A Framework for the Robust Estimation of Optical Flow

Michael J. Black [*]

Department of Computer Science
University of Toronto
Toronto, Ont, Canada M5S 1A4

P. Anandan

David Sarnoff Research Center
CN 5300
Princeton, NJ, USA 08543–5300

Abstract

We consider the problem of robustly estimating optical flow *from a pair of images using a new framework based on* robust estimation *which addresses violations of the brightness constancy and spatial smoothness assumptions. We also show the relationship between the* robust estimation framework *and line-process approaches for coping with spatial discontinuities. In doing so, we generalize the notion of a line process to that of an* outlier process *that can account for violations in both the brightness and smoothness assumptions. We develop a Graduated Non-Convexity algorithm for recovering optical flow and motion discontinuities and demonstrate the performance of the robust formulation on both synthetic data and natural images.*

1 Introduction

Algorithms for recovering optical flow embody a set of assumptions about the world which, by necessity, are simplifications and hence may be violated in practice. For example, the assumption of brightness constancy is violated when motion boundaries, shadows, or specular reflections are present. Motion boundaries also violate the common assumption that the optical flow varies smoothly. Violations such as these result in gross measurement errors which we refer to as *outliers*. To compute optical flow robustly we must reduce the sensitivity of the recovered optical flow to violations of the assumptions by detecting and rejecting outliers.

Many common solutions to the optical flow problem are formulated in terms of least-squares estimation which is well known to lack robustness in the presence of outliers. We show how a *robust statistical* formulation of these estimation problems makes the recovered flow field less sensitive to assumption violations. This robust formulation, combined with a deterministic optimization scheme, provides a framework for robustly estimating optical flow and allows assumption violations to be detected. We have applied the approach to three standard techniques for recovering optical flow [1]: area-based regression, correlation, and regularization techniques.

Previous work in optical flow estimation has focused on the violation of spatial smoothness at motion boundaries while ignoring violations of the brightness constancy assumption. Within the robust estimation framework, violations of both constraints are treated in a uniform manner and we will demonstrate that the "robustification" of the brightness constancy assumption greatly improves the flow estimates. The robust estimation framework is also closely related to "line-process" approaches for coping with spatial discontinuities [4]. We generalize the notion of a line process to that of an *outlier process* which can account for violations of both the brightness and smoothness assumptions.

2 Estimating Optical Flow

Most current techniques for recovering optical flow exploit two constraints on image motion: *data conservation* and *spatial coherence*. The data conservation constraint is derived from the observation that surfaces generally persist in time and, hence, the intensity structure of a small region in one image remains constant over time, although its position may change. The spatial coherence constraint embodies the assumption that surfaces have spatial extent and hence neighboring pixels in an image are likely to belong to the same surface. Since the motion of neighboring points on a smooth rigid surface changes gradually, we can enforce an implicit or explicit *smoothness constraint* on the motion of neighboring points in the image plane.

2.1 Data Conservation Constraint

Let $I(x, y, t)$ be the image intensity[1] at a point (x, y) at time t. The data conservation constraint can be expressed in terms of the standard *intensity constancy*

[*] This author 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] I may be a filtered version of the intensity image at time t.

assumption as follows:

$$I(x, y, t) \quad = \quad I(x + u\delta t, y + v\delta t, t + \delta t), \qquad (1)$$

where (u, v) is the horizontal and vertical image velocity at a point and δt is small.

