_{i,j} P(I(x_i,t_j) \,|\, v)) \;, \tag{2}$$

where the product is taken over all positions x_i and times t_j. The important quantity to calculate at every image location is the *likelihood* of a velocity, $P(I(x_i,t_j) \,|\, v)$.

Interestingly, there is growing evidence that the human visual system can be described in terms of computations like these. For example, in [15, 5] it was shown that a large number of visual illusions are explained by a model that maximizes Equation 2 when $P(v)$ is taken to be a prior distribution that favors slow speeds. Figure 2 shows an example from [15]. Each stimulus consisted of a translating rhombus whose endpoints are occluded. When the rhombus is "fat", it is indeed perceived as moving horizontally. But when the rhombus is "narrow" the percevied motion is illusory — subjects perceive it as moving diagonally rather than horizontally.

Why do humans misperceive the motion of a narrow rhombus but not a fat one? To address this question, let us first consider models that do not represent uncertainty about local motion measurements. In the case of the fat rhombus, the perceived motion can be explained by an intersection-of-constraints (IOC) model [2]. According to this model, each local analyzer extracts the constraint line corresponding to the local moving contour. Subsequent processing then finds the intersection of these two constraint lines. This procedure will always give the veridical motion for a translating 2D figure, so it can explain the motion of the fat rhombus.

But, this IOC model does not account for the motion percept with the narrow one. As an alternative model, if each local analyzer were to extract the normal velocity of the contour followed by a *vector average* of these normal velocities, this would predict the diagonal motion for the narrow rhombus [16]. But, this vector average model does not explain the percept of the fat rhombus.

Figures 3–4 show how both percepts can be accounted for by Equation 2. Here, each local analyzer extracts a likelihood from the local contour motion. As shown in the figures these likelihoods are fuzzy constraint lines, indicating that velocities on the constraint lines have highest likelihoods, and the likelihood decreases gradually with increasing distance from the constraint line. When these likelihoods are multiplied together with the prior, as dictated by Equation 2, the predicted motion is horizontal for fat rhombuses and diagonal for narrow rhombuses.

These results and others in [15] suggest that a Bayesian model with a prior favoring slow speeds can explain a range of percepts in human vision. But our original question concerning the right likelihood function remains.

3 What is the likelihood function for image velocity?

3.1 Previous approaches

In order to compute image velocity, one must first decide which property of the image to track from one time to the next. One common, successful approach in machine vision is based on the assumption that the light reflected from a object surface remains constant through time, in which case one can track points of constant image intensity (e.g., [3, 6, 7]). Mathematically, this can be expressed in terms of a path, $x(t)$, along which the image, $I(x(t), t)$, remains constant: i.e.,

$$I(x(t), t) = C , \tag{3}$$

where C is a constant. Taking the temporal derivative of both sides of Equation 3, and assuming that the path $x(t)$ is sufficiently smooth to be differentiable, with

