A Robust Gradient Method for Determining Optical Flow

Michael J. Black

Research Report YALEU/DCS/RR-891
October 1991

A Robust Gradient Method for Determining Optical Flow

Michael J. Black*

Department of Computer Science
Yale University
P.O. Box 2158 Yale Station
New Haven, CT 06520–2158

Phone: (203) 432–1223
Fax: (203) 432–0593
Email: black-michael@cs.yale.edu

Abstract

This paper presents a simple, robust, gradient-based, approach to computing optical flow in the presence of noise and motion discontinuities. We begin with the standard gradient-based method of Horn and Schunck which corresponds to a least-squares estimate of the optical flow. Such estimates are susceptible to outliers which do not conform the statistical assumptions of the approach; for example, measurements at motion discontinuities. The result is that the least-squares solution has the undesirable property of smoothing the flow field across motion boundaries. Robust statistics can be used to address this problem. By reformulating the optical flow equation in terms of robust estimation, the problems of image noise and over-smoothing are reduced. The reformulation is straightforward, and results in a remarkable improvement in the estimated flow, particularly at motion discontinuities. The flow field is recovered using a simple, deterministic, relaxation scheme which is described in sufficient detail to allow it to be easily implemented.

*This work was supported by a grants from the National Aeronautics and Space Administration (NGT–50749) and the Office of Naval Research (N00014–91–J–1577).

1 Introduction

This paper presents a simple and elegant solution to the problem of robustly recovering optical flow in the presence of noise and motion discontinuities. The standard optical flow equation is reformulated in a straightforward way to make it robust in a formal sense. The resulting *robust gradient* equation is less sensitive to noise and smoothing across motion discontinuities. The flow field is recovered using a simple iterative relaxation scheme.

The standard Horn and Schunck [11] formulation of optical flow is ubiquitous in motion analysis. The formulation represents a least squares estimate of the flow field. Such estimates are commonly known to be sensitive to measurements that do not conform to the statistical assumptions of the model. In a neighborhood about a motion discontinuity, the local differences in flow across the boundary can be viewed as statistical *outliers* that corrupt the least squares solution, resulting in smoothing of the flow field across the discontinuity. The least-squares formulation is also sensitive to errors in the intensity constraint equation resulting from occlusion, disocclusion, shadows, and specular reflections.

The problem is one of fitting a model to data when we know that the data may contain outliers which differ statistically from the model assumptions. We make the simple observation that the field of *robust statistics* has developed techniques which address this problem. By recasting the least-squares formulation of optical flow in the robust statistics framework, the recovery of optical flow is made less sensitive to gross violations of the intensity constraint equation and smoothing across motion boundaries.

The least-squares formulation results in a convex objective function which is easy to minimize. The *robust gradient* formulation, however, generally results in a non-convex objective function. A Graduated Non-Convexity (GNC) algorithm [6] is presented for recovering the optical flow. Experimental results on real and synthetic image sequences are very promising. The robust gradient approach does an excellent job of preserving motion discontinuities while reducing the effects of image noise.

1.1 Previous Work

While qualitative estimates of optical flow may be sufficient for a number of problems like obstacle detection, and coarse motion segmentation, many uses of optical flow require robust estimates; for example structure from motion. The lack of robustness has been seen as a major weakness of optical flow [1].

A number of approaches have been developed to recover discontinuous optical flow without over-smoothing. The most notable techniques are the Markov random field (MRF) formulations [4, 15, 17]. These approaches represent discontinuities either explicitly with the use of a "line process" or by using *weak continuity constraints* [6, 7]. These approaches achieve good results but typically rely on expensive stochastic minimization procedures. In a similar vein, there have been attempts to incorporate ideas like weak continuity in analog devices [9, 14]. Another approach uses statistics about the flow in a neighborhood to control the smoothing process [21].

The robust gradient approach is similar in spirit to the work of Black and Anandan [4, 5], in that both approaches take a familiar formulation of optical flow and reformulate it to account for motion discontinuities. The approach here, however, is conceptually simpler,

easier to motivate from the standpoint of robust statistics, achieves better results, and is more efficient. The efficiency gain is due to the fact that with the robust gradient formulation the objective function is differentiable; a fact that can be exploited when minimizing the function.

There are many similarities between the proposed approach and the work of Blake and Zisserman on image reconstruction [6]. The motivations however are quite different. Here we have taken a well known least-squares optimization problem and reformulated it using robust statistics. The resulting approach to dealing with spatial discontinuities turns out to be closely related to the notion of weak continuity. Blake and Zisserman however, ignore the possibility of outliers in the data measurements and concentrate on spatial discontinuities.

Finally, Schunck [20] proposed the idea of using robust statistics in his work on constraint line clustering, but avoided the direct robust formulation presented here. He does, however suggest that "further experiments should be conducted to compare robust estimates with constraint line clustering," ([20], p. 1018). In the case of constraint line clustering, the data error and over-smoothing problems are treated separately. Our approach has the advantage of being very straightforward in that is brings robust statistics to bear, in a uniform way, on the problems of data errors and motion discontinuities.

In the next section, we review the standard Horn and Schunck formulation of optical flow. In section 3 we introduce robust statistics. In section 4 we then reformulate the optical flow equation using robust statistics. Section 5 examines presents an iterative minimization technique for recovering the optical flow. We then describe experimental results with real and synthetic image sequences and compare the results with other approaches.

2 The Least Squares Formulation

This section reviews the standard least-squares formulation of the optical flow problem [10, 11]. Let $I(x, y, t)$ be the image intensity[1] at a point (x, y) at time t. The standard *intensity constancy assumption* is written as follows:

$$I(x, y, t) = I(x + u\delta t, y + v\delta t, t + \delta t), \tag{1}$$

where $\mathbf{u} = [u, v]^T$ is the horizontal and vertical image velocity at a point and δt is small.

By Taylor series expansion of the right hand side of (1) we have:

$$I(x, y, t) = I(x, y, t) + I_x u\delta t + I_y v\delta t + I_t \delta t + \epsilon, \tag{2}$$

where I_x, I_y, and I_t are the first partial derivatives of the brightness I with respect to x, y, and t respectively, and where ϵ contains the higher-order terms. Simplifying and dividing through by δt then taking the limit as $\delta t \rightarrow 0$ we obtain the standard optical flow constraint equation:

$$I_x u + I_y v + I_t = \nabla I^T \mathbf{u} + I_t = 0. \tag{3}$$

We refer to (3) as a *data conservation* assumption [4]. The equation is valid only for small displacements (ie. less than a pixel). Larger motions can be dealt using a hierarchical, coarse-to-fine, strategy [2].

[1] In fact, we always work with a Laplacian filtered version of the image.

It is well known that the recovery of optical flow is ill-posed [3, 18]. To recover the flow we regularize the ill-posed problem with the introduction of a *smoothness constraint.* Take as a measure of smoothness the square of the velocity field gradient:

$$u_x^2 + u_y^2 + v_x^2 + v_y^2. \tag{4}$$

We wish to find the optical flow field which minimizes the combined intensity and smoothness error:

$$E(\mathbf{u}) = \int\int \lambda(\nabla I^T \mathbf{u} + I_t)^2 + (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy, \tag{5}$$

where λ controls the relative importance of the two terms.

For an image of size $n \times n$ pixels we define a grid of *sites:*

$$S = \{s_1, s_2, \ldots, s_{n^2} \mid \forall w \; 0 \leq i(s_w), j(s_w) \leq n-1\},$$

where $(i(s), j(s))$ denotes the pixel coordinates of site s. Equation 5 is then easily discretized as:

$$E(\mathbf{u}) = \sum_{s \in S} [\lambda(\nabla I^T \mathbf{u}_s + I_t)^2 + \frac{1}{8} \sum_{n \in \mathcal{G}_s} (\mathbf{u}_s - \mathbf{u}_n)^T (\mathbf{u}_s - \mathbf{u}_n)], \tag{6}$$

where $\mathcal{G}_s$ represents the set of north, south, east, west neighbors of s in the grid.