From this we derive the data conservation constraint:

$$E_D(u, v) = \sum_{(x,y)\in\mathcal{R}} (I_x u + I_y v + I_t)^2. \qquad (2)$$

As the size of the region $\mathcal{R}$ tends to zero this error measure becomes the more familiar gradient-based constraint used in the Horn and Schunck algorithm [6] and the solution for (u, v) is underconstrained. A large region $\mathcal{R}$ is needed to sufficiently constrain the solution and provide some insensitivity to noise. The larger the region however, the less likely our assumptions about the motion will be valid over the entire region. For example, the constant velocity assumption used in E_D above will be violated by affine flow, transparency, motion boundaries, etc. The dilemma surrounding the appropriate size of $\mathcal{R}$ is referred to as the *generalized aperture problem.*

2.2 Spatial Coherence Constraint

When $\mathcal{R}$ is small, the solution for $\mathbf{u} = (u, v)$ may need to be further constrained by the addition of a spatial coherence assumption in the form of a regularizing term E_S; the objective function becomes:

$$E(\mathbf{u}) = E_D(\mathbf{u}) + \lambda E_S(\mathbf{u}), \qquad (3)$$

where λ controls the relative importance of the data conservation and spatial coherence terms. The most common formulation of E_S is the *first-order*, or *membrane*, model:

$$E_S(u, v) = u_x^2 + u_y^2 + v_x^2 + v_y^2, \qquad (4)$$

where the subscripts indicate partial derivatives of the flow in the x or y direction.

With this approach the local flow vector $\mathbf{u}_s$ is forced to be close to the average of its neighbors. When a motion discontinuity is present this results in smoothing across the boundary which reduces the accuracy of the flow field and obscures important structural information about the presence of an object boundary.

3 Robust Estimation

While much of the work in computer vision has focused on developing optimal strategies for exact parametric models, there is a growing realization that we must be able to cope with situations for which our models were not designed. This has resulted in a growing interest in the use of robust statistics in computer vision (see [7] for a discussion).

As identified by Hampel [5, page 11] the main goals

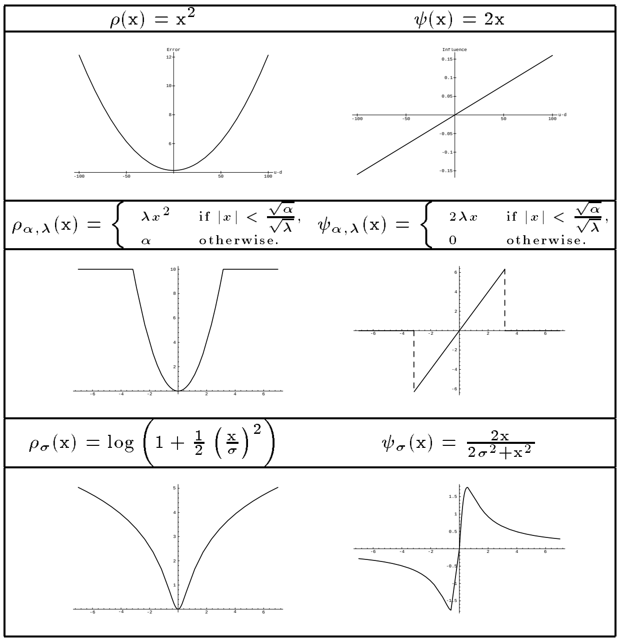


Figure 1: Example Estimators. Quadratic (top). Truncated quadratic (middle). Lorentzian (bottom).

of robust statistics are:

(i) *To describe the structure best fitting the bulk of the data.*

(ii) *To identify deviating data points (outliers) or deviating substructures for further treatment, if desired.*

Specifically, robust estimation addresses the problem of finding the values for the parameters, $\mathbf{a} = [a_0, \ldots, a_n]$, that best fit a model, $\mathbf{u}(s; \mathbf{a})$, to a set of data measurements, $\mathbf{d} = \{d_0, d_1, \ldots, d_S\}$, $s \in S$, in cases where the data differs statistically from the model assumptions. In fitting a model, the goal is to find the values for the parameters, $\mathbf{a}$, that minimize the size of the *residual* errors $(d_s - \mathbf{u}(s; \mathbf{a}))$:

$$\min_{\mathbf{a}} \sum_{s\in S} \rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s), \qquad (5)$$

where σ_s is a scale parameter, and ρ is our *estimator.* When the errors in the measurements are normally distributed, the optimal estimator is the quadratic:

$$\rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s) = \frac{(d_s - \mathbf{u}(s; \mathbf{a}))^2}{2\sigma_s^2}, \qquad (6)$$

which gives rise to the standard *least-squares* estimation problem. The function ρ is called an *M-estimator* since it corresponds to the *Maximum-likelihood* estimate. The *robustness* of a particular estimator refers to its insensitivity to outliers, or deviations, from the assumed statistical model.