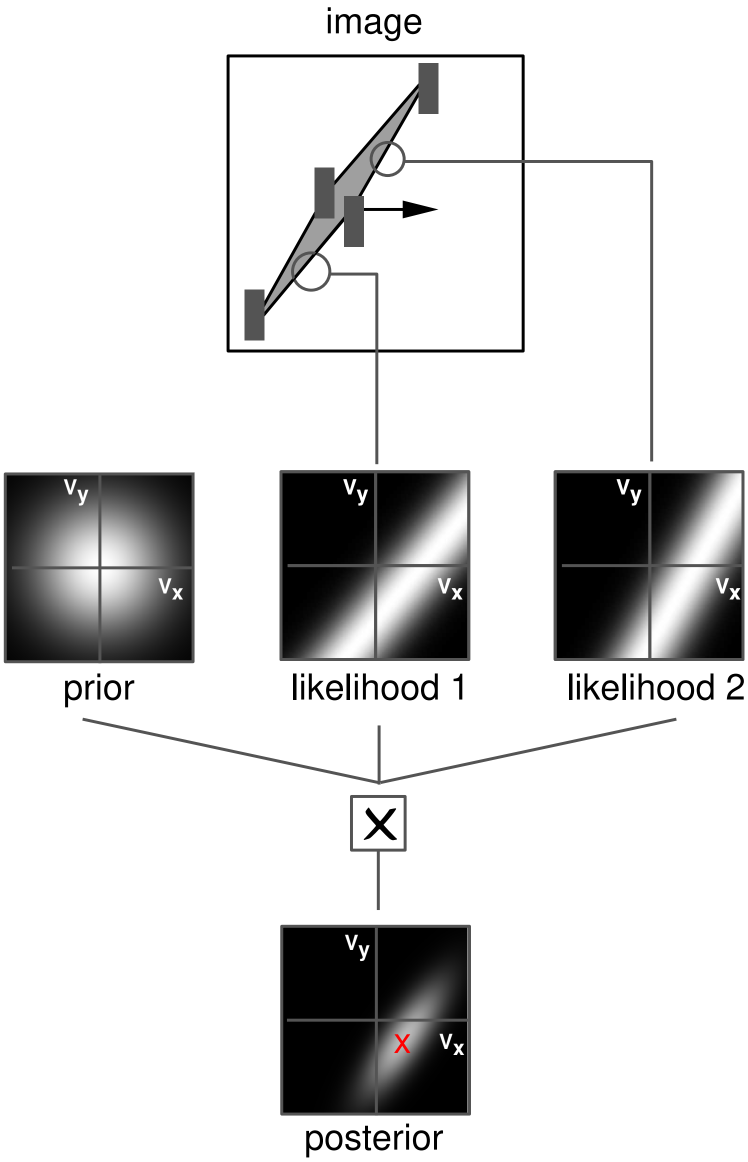


Figure 3: The response of the Bayesian estimator to a narrow rhombus. (replotted from Weiss and Adelson 98)

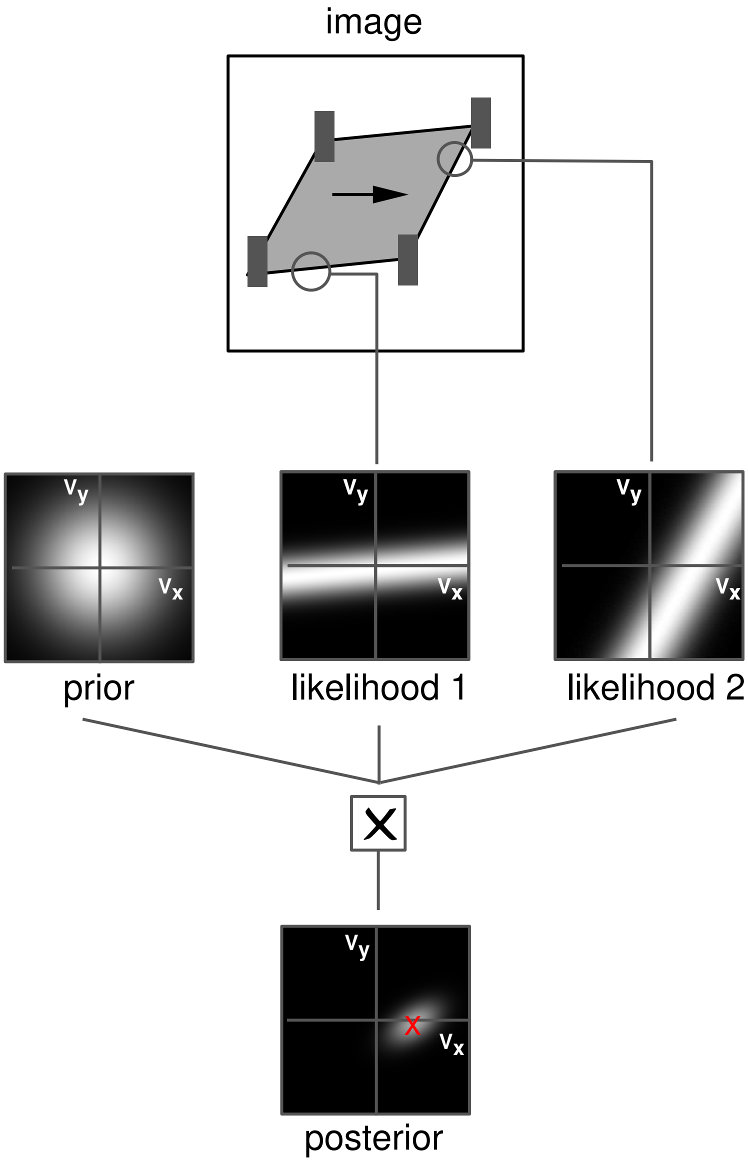


Figure 4: The response of the Bayesian estimator to a fat rhombus. (replotted from Weiss and Adelson 98)

$v \equiv (v_x, v_y) = (\frac{dx}{dt}, \frac{dy}{dt})$, provides us with the constraint

$$\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial x} v_y + \frac{\partial I}{\partial t} = 0 \,. \tag{4}$$

This is often refered to as the gradient constraint equation. When the exact solutions to Equation 4 are plotted in velocity space, one obtains a constraint line. This line represents all of the different 2d velocities that are consistent with the image derivative measurements, as given by Equation 4.

In estimating image velocity, it is the likelihood function that expresses our belief that certain velocities are consisten with the image measurements. Uncertainty in belief arises because the derivative measurements in Equation 4 only constrain velocity to somewhere along a line. Further uncertainty in belief arises because the partial derivative measurements are noisy. According to reasoning of this sort, most likelihood functions that have been proposed fall into one of two categories: "fuzzy constraint lines" (as in Figure 5c) and "fuzzy bowties" (as in Figure 5d). Examples of the two categories appeared in Simoncelli (93)[10].

The first one defines the likelihood to be

$$P(I \,|\, v) = \alpha \exp\left(-\frac{1}{2\sigma^2} \int \left(\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} \right)^2 dx dt \right) \,. \tag{5}$$

This likelihood function is often derived by assuming that the temporal derivative measurement is contaminated with mean-zero Gaussian noise, but the spatial derivative measurements are noise-free [11]. Figure 5c shows an example of the likelihood for the image sequence shown in Figure 5a. For this image, that contains only a single orientation, this looks like a fuzzy constraint line.