This formulation admits a simple iterative relaxation scheme for determining the optical flow:

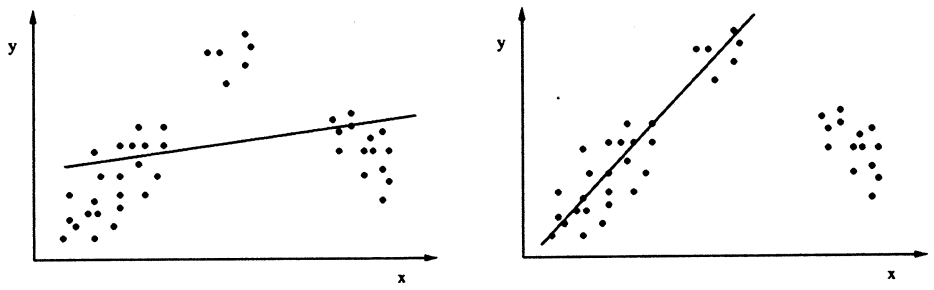
$$u_s^{(n+1)} = \bar{u}_s^n - \frac{I_x(I_x \bar{u}_s^n + I_y \bar{v}_s^n + I_t)}{1 + \lambda(I_x^2 + I_y^2)}, \tag{7}$$

$$v_s^{(n+1)} = \bar{v}_s^n - \frac{I_y(I_x \bar{u}_s^n + I_y \bar{v}_s^n + I_t)}{1 + \lambda(I_x^2 + I_y^2)}, \tag{8}$$

where $\bar{u}$ and $\bar{v}$ are local averages of u and v. The problem with this scheme however, is that to reduce the effects of noise one must oversmooth the flow field. Remaining faithful to the image measurements on the other hand results in a noisy flow field. What is needed is a way to ignore noisy measurements and at the same time prevent smoothing across discontinuities.

3 Robust Statistics

Robust statistics addresses the problem of fitting a model $\mathbf{u}$ to data, $\mathbf{d} = \{d_0, d_1, \ldots, d_S\}$, in cases where the data differs statistically from the model assumptions [8, 12]. For example, we may assume that measurement errors can be characterized by a Gaussian distribution. We may also know, however, that this assumption is only approximate; that there will be errors which do not fit the Gaussian assumptions and for which we have no statistical model. The *robustness* of a statistical estimator then refers to its insensitivity to deviations from the assumed statistical model. These deviations are called *outliers.*



a *b*

Figure 1: **Fitting a straight line.** *a)* Least-squares fit. *b)* Robust fit.

Assume that the error in the a data set $\mathbf{d}$ is normally distributed about the "true" model $\mathbf{u}$ with standard deviation σ. Then the probability of the data set is given by:

$$P = \prod_{s \in S} \{\exp[-\rho(d_s, u_s, \sigma_s)]\}, \tag{9}$$

where the *estimator* ρ is:

$$\rho(d_s, u_s, \sigma_s) = \frac{(d_s - u_s)^2}{2\sigma_s}. \tag{10}$$

Maximizing P is equivalent to minimizing the negative of its logarithm, which is equivalent to finding the $\mathbf{u}$ which minimizes:

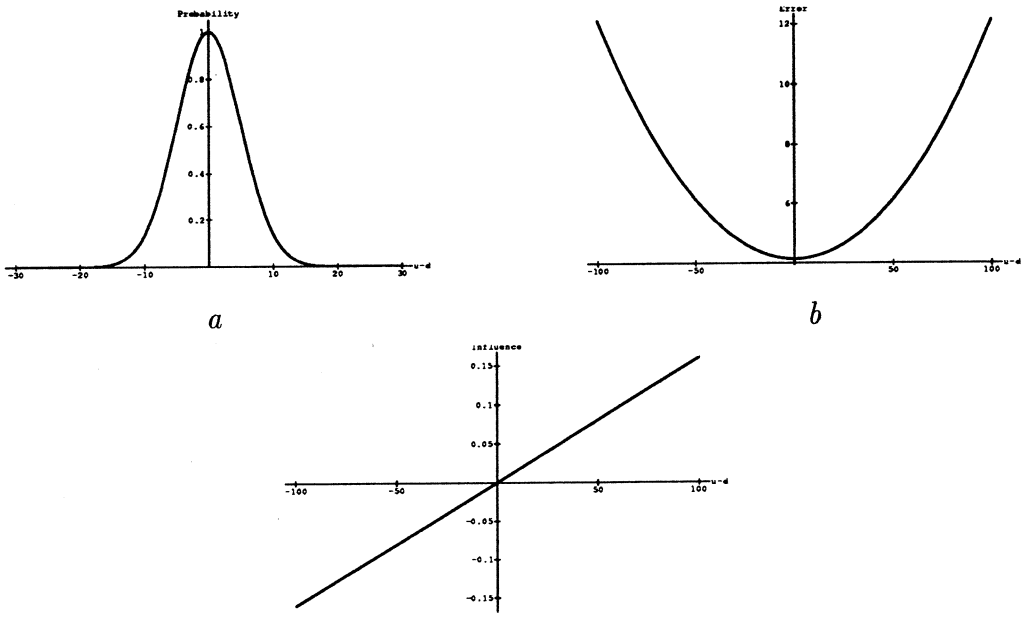
$$\min_{\mathbf{u}} \sum_{s \in S} \rho(d_s, u_s, \sigma_s) \tag{11}$$

For the case where P is Gaussian and ρ is quadratic, this is the standard least squares error. The function ρ is called an *M-estimator* since it corresponds to the *Maximum-likelihood* estimate.

3.1 Problems With the Least Squares Approach

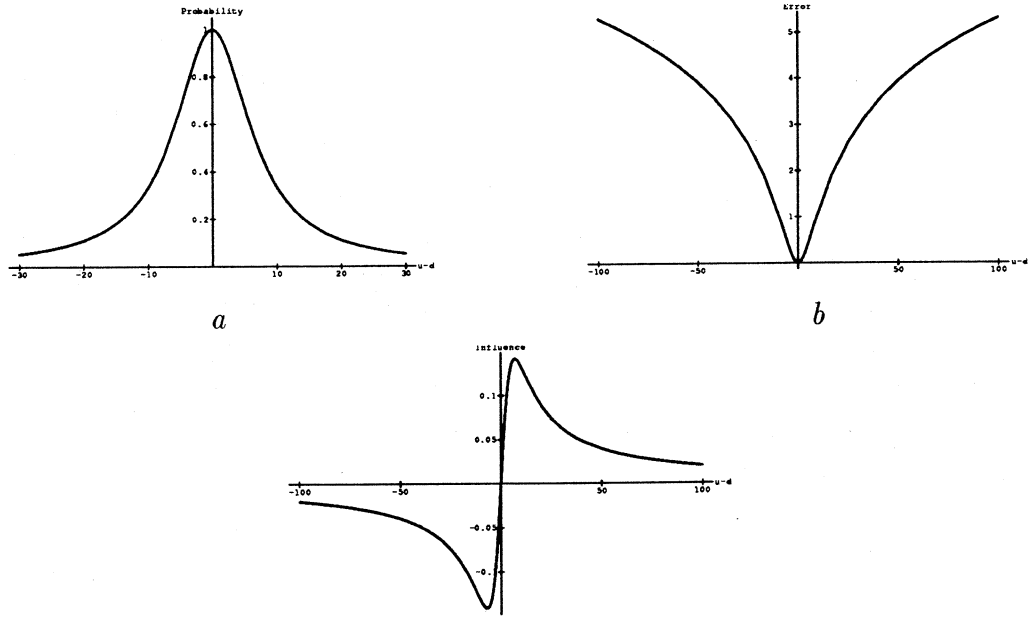
This standard approach is not without its problems. When the noise is, in fact, not Gaussian, the solution is skewed from the "true" solution. Figure 1 shows an example of fitting a line to data in the presence of outliers. Figure *a* shows how the least-squares fit is skewed in the direction of the outliers. The fit recovered in figure *b* is robust in the sense that it rejects the outliers and recovers a "better" fit to the *majority* of the data.