The problem with the least-squares solution is that the outlying points are assigned a high weight by the quadratic estimator (Figure 1, top left). One way to see this is by considering the *influence function* associated with a particular estimator. This function characterizes the bias that a particular measurement has on the solution and is determined by the derivative, ψ, of the estimator [5]. In the least-squares case, the influence of data points increases linearly and without bound (Figure 1, top right).

To increase robustness we will consider estimators for which the influence of outliers tends to zero. Many of these *redescending* M-estimators have been studied in robust statistics, but one of the most common in computer vision is the truncated quadratic [2] (Figure 1, middle). Up to a fixed threshold, errors are weighted quadratically, but beyond that, errors receive a constant value. By examining the ψ-function we see that the influence of outliers goes to zero beyond the threshold. For the remainder of the paper we will consider the Lorentzian estimator (Figure 1, bottom), but the treatment here could equally be applied to a wide variety of the other estimators. A discussion of various estimators can be found in [1].

3.1 Robust Estimation Framework

We make the simple observation that may common approaches to recovering optical flow are formulated as least-squares estimation (including: correlation, regularization, and area-base techniques). Because each approach involves pooling information over a spatial neighborhood these least-squares formulations are inappropriate at motion boundaries. By treating the problems in terms of robust estimation, we alleviate the problems of oversmoothing and noise sensitivity typically associated with the least-squares formulations.

To improve the robustness, we reformulate our minimization problems to account for outliers by using the robust estimators described above. We illustrate by considering a simple gradient-based formulation of optical flow [6]. For an image of size $m \times m$ pixels we define a grid of *sites:*

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

where $(i(s), j(s))$ denotes the pixel coordinates of site s. The objective function, $E(\mathbf{u})$, for the regularization approach, becomes:

$$\sum_{s\in S}[\sum_{\mathcal{R}} \rho_1(I_x u + I_y v + I_t, \sigma_1) \\ +\lambda \sum_{n\in\mathcal{G}_s} [\rho_2(u_s - u_n, \sigma_2) + \rho_2(v_s - v_n, \sigma_2)]], \quad (7)$$

where $\mathcal{G}_s$ represents the set of north, south, east, west neighbors of s in the grid, where σ_1 and σ_2 are scale parameters and where ρ_1 and ρ_2 may be different estimators. Rather than choosing $\rho_i(x) = x^2$, which gives the familiar least-squares formulation, we take the ρ_i to be robust estimators.

4 Relationship to Line Processes

We now examine how the robust estimation approach relates to line-process approaches in which first-order discontinuities in the flow are modeled by binary valued line processes $l_{s,n}$ which represent the presence, or absence, of a discontinuity between sites s and n [2, 4].[2] The new objective function, $E(\mathbf{u}, l)$, is then:

$$\sum_{s\in S}[(I_x u_s + I_y v_s + I_t)^2 \\ +\lambda \sum_{n\in\mathcal{G}_s} [\alpha_S(1 - l_{s,n})\|\mathbf{u}_s - \mathbf{u}_n\|^2 + \beta_S l_{s,n}]], \quad (8)$$

where α_S and β_S are constant factors controlling the weighting of the smoothness term and the "penalty term" respectively. Such a formulation allows violations of the spatial smoothness term, but does not account for violations of the data term. This prompts us to generalize the notion of a "line process" to that of an "outlier process" that can be applied to both data and spatial terms to perform *outlier rejection* in the same spirit as the robust estimators do. The objective function, $E(\mathbf{u}, l, d)$, is then reformulated as:

$$\sum_{s\in S}[\alpha_D(1 - d_s)(I_x u_s + I_y v_s + I_t)^2 + \beta_D d_s \\ +\lambda \sum_{n\in\mathcal{G}_s} [\alpha_S(1 - l_{s,n})\|\mathbf{u}_s - \mathbf{u}_n\|^2 + \beta_S l_{s,n}]], \quad (9)$$

where we have simply introduced a new process d_s and constant scaling factors α_D and β_D.

From Outlier Process to Robust Estimation: Blake and Zisserman [2] showed that line processes can be eliminated from the objective function by first minimizing over them, resulting in an objective function which is solely a function ρ of the actual variables under consideration. Exactly the same treatment can be applied to the outlier-process formulation to derive [1]:

$$\min_{\mathbf{u}} \sum_{s\in S}[\rho((I_x u_s + I_y v_s + I_t), \alpha_D, \beta_D) \\ +\lambda \sum_{n\in\mathcal{G}_s} \rho(\|\mathbf{u}_s - \mathbf{u}_n\|, \alpha_S, \beta_S)], \quad (10)$$

[2] For illustration, we consider a gradient-based formulation with a first order smoothness term applied to the norm of the local flow difference.

where, in the case of binary line processes, ρ is the truncated quadratic shown in Figure 1.[3] Notice that this is identical to a robust estimation formulation with the truncated quadratic as the estimator.

From Robust Estimators to Line Processes: For certain choices of robust estimators, we can convert a robust estimation problem into an equivalent problem involving binary or analog outlier processes (for a detailed treatment see [1, 8]). This allows spatial interactions between line processes to be explicitly modeled. Take, for example, a robust formulation of optical flow, $E(u, v)$, where ρ is the Lorentzian estimator:

$$\sum_{s\in S}[\rho(I_x u_s + I_y v_s + I_t, \sigma_1) + \lambda \sum_{n\in\mathcal{G}_s} \rho(\|\mathbf{u}_s - \mathbf{u}_n\|, \sigma_2)].$$

We can derive an equivalent cost function, $E(\mathbf{u}, d, l)$, containing analog line process, $z(x)$, [1, 8]:

$$\sum_{s\in S}[(1 - z(d))(I_x u + I_y v + I_t)^2 + P(d) + \lambda \sum_{n\in\mathcal{G}_s}[(1 - z(l))\|\mathbf{u}_s - \mathbf{u}_n\|^2 + P(l)]],$$

where $P(x)$ is a "penalty" term and $d, l \geq 0$. In the case of the Lorentzian estimator the outlier process $z(x)$ is defined as:

$$z(x) = 1 - \frac{1}{1+x}.$$

5 A Robust Gradient Method

We take as an example the robust gradient-based formulation in equation (7) with the Lorentzian as the estimator. Unlike the least-squares formulation, the robust objective function, $E(u, v)$, may be non-convex. A local minimum, however, can be found using Simultaneous Over-Relaxation (SOR). The iterative update equation for minimizing E at step $n+1$ is simply [2][4]:

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

where $0 < \omega < 2$ is an *overrelaxation parameter* that is used to *overcorrect* the estimate of $u^{(n+1)}$ at stage $n+1$. The first partial derivative of the robust flow equation (7) is simply:

$$\frac{\partial E}{\partial u_s} = \sum_{s\in S}[\lambda I_x \psi(I_x u + I_y v + I_t, \sigma_1) + \sum_{n\in\mathcal{G}_s} \psi(u_s - u_n, \sigma_2)],$$

where $\psi(x) = \partial\rho/\partial x$.