The second category of likelihood function is defined to be:

$$P(I \,|\, v) = \alpha \exp\left(-\frac{1}{2\sigma^2} \int \frac{(\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t})^2}{1 + v_x^2 + v_y^2} dx dt \right) \,. \tag{6}$$

This likelihood function has been shown to result from an assumption that mean-zero Gaussian noise is added to each of the spatial and temporal derivative measurements [9]. While this likelihood function has only recently been derived, the velocity at which it is maximal corresponds to what has been usually called the total-least-squares velocity estimate [14]. Figure 5d shows the picture in velocity space. For a sequence that contains only a single orientation, this looks like a fuzzy bowtie. Given the assumption of noise in both spatial and temporal derivatices, the fuzzy bowtie seems slightly more attractive — why should one direction of differentiation behave differently than another?

The fuzzy constraint line has other desirable qualities however. One nice property can be illustrated in Figure 5. Obviously a vertical edge moving with velocity v is indistinguishable from a vertical edge moving with velocity $(v_x, v_y) + \alpha(0,1)^T$. Thus if our image sequence contains only vertical edges, we might like the likelihood function to be invariant to an addition of a vertical component $P(I \,|\, (v_x, v_y)) = P(I \,|\, (v_x, v_y) + \alpha(0,1)^T)$. This means that curve of equal likelihood should be lines that are parallel to the constraint line, a property that fuzzy lines have but fuzzy bowties do not.

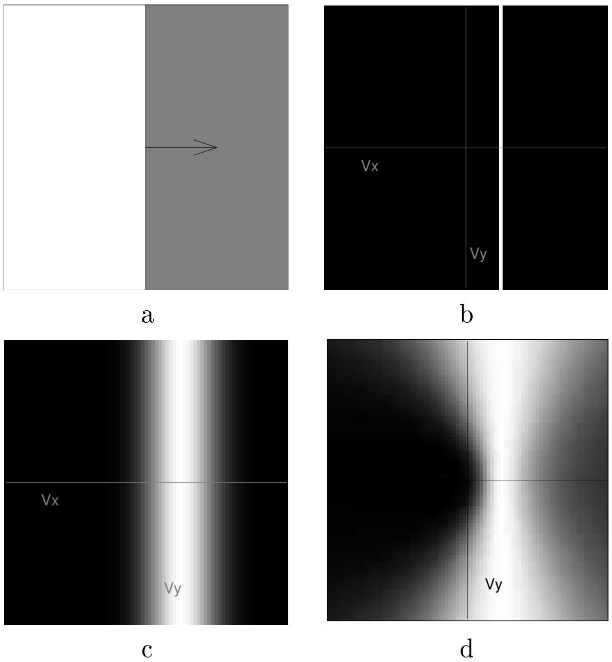


Figure 5: **a:** A moving edge. What is the likelihood of a velocity given this image sequence? **b.** The constraint line in velocity space. In the absence of noise, all velocities along the constraint line are consistent with the data. **c-d.** Likelihood functions in velocity space. White pixels correspond to high likelihood. Assuming only the temporal derivatives are noisy gives a fuzzy constraint line (as in b) but assuming all derivatives are equally noisy gives a fuzzy bowtie (as in c). What is the right likelihood to use?

Surprisingly, after many years of research into local motion analysis, there remains a lack of concensus regarding which likelihood to use, as these two and others have been suggested. To illustrate this, consider the recent paper of Fernmuller et al. [4] who have suggested yet another local likelihood function. It assumes that the noise in the spatial and temporal derivatives may be correlated. Specifically the noise in the two spatial derivatives is uncorrelated but the noise in the spatial and temporal derivatives is correlated with a correlation coefficient that depends on the sign of the derivatives, $E(I_x I_t) = -\sigma_{xt} sgn(I_x I_t)$. For $\sigma_{xt} = 0$ this reduces to the total-least-squares likelihood or the fuzzy bowtie. But when σ_{xt} is nonzero, they find that the likelihood is biased. That is, even if the noise is zero, the ML estimator using their likelihood function does not give the true velocity.

The problem with deriving these likelihoods from Equation 4 is that there is no generative model. The preceding discussion tries to derive noise models in the derivative domain rather than basing the noise assumptions in the imaging domain (where presumably we have better intuitions about what constitutes a reasonable noise model).