The problem with the least-squares solution is that the outliers contribute "too much" to the overall solution. Figure 2*a* shows a Gaussian probability distribution. The tails of the distribution drop off quickly and hence, outlying points have low probability of occurring. Consequently, when they do occur they are assigned a high weight by the quadratic estimator 2*b*. The *influence* ψ of measurements is determined by the derivative of the estimator [8]. In the least-squares case, the influence of data points increases linearly and without bound (figure 2*c*).



a *b* *c*

Figure 2: **Least Squares Estimation.** *a)* Gaussian probability distribution, *b)* Quadratic error measure (negative logarithm of the probability distribution), *c)* Influence function (derivative of the error measure).



a *b* *c*

Figure 3: **Robust Estimator (Lorentzian).** *a)* Probability distribution, *b)* Error measure (negative logarithm of the probability distribution), *c)* Influence function (derivative of the error measure).

To construct a robust estimator, we need to be more forgiving about outlying measurements. Figure 3*a* shows the *Lorentzian* distribution [19]:

$$\text{Prob}(d_s - u_s, \sigma_s) \sim \frac{1}{1 + \frac{1}{2}\left(\frac{d_s - u_s}{\sigma_s}\right)^2}. \tag{12}$$

Notice that the tails of this distribution decrease more gradually than the Gaussian. Once again, the appropriate estimator is defined as the negative logarithm of the probability distribution:

$$\rho(d_s - u_s, \sigma_s) = \log\left(1 + \frac{1}{2}\left(\frac{d_s - u_s}{\sigma_s}\right)^2\right). \tag{13}$$

The estimator is shown in figure 3*b*. Unlike the quadratic estimator, the robust estimator has a saturating property; as the measurement errors increase, the contribution of the measurements approaches a constant. This can be seen by examining the influence function:

$$\psi(d_s - u_s, \sigma_s) = \frac{2(d_s - u_s)}{2\sigma_s^2 + (d_s - u_s)^2}, \tag{14}$$

which is shown in figure 3*c*. As the error increases, the influence of the measurements approaches zero.

4 A Robust Formulation of Optical Flow

As we saw in the previous section, least-squares solutions are sensitive to outliers. Where do outliers affect the determination of optical flow? First, the intensity constraint equation (3) can be violated in a number of commonly occurring situations. For example, it will be violated at motion boundaries where occlusion or disocclusion takes place. Also, in cases of specular reflection or shadow boundaries, the prediction of the constraint equation does not correspond to the motion of the physical surfaces. In these cases it may be desirable to "ignore" the intensity constraint if doing so produces a "better" interpretation of the scene. With the least-squares approach the optimal estimate will be skewed towards these anomalous measurements.

To reduce the effects of noise in the recovered flow field, the relative importance of the smoothness constraint can be increased. This, however, results in over-smoothing at motion boundaries. Take, for example the motion boundary in figure 4. When trying to determine the flow at a point (i, j) we examine the flow at neighboring points to the north, south, east and west. The flow vectors at the points $(i, j+1)$, $(i-1, j)$, and $(i, j-1)$ are all consistent with the central point, and we can expect the differences between these neighboring points to be captured quite well by a Gaussian distribution.

The point $(i+1, j)$, however, belongs to a different surface. The flow here is inconsistent with the other flow vectors, and hence can be viewed as a statistical outlier. Employing a least squares smoothness constraint will force the flow at (i, j) away from its true value. What we would like to do is recognize that $(i+1, j)$ does not belong to the true distribution and ignore it.

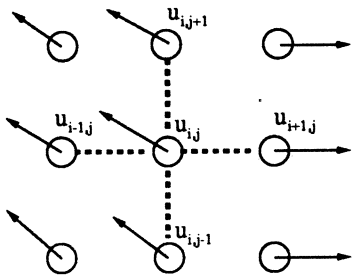


Figure 4: Smoothing across a flow discontinuity.

By reformulating the motion equation using the robust statistical techniques of the previous section, we can achieve a more robust computation of optical flow in the presence of erroneous image measurements and motion discontinuities. The least-squares form of the optical flow equation (6) can be reformulated as:

$$E(u,v) = \sum_{s \in S} [\lambda \rho(I_x u_s + I_y v_s + I_t, \sigma_1) + \sum_{n \in \mathcal{G}_s} \rho(u_s - u_n, \sigma_2) + \sum_{n \in \mathcal{G}_s} \rho(v_s - v_n, \sigma_2)]. \quad (15)$$

When ρ is the quadratic error measure, this is the least-squares optical flow equation (6). For the robust formulation, we simply replace the quadratic error measure by the more robust Lorentzian M-estimator. We explore the implications of this reformulation in the remainder of this and the following section.

4.1 Discontinuities and Parameter Estimation

We would like to be able to set thresholds τ_1 and τ_2 that determine what data and smoothness errors are considered outliers. To do so we need to determine the appropriate values for σ_1 and σ_2. These values determine the point at which measurements are considered outliers. This is taken to be point where the influence of the measurements begins to decrease; that is where the derivative of the influence function:

$$\frac{\partial^2 \rho}{\partial x^2} = \frac{\partial \psi}{\partial x} = \frac{2(2\sigma^2 - x^2)}{(2\sigma^2 + x^2)^2}, \quad (16)$$

equals zero. This occurs when:

$$x = \pm \sqrt{2}\, \sigma. \quad (17)$$

So to define an outlier threshold τ, we set $\sigma = \tau / \sqrt{2}$.

For example, in the case of the smoothness constraint, if a difference of 0.05 pixels is considered a discontinuity then $\sigma_2 = 0.0353553$. This threshold could presumably be set on the basis of psychophysical evidence. Motion discontinuities can be trivially recovered from the computed flow field by examining where this threshold is exceeded.

For the data term we make a conservative estimate of the variance in the intensity error for the optimal flow field. We do this by computing the intensity error $\nabla I^T \mathbf{u} + I_t$ in the case where the flow is zero everywhere; ie. the error is simply I_t. We then compute the variance of this initial error and take that as the value of σ_1.

4.2 Convexity

The least squares formulation of optic flow is relatively straightforward to solve since the objective function is convex. The robust formulation, however, may not be convex. This occurs since, if the data and smoothness terms disagree, we can minimize for either one and treat the other as an outlier.

Formally, the objective function is convex when the *Hessian matrix*:

$$H = \begin{bmatrix} \frac{\partial^2 E}{\partial u^2} & \frac{\partial^2 E}{\partial u \partial v} \\ \\ \frac{\partial^2 E}{\partial v \partial u} & \frac{\partial^2 E}{\partial v^2} \end{bmatrix} \tag{18}$$

is positive definite [6]. This condition is met if and only if both eigenvalues of the matrix H are positive. This gives us a simple test for convexity. It is easy to show that E is convex when:

$$\max_{s \in S} |(\nabla I^T \mathbf{u}_s + I_t)| \leq \sqrt{2}\sigma_1 = \tau_1, \text{ and} \tag{19}$$

$$\max_{s \in S} \max_{n \in \mathcal{G}_s} |\mathbf{u}_s - \mathbf{u}_n| \leq \sqrt{2}\sigma_2 = \tau_2. \tag{20}$$

This corresponds to the case where there are no data or spatial outliers. In this range, the influence function ψ is roughly linear and the error function ρ is roughly quadratic. On the other hand, E is non-convex when:

$$\max_{s \in S} |(\nabla I^T \mathbf{u}_s + I_t)| > \sqrt{2}\sigma_1 = \tau_1, \text{ and} \tag{21}$$

$$\max_{s \in S} \max_{n \in \mathcal{G}_s} |\mathbf{u}_s - \mathbf{u}_n| > \sqrt{2}\sigma_2 = \tau_2. \tag{22}$$

Finally, when one term is above the outlier threshold and the other is below, the function may or may not be convex. The minimization of this objective function is the topic of the following section.