The term $T(u_s)$ is an upper bound on the second partial derivatives of E which implies:

$$T(u_s) = \frac{\lambda I_x^2}{\sigma_1^2} + \frac{4}{\sigma_2^2} \geq \frac{\partial^2 E}{\partial u_s^2}, \quad \forall s \in S.$$

5.1 Graduated Non-Convexity

We now turn to the problem of finding a globally optimal solution when the function is non-convex. We can construct a convex objective function by choosing σ_1 and σ_2 to be sufficiently large so that the Hessian matrix of E at each point in the image is positive definite. These σ_i determine the point at which measurements are considered outliers; that is, the point at which the influence of the measurements begins to decrease. This occurs when the derivative of $\psi(x)$ equals zero or $x = \pm\sqrt{2}\,\sigma$. For the convex approximation we take $\sigma_i = \tau_i/\sqrt{2}$, where τ_i is the largest expected outlier. The minimum of this convex formulation is readily obtained using SOR.

We use the *Graduated Non-Convexity (GNC)* continuation method of Blake and Zisserman [2] to track the minimum over a sequence of objective functions with decreasing values for the σ_i which gradually introduce discontinuities in the data and spatial terms. The SOR algorithm is used to converge to the minimum for each new value of σ_i. The minimum values for the σ_i are determined from prior expectations of motion discontinuities and sensor noise.[5]

6 Experimental Results

We have conducted a number of experiments using synthetic and natural image sequences to compare the performance of the least-squares and robust formulations of the optical flow equation (7). All experiments were performed using 200 iterations[6] of each algorithm. The parameter λ was empirically determined and remained unchanged for all the experiments: $\lambda = 10$ for the robust-gradient approach, and $\lambda = 50$ for the least-squares approach[7]. The spatial and temporal derivatives (I_x, I_y, I_t) were estimated using simple image differencing and the images were prefiltered with a Laplacian.

6.1 Synthetic Sequence

The first experiment involves a synthetic sequence containing two textured surfaces, one which is stationary

[3] Geman and Reynolds [3] showed that this approach can be generalized to *analog* line processes that assume continuous non-negative values.

[4] Only the equations for the horizontal component of the flow are show; the treatment of the vertical component is identical.

[5] A coarse-to-fine strategy for coping with large motions is described in [1].

[6] An iteration involves the updating of every site in the flow field.

[7] The different values of λ are due to the different ρ functions used; that is, the quadratic for the least-squares approach, and the Lorentzian for the robust-gradient method.

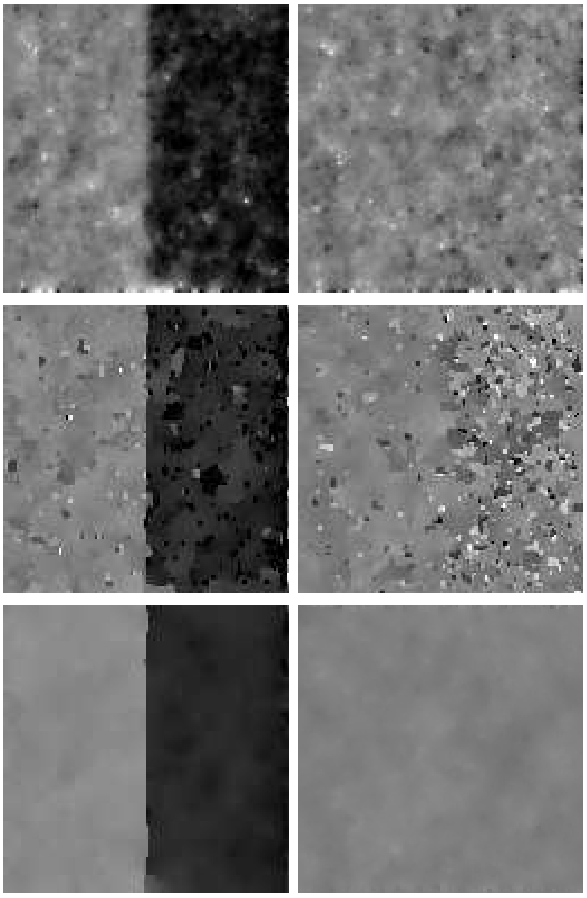


Figure 2: **Effect of robust data term,** (10% uniform noise). *(top)* Least-squares. *(middle)* Quadratic data and robust smoothness. *(bottom)* Robust formulation.