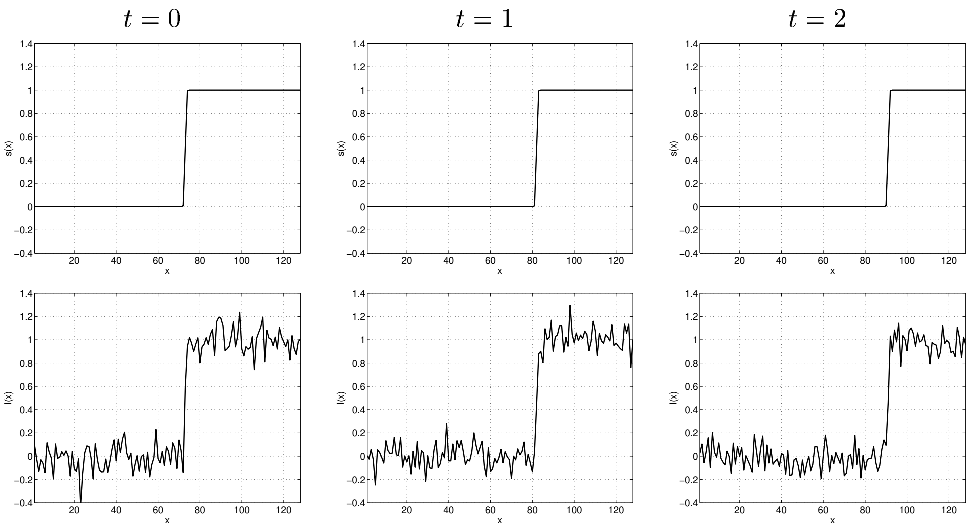


Figure 6: The generative model that we use to derive the likelihood function. The signal function (top panels) translates and conserves brightness. The image (bottom panels) equals signal plus imaging noise.

3.2 Generative model

In what follows we derive a likelihood function from a generative model of images. It is a natural extension of intensity conservation to a noisy imaging situation (see Figure 6). For notational simplicity we consider the generation of 1d images. The extension to 2d images is straightforward.

Let us assume that an unknown scene function $s(x)$ is first generated with probability $P(s)$. It is then translated with velocity v:

$$S(x,t) \;=\; s(x - vt)\ . \tag{7}$$

In what follows we use capital $S(x,t)$ to denote the ideal, noiseless image sequence and $s(x) = S(x,0)$ to denote a single image from that sequence.

Finally, to model the process of image formation, we assume that the observed image is equal to the translating scene plus imaging noise:

$$I(x,t) = S(x,t) + \sigma\eta \tag{8}$$

where η denotes zero mean Gaussian noise with variance 1 that is independent across time and space, and independent of S. We assume that we observe $I(x,t)$ for a fixed time interval $|t| < t_m$ and for all x. Also, we will use the notation $\|f\|^2$ to denote the energy in the signal f; that is,

$$\|f\|^2 \;=\; \int_{|t|<t_m,x} f^2(x,t)\,dxdt\ . \tag{9}$$

Claim 1: Assuming a uniform prior over scene functions ($P(s)$ is independent of s) then

$$P(I\,|\,v) \;=\; \alpha \exp\left(-\frac{1}{2\sigma^2}\|I - \hat{S}_v\|^2\ ,\right) \tag{10}$$

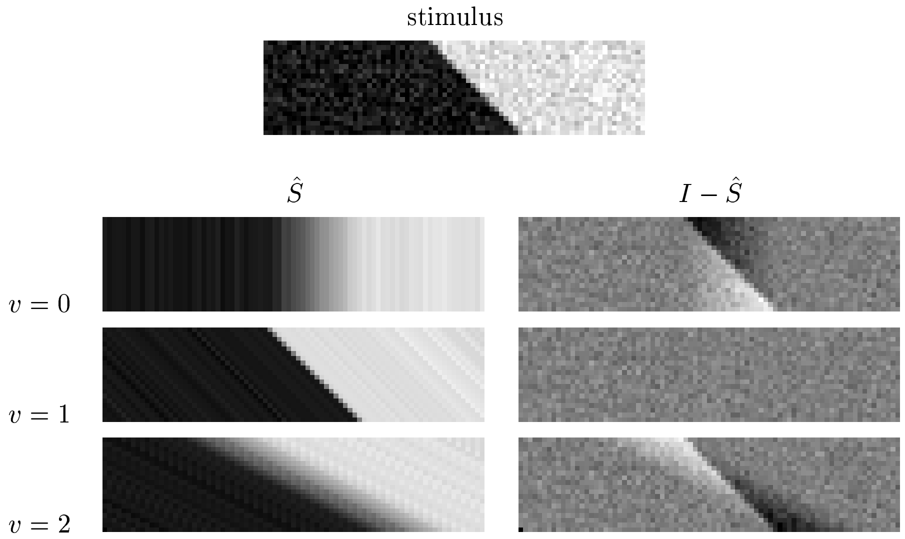


Figure 7: **Top:** A space versus time (xt) plot of an edge moving to the right. **Bottom:** Calculations of the log likelihood for different velocities. For each velocity we calculate the predicted intensity assuming a scene function moving at that velocity (shown in the left column). The residual intensity (shown in the right column) is explained as noise: the less energy in the residual the more likely the velocity.

with

$$\hat{S}_v(x,t) \;=\; \hat{s}(x - vt) \;, \tag{11}$$

and

$$\hat{s}_v(x,t) \;=\; \frac{1}{2t_m} \int_{-t_m}^{t_m} I(x + vt, t)\, dt \;, \tag{12}$$

Figure 7 illustrates this calculation. For each velocity we calculate the predicted intensity assuming a scene function moving at that velocity (shown in the left column). The residual intensity (shown in the right column) is explained as noise: the less energy in the residual the more likely the velocity.