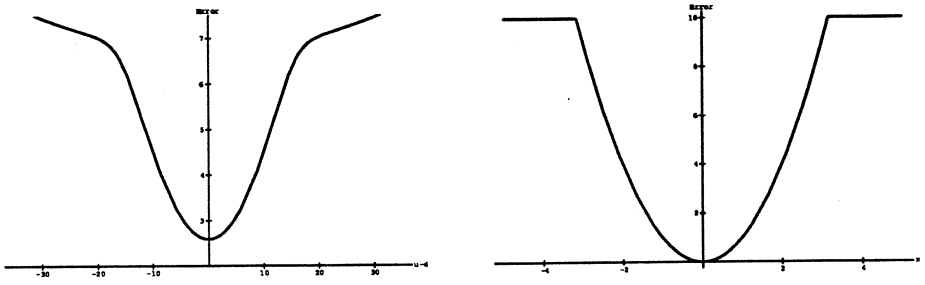
4.3 Modeling Outliers

The robust statistical formulation of the problem seems a natural evolution for the standard least squares formulation. Yet, a robust formulation makes no explicit assumptions about the statistical nature of the outliers. In motion estimation, however, the outliers correspond to physically meaningful events and hence will not be simply uncorrelated noise. The question arises: Can we do better by explicitly modeling the statistics of the outliers?

For example, a *contaminated Gaussian* model [23] provides another model of the expected data errors:

$$p(d_s|\mathbf{u}) = \frac{1-\epsilon}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(d_s - u_s)^2}{2\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(d_s - u_s)^2}{2\sigma_2^2}\right) \tag{23}$$

where $\sigma_1 << \sigma_2$ and $\epsilon << 1.0$. This model assumes that measurements are typically governed by a Gaussian distribution with small variance, but occasionally are characterized by a Gaussian with a large variance. The negative logarithm of this distribution gives the



a *b*

Figure 5: **Other estimators.** *a)* Contaminated Gaussian estimator. *b)* Truncated quadratic estimator.

estimator in figure 5*a*. This estimator does not have the saturating property that we require to reject outliers and hence be robust.

To model discontinuities, the notion of *weak continuity* can be employed [4, 6, 7]. Here, discontinuities are modeled by binary valued "line processes" (l) in the objective function:

$$E(d_s, u_s, l) = (1 - l)\lambda^2(d_s - u_s)^2 + \alpha l, \tag{24}$$

where α and λ are constant, and, where $l = 0$ if there is no discontinuity and $l = 1$ if there is. Blake and Zisserman [6] showed that the line processes can be eliminated from the objective function by first minimizing over it, resulting in an objective function which is solely a function ρ of the actual variables under consideration:

$$E^*(d_s, u_s) = \rho_{\alpha,\lambda}(d_s - u_s), \text{ where,} \tag{25}$$

$$\rho_{\alpha,\lambda}(x) = \begin{cases} \lambda^2 x^2 & \text{if } |x| < \sqrt{\alpha}/\lambda, \\ \alpha & \text{otherwise.} \end{cases} \tag{26}$$

The resulting *truncated quadratic* estimator (figure 5*b*) is similar to our Lorentzian M-estimator. The truncated quadratic has an advantage in that it is motivated by a model of discontinuities. Unfortunately, it is not smooth, which makes it less appealing for our purposes. Certainly, the relationship between weak continuity constraints and robust statistics deserves further study.

5 Minimization

The robust formulation of the optical flow equation is straightforward. However, as was shown in the previous section, the resulting objective function may be non-convex. The situation, however, is not without hope. There are a number of minimization techniques which can be brought to bear on the problem. First we will describe a local optimization technique which rapidly converges to a local minimum. We then consider a number of global optimization strategies.

5.1 Simultaneous Over-Relaxation

Simultaneous Over-Relaxation (SOR) belongs to a family of relaxation techniques which include *Jacobi's* method and the *Gauss-Seidel* method [19, 22, 24]. We compute the first partial derivatives of the robust flow equation (15):

$$\frac{\partial E}{\partial u_s} = \sum_{s \in S}[\lambda I_x \psi(\nabla I^T \mathbf{u}_s + I_t, \sigma_1) + \sum_{n \in \mathcal{G}_s} \psi(u_s - u_n, \sigma_2)], \tag{27}$$

$$\frac{\partial E}{\partial v_s} = \sum_{s \in S}[\lambda I_y \psi(\nabla I^T \mathbf{u}_s + I_t, \sigma_1) + \sum_{n \in \mathcal{G}_s} \psi(v_s - v_n, \sigma_2)]. \tag{28}$$

Then the iterative update equations for minimizing E at step $n+1$ are simply [6]:

$$u_s^{(n+1)} = u_s^{(n)} - \omega \frac{1}{T(u_s)} \frac{\partial E}{\partial u_s}, \tag{29}$$

$$v_s^{(n+1)} = v_s^{(n)} - \omega \frac{1}{T(v_s)} \frac{\partial E}{\partial v_s}, \tag{30}$$

where ω is an *overrelaxation parameter* which is used to *overcorrect* the estimate of $u^{(n+1)}$ at stage $n+1$.

The terms $T(u_s)$ and $T(v_s)$ are upper bounds on the second partial derivatives of E:

$$T(u_s) \geq \frac{\partial^2 E}{\partial u_s^2}, \quad \forall s \in S, \tag{31}$$

$$T(v_s) \geq \frac{\partial^2 E}{\partial v_s^2}, \quad \forall s \in S. \tag{32}$$

The second derivative is maximized when both the data and smoothness errors are zero everywhere, which implies:

$$T(u_s) = \frac{\lambda I_x^2}{\sigma_1^2} + \frac{4}{\sigma_2^2}, \tag{33}$$

$$T(v_s) = \frac{\lambda I_y^2}{\sigma_1^2} + \frac{4}{\sigma_2^2}. \tag{34}$$

When $0 < \omega < 2$ the method can be shown to converge [24] and when $\omega = 1$ the method is equivalent to the more familiar *Gauss-Seidel* method. To achieve the fastest convergence, the optimal value of ω should be chosen. While determining the optimal ω is difficult in the case of a non-linear problem, we can get a rough approximation by computing the optimal value for the linear Jacobi version of the problem. The optimal ω is then defined in terms of the largest eigenvalue ($\mu_{\max}$) of the Jacobi iteration matrix which can be shown to be:

$$\mu_{\max} = \cos \pi h, \tag{35}$$

$$h = \frac{1}{(n+1)}, \tag{36}$$

for an $n \times n$ problem [22]. The optimal overcorrection is then:

$$\omega_{opt} = \frac{2(1 - \sqrt{1 - \mu_{\max}^2})}{\mu_{\max}^2}. \tag{37}$$

For example, a 128×128 image would have an overcorrection of $\omega_{opt} = 1.95209$, and a 64×64 image would have $\omega_{opt} = 1.90645$. In practice, this approximation works well and for an $n \times n$ problem acceptable convergence is reached within only n iterations.

The algorithm can be implemented sequentially, but is inherently parallel. When updating a site, its four neighbors must remain fixed. A parallel algorithm is achieved by dividing the grid into black and white sites in a chessboard pattern, and updating first the black and then the white sites.

Faster convergence can be achieved using *Chebyshev acceleration* [19]. Here the value of ω is updated after each half-iteration (ie. after updating all the white, or all the black, sites) using the following scheme:

$$\begin{aligned}
\omega^{(0)} &= 1, \\
\omega^{(1/2)} &= 1/(1 - \mu_{\max}^2/2), \\
\omega^{(n+1/2)} &= 1/(1 - \mu_{\max}^2 \omega^{(n)}/4), \quad n = 1/2,\, 1, \ldots, \infty, \\
\omega^{(\infty)} &\rightarrow \omega_{opt}.
\end{aligned}$$

5.2 Global Minimization

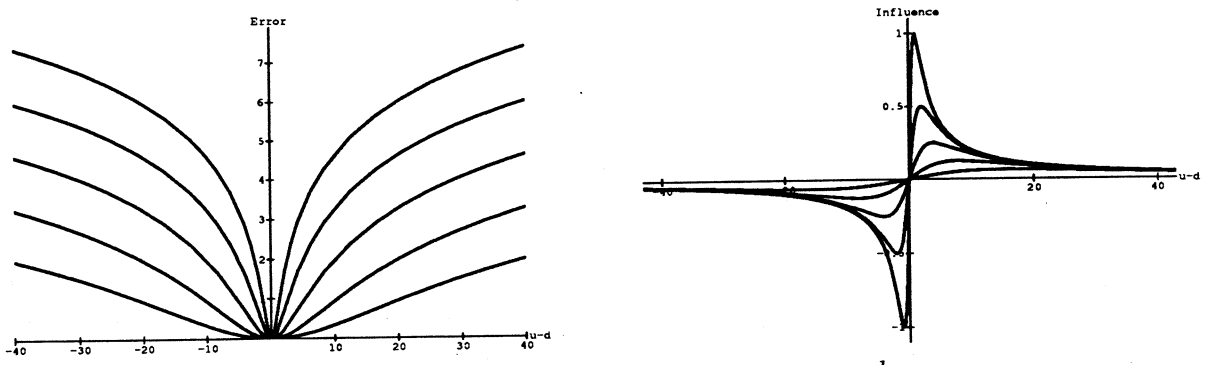
We now turn to the problem of finding a globally optimal solution when the function is non-convex. Two general approaches are considered; graduated non-convexity, and stochastic relaxation.