and one which is translating one pixel to the left. The second image in the sequence has been corrupted with 10% uniform random noise. To evaluate the effect of the robust formulation of the data and smoothness terms, we compare the performance of three different formulations: least-squares (Horn and Schunck), a version with a quadratic data term and robust smoothness term (eg. Blake and Zisserman), and the fully robust formulation.

The results are illustrated in Figure 2. The left column shows the horizontal motion and the right column shows the vertical motion recovered by each of the approaches (black $= -1$ pixel, white $= 1$ pixel, gray $= 0$ pixels). Figure 2 (top) shows the noisy, but smooth, results obtained by least-squares. Figures 2 (middle) shows the result of introducing a robust smoothness term alone; the recovered flow is piecewise smooth, but the gross errors in the data produce spurious motion discontinuities. Finally, Figure 2 (bottom) shows the improvement realized when both the data and spatial terms are robust.

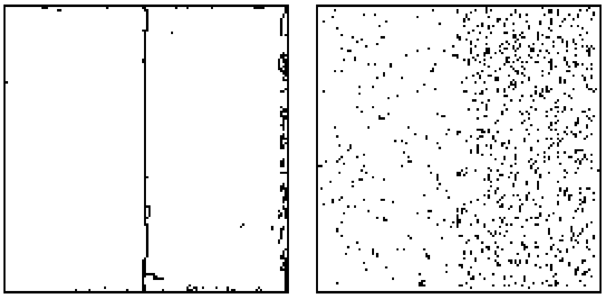


Figure 3: **Outliers in the smoothness and data terms,** (10% uniform noise). *a)* Flow discontinuities. *b)* Data outliers.

We can detect outliers where the final values of the data coherence and spatial smoothness terms are greater than the outlier thresholds $\sqrt{2}\sigma_1$ and $\sqrt{2}\sigma_2$. Motion discontinuities are simply outliers with respect to spatial smoothness (Figure 3*a*). A large number of image measurements are treated as outliers by the data term; especially when the motion is large (Figure 3*b*).

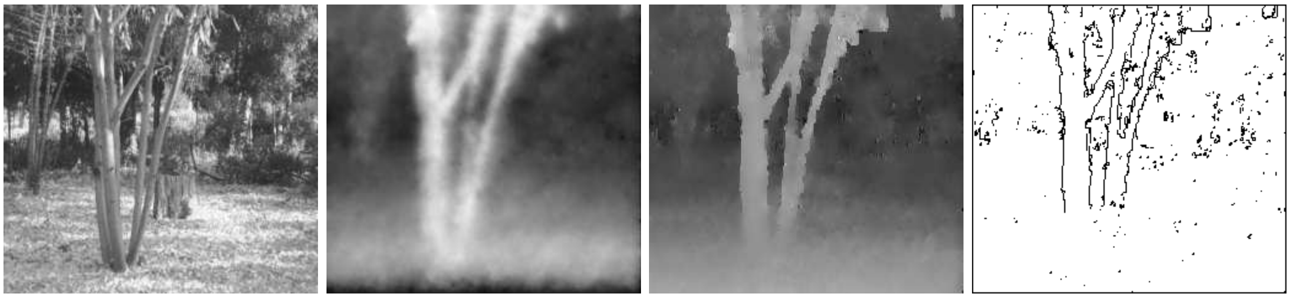
6.2 The Pepsi Sequence

We next consider a natural image sequence in which a Pepsi can and textured background move approximately 0.8 and 0.35 pixels to the left between frames respectively (Figure 4, top left). Figure 4 (bottom) shows that the flow recovered with the robust formulation does an excellent job of preserving sharp motion discontinuities. Figure 4 (top right) shows the locations where the smoothness constraint is violated (ie. the change in flow across the boundary is treated as an outlier). The boundaries correspond well to the physical boundaries of the can.