Proof: A proof of claim 1 is obtained by first formulating the likelihood, $P(I|v)$, as the marginalization of the joint distribution over both I and the unknown scene function s, conditioned on v. More formally,

$$\begin{aligned}
P(I \,|\, v) &= \int_s P(s, I \,|\, v) && (13)\\
&= \int_s P(s \,|\, v) P(I \,|\, s, v) && (14)\\
&= \int_s \alpha \exp\left(-\frac{1}{2\sigma^2} \int (I(x,t) - s(x - vt))^2 dx dt \right) && (15)\\
&= \max_s \alpha \exp\left(-\frac{1}{2\sigma^2} \int (I(x,t) - s(x - vt))^2 dx dt \right) && (16)
\end{aligned}$$

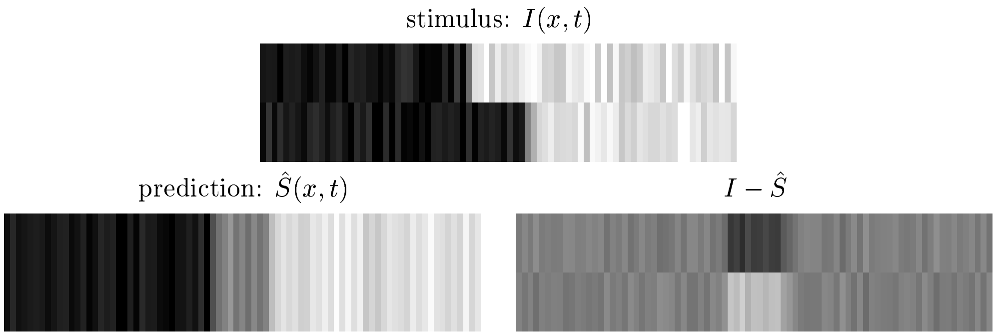


Figure 8: When the image is sampled temporally to yield two frames. The likelihood of Equation 10 is monotonically related to the sum of squared difference (SSD) criterion.

$$= \quad \alpha \exp\left(-\frac{1}{2\sigma^2}\int (I(x,t) - \hat{s}(x,t))^2 dx dt\right) \tag{17}$$

where we have used the fact that $P(s|v)$ is independent of s and of v, and that for jointly Gaussian random variables, marginalization can be replaced with maximization: $\int_z P(x,z)dz = \alpha/\sqrt{V(z|x)}\max_z P(x,z)$ where $V(z|x)$ denotes the conditional variance of z given x. The maximization over s turns into a separate maximization over $s(x)$ for each x and it is easy to see that $s(x)$ is most likely when it is equal to the mean of $I(x+vt)$ over t. $\Box$

3.3 Extensions

Of course the derivation given above makes several assumptions, many of which are somewhat restrictive. However, many of them can be relaxed in straightforward ways:

- Colored noise: If the noise is not white, then it can be shown that the likelihood becomes:

$$P(I\,|\,v) \;=\; \alpha \exp\left(-\frac{1}{2\sigma^2}\|I - \hat{S}_v\|_W^2\right) \quad . \tag{18}$$

  That is, rather than calculating the energy of the residual, we calculate a weighted energy; the weight of an energy band is inversely proportional to the expected noise variance in that band.

- Non-uniform prior over scene functions: Above we assumed that all scene functions are equiprobable. However, if we have some prior probability over the scene function, it can be shown that Equation 10 still holds but $\hat{S}_v$ is different. The estimated scene function is the one that is most probable given the prior scene probability and the observed data (unlike the present case where just the observed data determine the estimated scene function)

3.4 Connection to other models of early motion analysis

Sum of squared differences (SSD): In many computer vision applications motion is estimated using only two frames $I_1(x) = I(x,t_1)$ and $I_2(x) = I(x,t_2)$. Velocity is

Figure 9: When the image sequence is perfectly described by its linear Taylor series approximation, the likelihood of Equation 10 is a function of the gradient constraint.

chosen by minimizing:

$$SSD(v) \;=\; \int (I_1(x) - I_2(x+v))^2 dx \tag{19}$$

It is straightforward to show that if we only observe $I(x,t)$ at two distinct times t_1, t_2 then:

$$P(I_1, I_2 \,|\, v) \;=\; \alpha \, exp(-SSD(v)/4\sigma^2) \tag{20}$$

so that minimizing $SSD(v)$ is equivalent to maximizing the likelihood.