5.2.1 Graduated Non-Convexity

Graduated Non-Convexity (GNC) was proposed by Blake and Zisserman [6] as a deterministic minimization strategy for non-convex functions. The general idea is to take the non-convex objective function and construct a convex approximation. This approximation is then readily minimized; using, for example, the SOR technique above. Successively better approximations of the true objective function are then constructed and minimized starting from the solution of the previous approximation. For a given objective function the challenge is to construct the sequence of approximations.

Given the Lorentzian as the robust estimator there is a natural sequence of approximations. In the previous section, it was noted that E is convex if the outlier thresholds τ_1 and τ_2 are set to be greater than maximum data and smoothness errors. Assume that the motion in the scene is constrained to be less than some constant; for example, by using a hierarchical approach [2]. Then we choose τ_2 to be twice the largest allowable motion. The maximum data error can be conservatively estimated from the images. First assume that the flow is zero everywhere, so $\nabla I^T \mathbf{u} + I_t = I_t$. Now we take as our estimate $\tau_1 = \max |I_t|$.

The minimization can begin with this convex approximation and the resulting coarse flow field approximation will contain no flow discontinuities. In this sense it will be very much



a *b*

Figure 6: **Graduated Non-Convexity.** $\rho(x,\sigma)$ and $\psi(x,\sigma)$ plotted for thresholds $\tau \in \{16,8,4,2,1\}$. *a)* Error measure $\rho(x,\sigma)$. *b)* Influence function $\psi(x,\sigma)$.

like the least-squares flow estimate. Discontinuities can be gradually added by lowering the thresholds τ_1 and τ_2 and repeating the minimization. Figure 6 shows the error function (figure 6*a*) and the influence function (figure 6*b*) for various values of τ. In practice, a two stage minimization works well. First the coarse convex approximation is used, followed by the original objective function.

5.2.2 Stochastic Minimization

Stochastic approaches provide another way of minimizing a non-convex objective function. They have been used by a number of authors for computing optic flow with non-convex objective functions [4, 15, 17]. The approach taken is to construct a probability distribution Π over the space of displacements using a *Gibbs distribution*:

$$\Pi(\mathbf{u}) = Z^{-1} e^{-E(\mathbf{u})/T}, \tag{38}$$

where Z is the normalizing constant:

$$Z = \sum_{\mathbf{u}} e^{-E(\mathbf{u})/T}, \tag{39}$$

and where T is a *temperature* constant at time t which serves to sharpen (or flatten) the distribution.

The minimum of E can then be found using *simulated annealing* [13]. This is done by choosing values of $\mathbf{u}$ according to the distribution Π with logarithmicly decreasing temperatures. As the temperature is lowered, the probability distribution Π becomes concentrated about the minimum while the stochastic nature of the process prevents the estimate from getting trapped in local minima. The result is that at high temperatures the sampling process freely chooses values of $\mathbf{u}$, but as the temperature is lowered, the minimum is chosen with increasing probability. In the limit, this process converges to the correct solution when a logarithmic cooling schedule is used.

Simulated annealing is expensive in practice. While it is useful for a global exploration of the energy landscape, it does not exploit the fact that the local minima of E can be explored

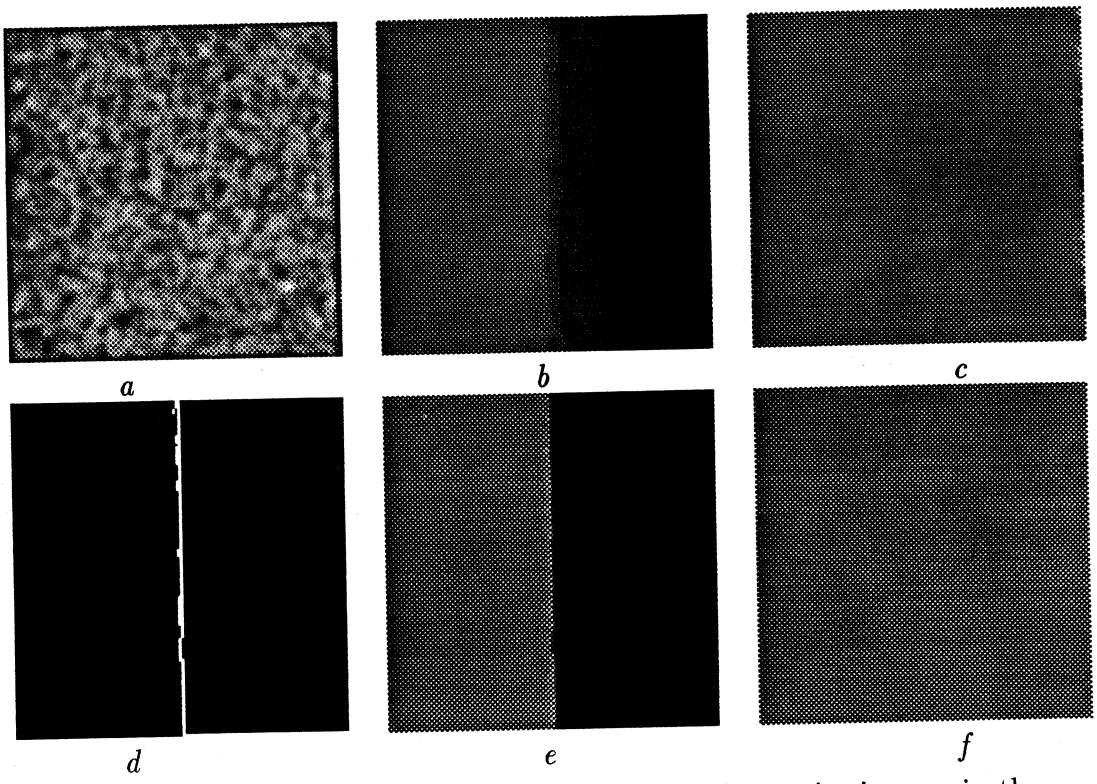


Figure 7: **Random Noise Sequence.** *a)* One of the random noise images in the sequence. *b, c)* Least-squares solution; horizontal and vertical components of the flow. *d)* Motion discontinuities where the smoothness constraint is violated. *e, f)* Robust gradient solution; horizontal and vertical components of the flow.

more in a directed fashion. A more efficient procedure would exploit the differentiable nature of E to perform a directed gradient search (like the SOR algorithm) of local minima, while occasionally making large jumps to avoid getting trapped in a local minimum. Such a procedure is similar to the *large-step Markov chain* approach to solving the Traveling Salesman Problem [16].