6.3 The Tree Sequence

Finally, we consider a more complex example with many discontinuities and motion greater than a pixel. The first 233×256 image in the SRI tree sequence is seen in Figure 5*a*. As expected, the least-squares flow estimate (Figure 5*b*) suffers from over-smoothing.[8] The robust flow, shown in Figure 5*c* exhibits sharp motion boundaries, yet still recovers the smoothly varying flow of the ground plane. Figure 5*d* shows the motion discontinuities where the outlier threshold is exceeded for the smoothness constraint.

[8] Only the horizontal component of the flow is shown.



a *b* *c* *d*

Figure 5: **The SRI Tree Sequence.** *a)* First intensity image. *b)* Least-squares (horizontal component). *c)* Robust gradient. *d)* Spatial outliers.

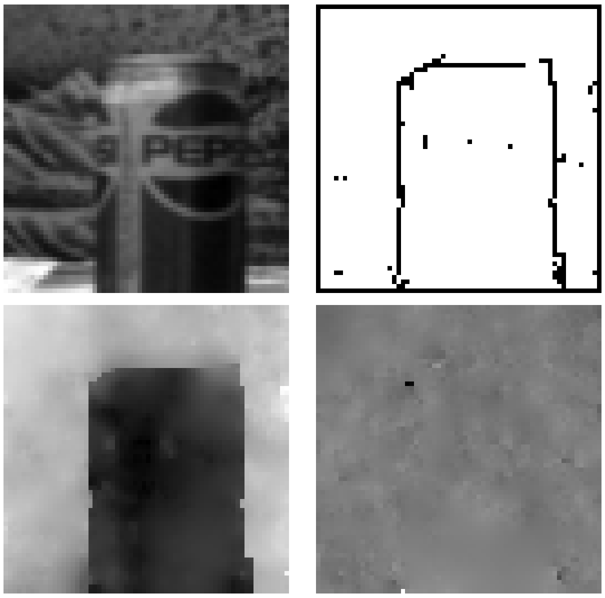


Figure 4: **The Pepsi Sequence.** Image 1 (top left); Robust flow field (u, v) (bottom left and right respectively); Spatial smoothness outliers (top right).

7 Conclusion

This paper has considered the issues of robustness related to the recovery of optical flow with motion discontinuities. In this regard, it is important to recognize the generality of the problems posed by motion discontinuities; measurements are corrupted whenever information is pooled from a spatial neighborhood which spans a motion boundary. This applies to both the data conservation and spatial coherence assumptions. These violations of the constraints cause problems for the standard least-squares formulations of optical flow. By recasting these formulations within our robust estimation framework, erroneous measurements at motion boundaries are treated as outliers and their influence is reduced.

Finally, it should be noted that the robust estimation framework has more general applicability than the recovery of optical flow. It provides a general framework for dealing with model violations which can be applied to a wide class of problems in early vision.

Acknowedgements

We thank D. Heeger, A. Rangarajan, G. Hager, S. Engelson, and J. MacLean for reading and commenting on this work.

References

[1] M. J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale Univeriesty, New Haven, CT, 1992. Research Report YALEU/DCS/RR-923.

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

[3] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):376–383, March 1992.

[4] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

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

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

[7] P. Meer, D. Mintz, and A. Rosenfeld. Robust regression methods for computer vision: A review. *Int. J. of Computer Vision*, 6(1):59–70, 1991.

[8] A. Rangarajan and R. Chellapa. A continuation method for image estimation and segmentation. Tech. Report CAR–TR–586, Univ. of Maryland, Oct. 1991.