The gradient constraint: A popular computer vision algorithm for estimating local velocities [7] is to find the vector v that minimizes:

$$J_{LK}(v) \;=\; \int_x \left(\frac{\partial I}{\partial x} v + \frac{\partial I}{\partial t} \right)^2 \tag{21}$$

It can be shown that when $I(x,v)$ is well approximated by its Taylor series, i.e. $I(x+vt,t) = I(x,0) + vt\frac{\partial I}{\partial x} + t\frac{\partial I}{\partial t}$ then:

$$P(I \,|\, v) \;=\; \alpha \exp\left(-\frac{1}{2\sigma^2} \frac{2t_m^3}{3} J_{LK}(v) \right) \tag{22}$$

This derivation is based on the assumption that $I(x,t)$ is perfectly approximated by its Taylor series, an assumption that will never hold with white noise, nor exactly in practice. In most situations, thus, Equation 22 will only be a rough approximation to Equation 10. Equation 22 is also based on the assumption that the image is observed for $|t| < t_m$ and for all x. When the image is observed within a spatial window of finite extent, then the likelihood changes.

4 Connection to physiology

The most popular model for early motion calculations in primate visual cortex is based on the idea that motion is related to orientation in space-time. Accordingly, velocity tuned cells could be used to extract "motion energy" by applying space-time oriented filters to the spatiotemporal image sequence, followed by a squaring

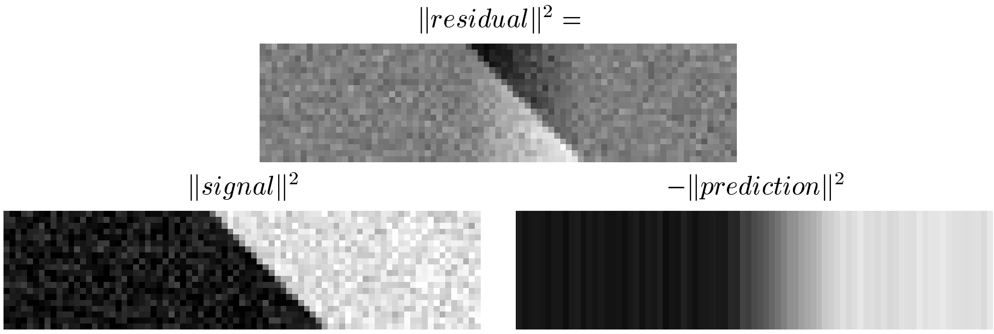


Figure 10: The energy of the residual is equal to the energy of the sequence minus that of the predicted sequence. This means that the likelihood of Equation 10 is montonically related to the motion energy of the sequence.

nonlinearity [1, 12]. The term "motion energy" refers to the fact that summing the squared output of oriented filters in all spatiotemporal bands is equivalent (by Parseval's theorem) to calculating the energy in an oriented hyperplane in the frequency domain.

To formalize this notion, we define the motion energy of a stimulus f as the energy of that stimulus convolved with an ideal oriented filter; i.e.,

$$ME(v; f) \;=\; ||f * \delta(x - vt)||^2 \; . \tag{23}$$

Equation 23 can also be interperted in the Fourier domain, where convolving f and $\delta(x - vt)$ are equivalent to multiplying their Fourier transforms $\hat{f}$ and $\delta(\omega_t + v\omega_x)$. Thus if we used infinite windows to analyze the stimulus, motion energy can be thought of as the total power of spatiotemporal frequencies that lie along the plane $\omega_t + v\omega_x = 0$. Recall, however, that our definition of energy integrates $(f * \delta(x - vt))^2$ over the window $|t| < t_m$, so that we also include spatiotemporal frequencies that are close to the plane $\omega_t + v\omega_x = 0$ but lie off it.

Claim 2: Let $P(I \,|\, v)$ be as defined in Equation 10. Then:

$$P(I \,|\, v) \;=\; \alpha \exp\left(\frac{ME(v; f)}{8\sigma^2 t_m^2}\right) \tag{24}$$

with $f = I(x, t)$ for $|t| < t_m$ and zero otherwise.

Claim 2 follows from the fact that the residual, $I - \hat{S}$, and the predicted signal $\hat{S}$ are *orthogonal* signals (see Figure 10):

$$||I - \hat{S}_v||^2 \;=\; ||I||^2 - ||\hat{S}_v||^2 \tag{25}$$

Equation 25 can be derived by performing the integration along lines of constant $x - vt$. Along such lines $\hat{S}_v$ is equal to the mean of I so that cross terms of the form $(I - \hat{S})\hat{S}$ cancel out. Using the fact that $||I||^2$ is independent of v and $||\hat{S}_v||^2 = \frac{ME(f;v)}{4t_m^2}$ gives Equation 24.

This shows that the likelihood of a velocity v can be computed as follows:

- compute the responses of a band of filters that are oriented in space-time with orientation dependent on v. The filters are shifted copies of $f(x,t) = \delta(x - vt)$
- square the output of the filters.
- pool the squared output over space.
- pass the pooled response through a pointwise nonlinearity

If the input signal is band-limited we can replace $\delta(x - vt)$ with a sufficiently skinny oriented Gaussian. Thus the log likelihood can be calculated exactly by summing the squared response of space-time oriented filters.

The main difference between this calculation and the Adelson and Bergen [1] model is that the oriented filters are not band-pass. That is, the idealized filters $\delta(x - vt)$ respond to oriented structure in any spatiotemporal frequency band. The oriented filters in Adelson and Bergen as well as in Simoncelli and Heeger [12] respond to orientation only in a band of spatiotemporal frequencies. Note that the squared response to an all-pass oriented filter can be computed by adding the squared responses to band-pass oriented filters (assuming the band-pass filters are orthogonal). It would be interesting to find conditions under which the likelihood calculation requires band-pass oriented filters.