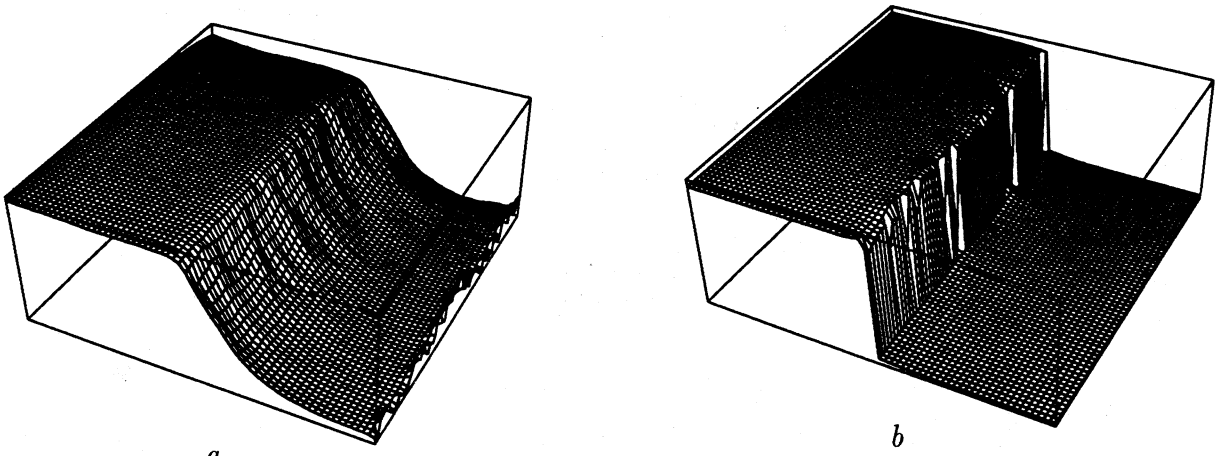
6 Experimental Results

The robust gradient technique is implemented using the GNC algorithm of the previous section. The current Connection Machine implementation fully exploits the parallelism inherent in the formulation. The algorithm, however, is easy to implement and the previous section should provide sufficient detail to implement a serial version.

The algorithm is compared with a standard implementation of the Horn and Schunck least-squares solution. It is not surprising that the robust version performs better, but it is surprising that such a simple reformulation results in such radical improvements.

All experiments were performed using 200 iterations[2] of each algorithm even though 200 iterations are not typically necessary in the case of SOR. The only parameters which need

[2]An iteration is taken to mean the updating of every site in the flow field.



a *b*

Figure 8: **Horizontal Displacement.** The horizontal component of motion is plotted to illustrate the over-smoothing of the least-squares solution (a), and the sharp discontinuity which is preserved by the robust gradient technique (b).

to be empirically determined are τ_2 and λ. These were chosen, and remained unchanged for all the experiments: $\tau_2 = 0.05$, $\lambda = 10$ for the robust gradient approach, and $\lambda = 50$ for the least-squares approach. All other parameters were determined as specified in the previous section.

6.1 Synthetic Sequence

The first experiment explores an easy case with no noise. The synthetic sequence contains two highly textured surfaces, one which is stationary and one which is translating to the left (figure 7a). The horizontal and vertical components of the computed flow are shown with the magnitude flow coded by intensity, where black indicates motion to the left and up and, similarly, white indicates motion to the right and down. Figures 7b and c show the flow computed with the least squares formulation. Notice how the horizontal flow is smoothed across the motion boundary. The robust gradient technique does not suffer from over-smoothing (figure 7e,f). Motion discontinuities can be detected (figure 7d) by examining where the smoothness assumption is violated (ie. when the absolute difference in neighboring flow vectors is greater than τ_2).

This example illustrates how, even when no noise is present, the least-squares approach can perform badly by smoothing across discontinuities. The contrast between the approaches is vividly observable in figure 8. Plotting the horizontal component of the flow field graphically shows the behavior of the two algorithms at motion boundaries.

Figure 9 shows the convergence behavior of the two algorithms. The faster convergence of the robust gradient algorithm is due to the use of over-relaxation. This was not used in the least-squares case. The least-squares approach, however, does not approach the error of the robust gradient technique. This is a result of over-smoothing error at the motion boundary.

The effects of noise are explored in figures 10 and 11. The figures show the effects of adding 5 percent uniform noise to the second image in the sequence. The discontinuity is still clearly preserved by the robust approach (figure 10c). With the standard smoothness

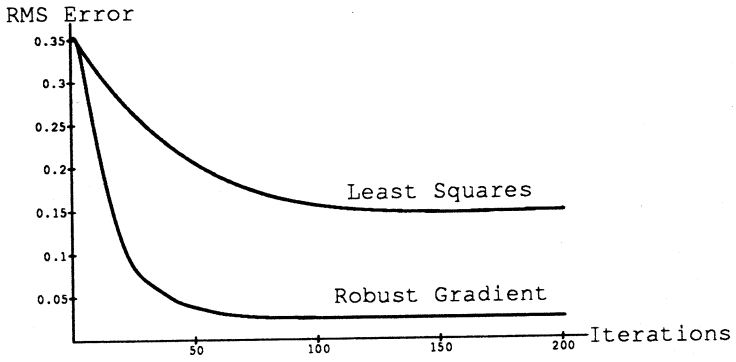


Figure 9: **Convergence.** Root mean squared error is plotted as a function of iterations for the standard Horn and Schunck scheme and the robust gradient scheme using simultaneous over-relaxation on the random dot sequence shown in figure 7. Notice that for an image of size 128×128 that SOR, with Chebyshev acceleration, converges in less than 128 iterations.

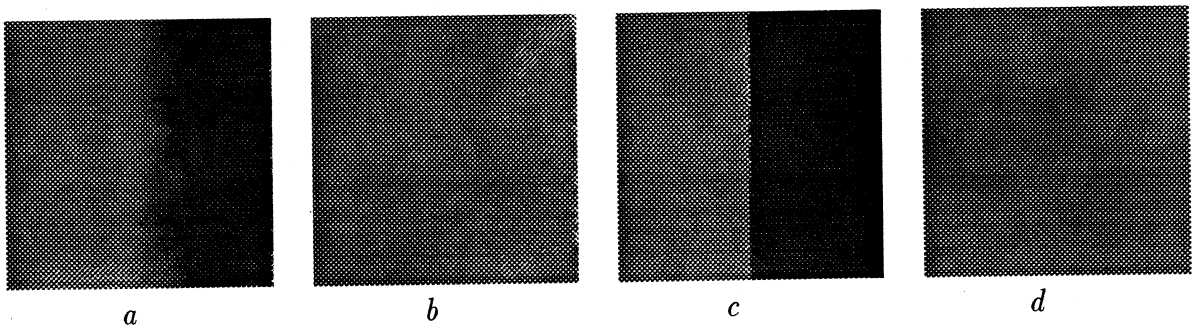


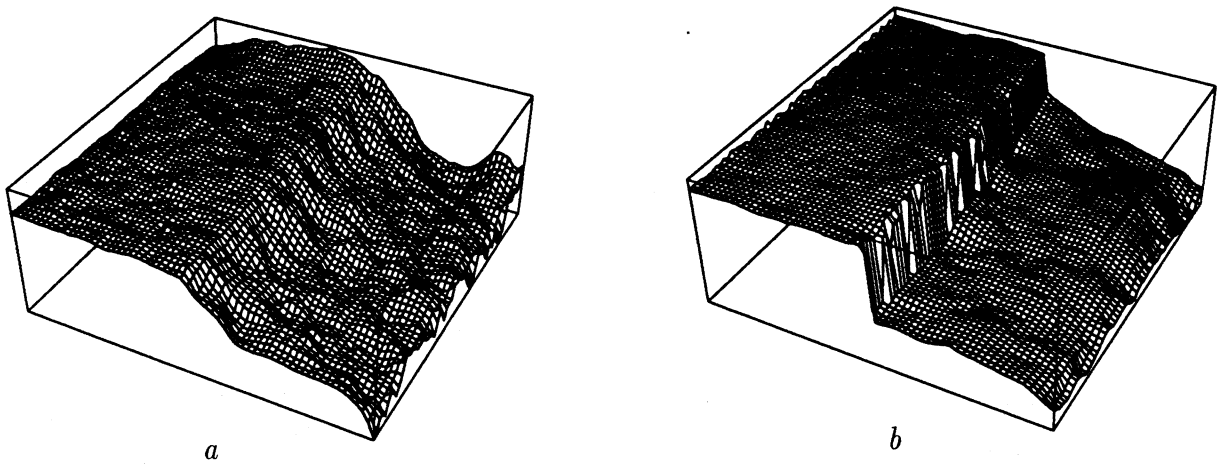
Figure 10: **Random Noise Sequence.** Computed flow in the case where 5 percent uniform noise is added to the second image. *a, b)* Horizontal and vertical least-squares flow. *c, d)* Horizontal and vertical robust flow.

constraint there is a tradeoff between smoothing the noise and over-smoothing the motion boundaries. The robust gradient approach allows us to smooth the data and preserve the discontinuities (figure 11*b*).