5 Examples of likelihoods on specific stimuli

In the derivation so far, we have assumed that the sequence is observed for infinite space and finite time. Any calculation on real images, of course, will have to work with finite spatial windows. Finite spatial windows present the problem of window boundary effects. The predicted scene function at a point is the mean of all samples of this point in the window, but for finite sized windows, different velocities will have a different number of independent samples. This introduces a bias in favor of fast velocities.

To get an unbiased estimate as possible, we use windows whose spatial extent is much larger than the temporal extent. For these simulations we used rectangular windows of size $64 \times 64 \times 5$ pixels. The data for each window was obtained by zooming in on the moving square sequence shown in Figure 11.

Figures 12–15 show the results. We compare the likelihood from the generative model (Equation 10) to the likelihood from the bowtie equation (6) and the likelihood from the fuzzy line equation (5). Gradients for the fuzzy bowties and fuzzy line equations were estimated by convolving the signal with derivatives of Gaussians. It can be seen that for edge locations the generative model likelihood is approximately a fuzzy line and for corner locations it is a fuzzy blob centered on the correct velocity. When contrast is decreased the likelihood becomes more fuzzy; uncertainty increases.

The fuzzy line likelihood gives qualitatively similar likelihood functions while the fuzzy bowtie equations give a very poor approximation.

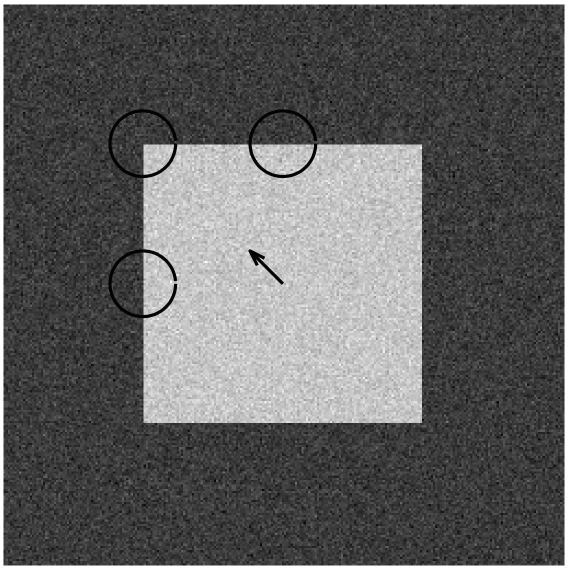
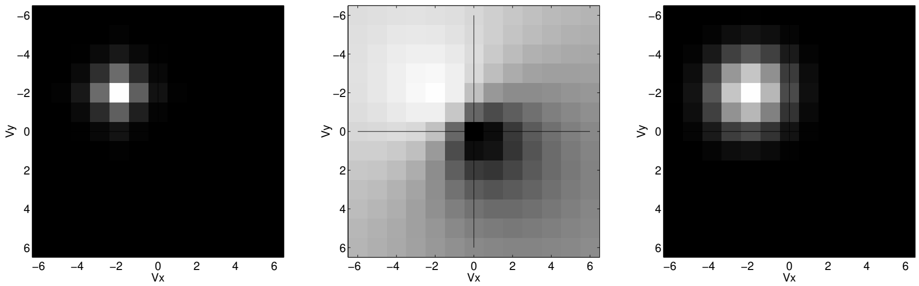
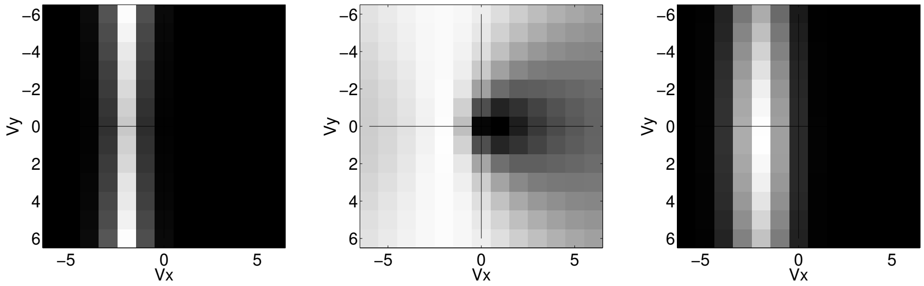


Figure 11: A single frame from the simple stimulus on which we calculated local likelihoods. The likelihoods were calculated at three locations: at the corner, side edge, and top edge. The image sequence was constructed by moving a square with velocity $(2, 2)$ and adding Gaussian noise with standard deviation 10% of the square contrast.