6.2 The Pepsi Sequence

Figure 12 shows the flow computed for a real image sequence. The sequence contains a Pepsi can (figure 12*a*) in front of a textured background. The camera is translating to the right, resulting in the can being displaced approximately one pixel to the left in each frame and the background being displaced by approximately a third of a pixel between frames. Figures 12*b* and 12*c* are the results of applying the least-squares algorithm. The flow is clearly smoothed across the motion boundary.

Figures 12 *d*, *e*, and *f* show the results of the incremental stochastic minimization algorithm of Black and Anandan [4]. The results are obtained by incrementally processing the 10 image sequence. Figure *e* shows that the algorithm does not smooth across motion boundaries and figure *d* shows that the motion boundaries are classified as occluding (white)



a *b*

Figure 11: **Horizontal Displacement (Noise added).** The horizontal component of motion is plotted for the case where 5 percent uniform noise is added to the second image. *a)* The least-squares solution; smoothing to reduce noise, smooths the motion boundary. *b)* The robust gradient approach smoothes the data while preserving the discontinuity.

or disoccluding (black).

The results of the robust gradient approach are shown if figures 12 *g*, *h*, and *i*. From figure *h* it is clear that the approach does an excellent job at preserving sharp motion discontinuities. The results appear better than the Black and Anandan results in that the motion boundaries correspond more closely to the surface boundary of the can. We believe that the poorer definition in the Black and Anandan results is due to the use of correlation, instead of spatial and temporal gradients, in the formulation of the data conservation constraint. This is due to the fact that correlation windows typically cover larger areas of the image and are hence more likely to be corrupted by multiple motions. Figure *g* shows the locations where the smoothness constraint is violated (ie. the motion discontinuity is greater than τ_2).

The least-squares and robust gradient solutions can be compared by examining the plots in figure 13. Here, the magnitude of the flow vectors is plotted. The sharp discontinuity present in the robust solution (figure *b*) is lacking in the least-squares estimate (figure *a*).

6.3 The Tree Sequence

Finally, we consider a more complex example with many discontinuities. The first image in the SRI tree sequence is seen in figure 14*a*. As expected, the least-squares flow estimate (figures *b* and *c*) shows a good deal of over-smoothing. The robust flow, shown in figures *e* and *f* exhibits sharp motion boundaries, yet still recovers the smoothly varying flow of the ground plane. Figure *d* shows the motion discontinuities where the outlier threshold is exceeded for the smoothness constraint. The results are noisier than in the Pepsi can example. This may in part be due the resolution of the images; the original images were smoothed and subsampled to be 116×128 pixels.

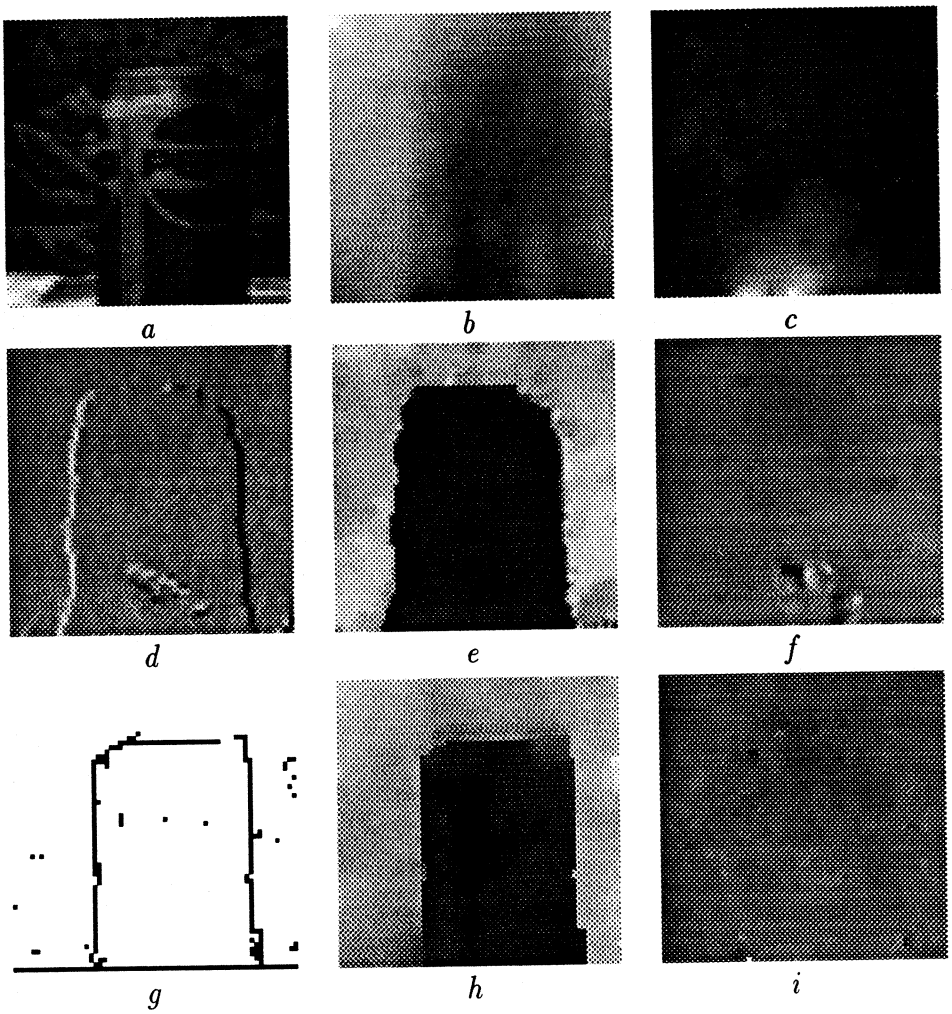
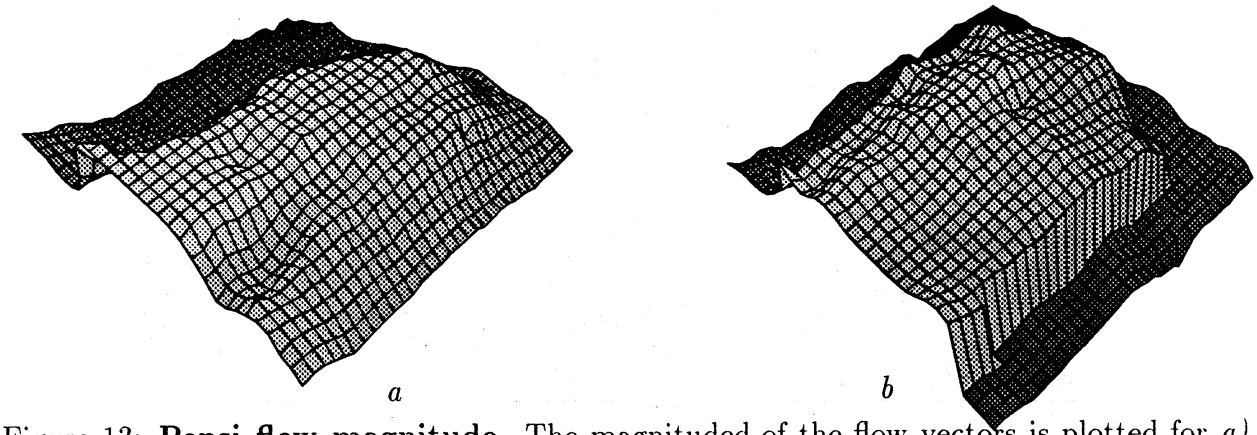


Figure 12: **The Pepsi Sequence.** *a)* The first intensity image. *b, c)* Horn and Schunck optical flow. *d, e, f)* Black and Anandan motion discontinuities and optical flow at the end of a 10 image sequence. *g, h, i)* Robust gradient results: flow discontinuities and optical flow.