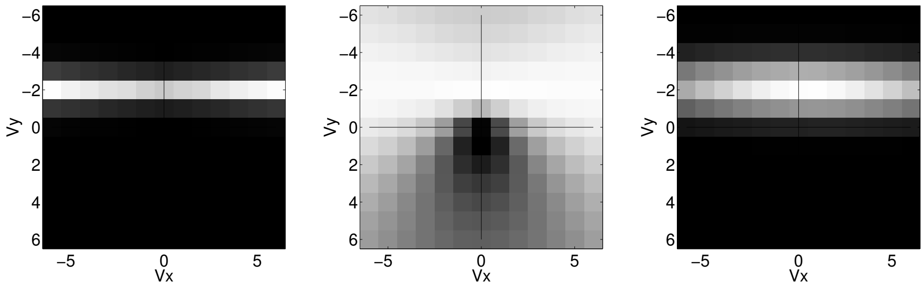
a b c

Figure 12: The likelihoods at the corner of the square calcuated using three equations. (a) The generative model likelihood (Eqn. 10) (b) the bowtie equation (Eqn. 6) (c) the fuzzy line equation (Eqn. 5).



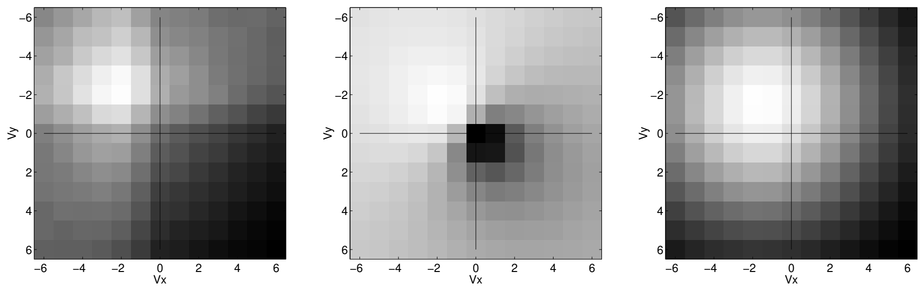
a b c

Figure 13: The likelihoods at the side edge of the square calcuated using three equations. (a) The generative model likelihood (Eqn. 10) (b) the bowtie equation (Eqn. 6) (c) the fuzzy line equation (Eqn. 5).



a b c

Figure 14: The likelihoods at the top edge of the square calcuated using three equations. (a) The generative model likelihood (Eqn. 10) (b) the bowtie equation (Eqn. 6) (c) the fuzzy line equation (Eqn. 5).



a b c

Figure 15: The likelihoods at the corner of the square calcuated using three equations. Here the contrast of the square was reduced by a factor of four and noise stays the same. Note that the likelihood becomes more fuzzy: uncertainty increases. (a) The generative model likelihood (Eqn. 10) (b) the bowtie equation (Eqn. 6) (c) the fuzzy line equation (Eqn. 5).

6 Discussion

We have briefly reviewed the successes of Bayesian models in accounting for human motion perception. These models require a formula for the likelihood of a velocity given image data. We have shown that such a formula can be derived from a simple generative model — the scene translates and conserves noise while the image equals the projected scene plus independent noise. We reviewed the connection between the likelihood function derived from this generative model and commonly used cost functions in computer vision. We also showed that the likelihood function can be calculated by summing the squared outputs of spatiotemporal oriented filters.

There are intriguing similarities between the calculation implied by the ideal likelihood function and common models for motion analysis in striate cortex. To a first approximation, complex cells in V1 can be modeled as squared outputs of spatiotemporal oriented filters. Again to first approximation, MT pattern cells can be modelled as pooling these squared responses over space [12]. This is consistent with the idea that a population of velocity tuned cells in area MT represent the likelihood of a velocity.

Acknowledgements

YW is supported by MURI-ARO-DAAH04-96-1-0341. DJF, on leave from Queen's University, is supported in part by an Alfred P. Sloan Research Fellowship.

References

[1] E. H. Adelson and J. R. Bergen. The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of the Workshop on Motion: Representation and Analysis*, pages 151–155, Charleston, SC, 1986.

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

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[4] C. Fernmuller, R. Pless, and Y. Aloimoinos. The statistics of visual correspondence: Insights into the visual system. In *Proceedings of SCTV 99*. 1999. http://www.cis.ohio-state.edu/ szhu/workshop/Aloimonos.html.

[5] D. J. Heeger and E. P. Simoncelli. Model of visual motion sensing. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*. Cambridge University Press, 1991.

[6] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, 1981.

[7] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. DARPA Image Understanding Workshop*, pages 121–130, 1981.

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

[9] O. Nestares, D.J. Fleet, and D.J. Heeger. Likelihood functions and confidence bounds for total-least-squares problems. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, Vol. II, pp. 760-767, 2000.

[10] E. P. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts of Technology, Cambridge, January 1993.

[11] E.P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, pages 310–315, 1991.

[12] E.P. Simoncelli and D.J. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, 1998.

[13] H. Wallach. Ueber visuell whargenommene bewegungrichtung. *Psychologische Forschung*, 20:325–380, 1935.

[14] 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.

[15] Y. Weiss and Edward H. Adelson. Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision. Technical Report 1624, MIT AI lab, 1998.

[16] C. Yo and H.R. Wilson. Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Research*, 32(1):135–147, 1992.