6.4 Commentary

The idea proposed is a very simple one: take a least-squares problem and make it robust. Despite this simplicity, the experimental results indicate that the improvement in the flow field, particularly at discontinuities, is substantial. With the robust approach, the flow field can be smoothed without loosing the motion boundaries.

The only apparent drawback is the non-convex nature of the objective function. But good algorithms exist to cope with this minimization problem and our experiments indicate that, with SOR, we achieve faster convergence than standard Horn and Schunck implementation. Consequently, current applications of optic flow which rely on gradient based approaches, can be readily updated to incorporate this new robust scheme. The result will be more robust flow estimates as well as information about motion discontinuities.



a *b*

Figure 13: **Pepsi flow magnitude.** The magnituded of the flow vectors is plotted for *a)* the least-squares solution, *b)* the robust gradient solution.

7 Conclusion

We have presented a simple, robust, reformulation of the familiar least-squares optical flow equation of Horn and Schunck. The robust formulation takes into motion discontinuities and errors in the intensity constraint. A graduated non-convexity algorithm is presented for recovering the optical flow. The results are exciting. The new robust gradient formulation is less sensitive to image noise and does not smooth across motion discontinuities. In addition, the technique is easy to implement and the paper presents all the details necessary for such an implementation.

The robust gradient technique is currently being extended in a number of directions. First, to cope with motions larger than a single pixel, a hierarchical coarse-to-fine scheme is required [2]. Second, the minimization scheme presented can be made incremental by using the incremental minimization framework of Black and Anandan [4, 5]. This will improve performance by allowing the minimization to take place over the length of an image sequence. Additionally, alternative minimization schemes are being explored; particularly the idea of large-step Markov chains [16].

References

[1] J. Aloimonos, "Purposive and qualitative active vision," *Proc. Int. Conf. on Pattern Recognition,* Vol. 1, Atlantic City, NJ, June 1990, pp. 346–360.

[2] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *Int. Journal of Computer Vision*, 2, 1989, pp. 283–310.

[3] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proceedings of the IEEE*, Vol. 76, No. 8, August 1988, pp. 869–889.

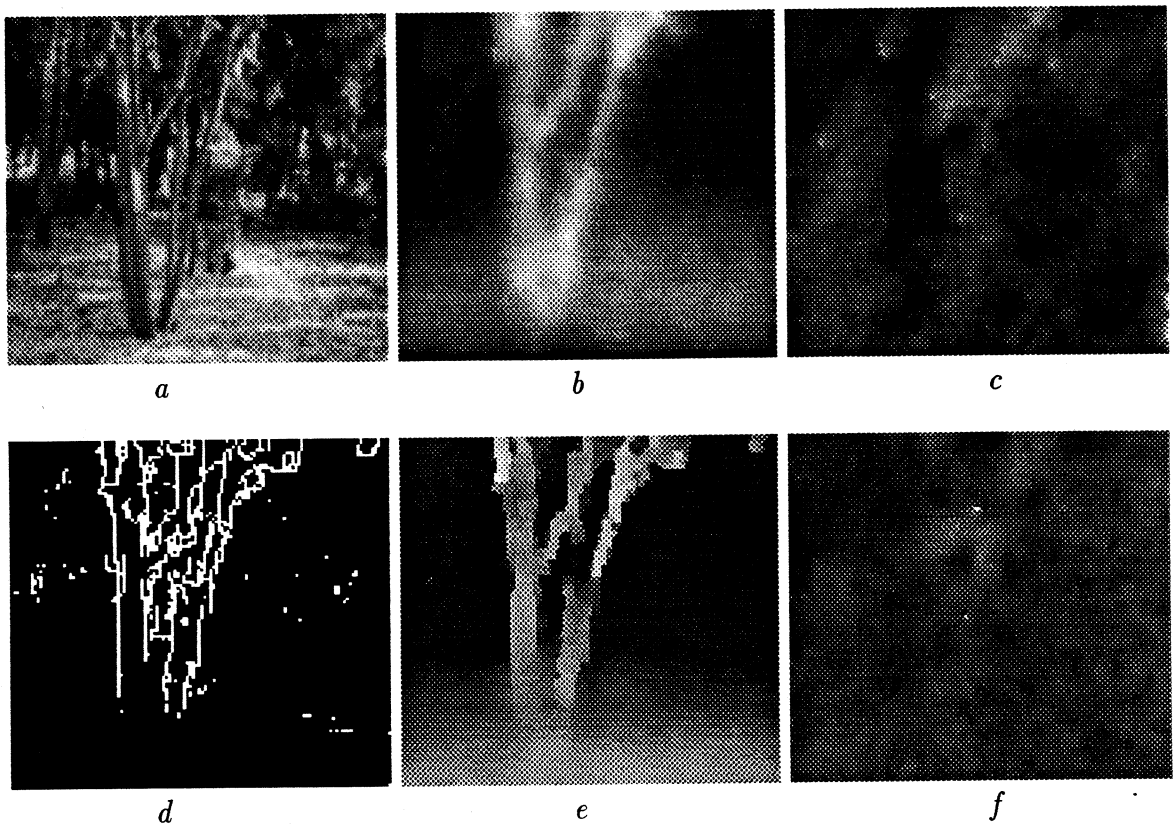


Figure 14: **The SRI Tree Sequence.** *a)* First intensity image. *b, c)* Least-squares flow. *d)* Flow discontinuities. *e, f)* Robust gradient flow.

[4] M. J. Black and P. Anandan, "Robust dynamic motion estimation over time," *Proc. Computer Vision and Pattern Recognition*, CVPR-91, Maui, Hawaii, June 1991, pp. 296–302.

[5] M. J. Black and P. Anandan, "A model for the detection of motion over time," *Proc. Int. Conf. on Computer Vision*, ICCV-90, Osaka, Japan, Dec. 1990, pp. 33–37.

[6] A. Blake and A. Zisserman, *Visual Reconstruction,* Cambridge, Massachusetts: The MIT Press, 1987.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-6, No. 6, November 1984, pp. 721–741.

[8] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions,* New York, NY: John Wiley and Sons, 1986.

[9] J. G. Harris, C. Koch, E. Staats, and J. Luo, "Analog hardware for detecting discontinuities in early vision," *Int. Journal of Comp. Vision,* IJCV, Vol. 4, No. 3, June 1990, pp. 211-223.

[10] B. K. P. Horn, *Robot Vision*, Cambridge, Massachusetts,: The MIT Press, 1986.

[11] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, Vol. 17, Nos. 1–3, August 1981, pp. 185-203.

[12] P. J. Huber, *Robust Statistics,* New York, NY: John Wiley and Sons, 1981.

[13] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science,* Vol. 220, No. 4598, May 1983, pp. 671–680.

[14] C. Koch, J. Luo, C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computer*, March, 1988, pp. 52–63.

[15] J. Konrad and E. Dubois, "Miltigrid Bayesian estimation of image motion fields using stochastic relaxation," *Int. Conf. on Computer Vision,* ICCV-88, pp. 354–362, 1988.

[16] O. Martin, S. W. Otto, and E. W. Felten, "Large-step Markov chains for the Traveling Salesman Problem," *Complex Systems,* Vol. 5, No. 3, pp. 299–326, 1991.

[17] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. PAMI-9, No. 2, March 1987, pp. 220–228.

[18] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, Vol. 317, No. 26, Sept. 1985, pp. 314–319.

[19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1988.

[20] B. G. Schunk, "Image flow segmentation and estimation by constraint line clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol 11, No. 10, Oct. 1989, pp. 1010–1027.

[21] A. Singh, "An estimation-theoretic framework for image-flow computation," *Proc. Int. Conf. on Comp. Vision*, ICCV-90, Osaka, Japan, Dec. 1990, pp. 168–177.

[22] G. Strang, *Linear Algebra and its Applications*, Academic Press, New York, 1976.

[23] R. S. Szeliski, "Bayesian modeling of uncertainty in low-level vision," Ph.D. Thesis, Carnegie Mellon University, 1988.

[24] R. S. Varga, *Matrix Iterative Analysis*, Prenctice-Hall, Inc, New Jersey, 1962.