ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/220135301

The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields

Article *in* Computer Vision and Image Understanding · January 1996

DOI: 10.1006/cviu.1996.0006 · Source: DBLP

CITATIONS
1,601

READS
1,473

2 authors:

Michael J Black
Max Planck Institute for Intelligent Systems
434 PUBLICATIONS **40,353** CITATIONS

SEE PROFILE

Padmanabhan Anandan
Microsoft
60 PUBLICATIONS **7,424** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Multimodal Motion Capture View project

Project Image Analytics - HPNN View project

All content following this page was uploaded by Michael J Black on 04 November 2017.

The user has requested enhancement of the downloaded file.

COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 63, No. 1, January, pp. 75–104, 1996
ARTICLE NO. 0006

The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields

MICHAEL J. BLACK*

Xerox Palo Alto Research Center, Palo Alto, California 94304

AND

P. ANANDAN

David Sarnoff Research Center, Princeton, New Jersey 08543-5300

Received December 27, 1993; accepted December 5, 1994

Most approaches for estimating optical flow assume that, within a finite image region, only a single motion is present. This *single motion assumption* is violated in common situations involving transparency, depth discontinuities, independently moving objects, shadows, and specular reflections. To robustly estimate optical flow, the single motion assumption must be relaxed. This paper presents a framework based on *robust estimation* that addresses violations of the brightness constancy and spatial smoothness assumptions caused by multiple motions. We show how the *robust estimation framework* can be applied to standard formulations of the optical flow problem thus reducing their sensitivity to violations of their underlying assumptions. The approach has been applied to three standard techniques for recovering optical flow: area-based regression, correlation, and regularization with motion discontinuities. This paper focuses on the recovery of multiple parametric motion models within a region, as well as the recovery of piecewise-smooth flow fields, and provides examples with natural and synthetic image sequences. © 1996 Academic Press, Inc.

1. INTRODUCTION

When estimating 2D image velocity, or optical flow, from image sequences, it is common to assume that only a single motion is present within any finite image region. This *single motion assumption* appears in a variety of forms. For example, the standard *data conservation constraint* assumes that the image brightness of a region remains constant while its location may change. This assumption underlies the common correlation and gradient-based approaches and is violated when a region contains transparency, specular reflections, shadows, or fragmented occlusion (for example, looking through the branches of a tree). The assumption may also be violated when the region spans a depth discontinuity in the scene. The standard *spatial coherence constraint* suffers from the same problem. It assumes that the flow within a neighborhood changes gradually since it is caused by a single motion. Optical flow, however, is typically only piecewise smooth since depth boundaries in the scene may give rise to discontinuities in the flow field as a result of motion parallax or the independent motion of objects. In these common situations, most current approaches cannot recover the multiple motions, give erroneous motion estimates for the region, and fail to signal that the underlying assumption of a single motion has been violated.

The fundamental problem is that the estimation of optical flow involves the pooling of constraints over some spatial neighborhood. To gain accurate estimates this region must be sufficiently large to constrain the solution (the standard aperture problem), yet the larger the region of integration the more likely it is to contain multiple motions with competing constraints. This has been referred to as the *generalized aperture problem* [30]. The pooling of constraints over a neighborhood is typically performed using least-squares estimation. These least-squares formulations average the motions present in a region which results in incorrect motion estimates from the data conservation constraint and over-smoothing by the spatial coherence constraint.

When, for example, two motions are present in a neighborhood, one set of constraints will be consistent with one of the motions and another set will be consistent with the other motion. When examining one of the motions, the constraints for the other motion will appear as gross errors which we refer to as *outliers*. Least-squares estimation is

* To whom correspondence should be addressed. E-mail: black@parc.xerox.com.

1077-3142/96 $18.00
Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

well known to lack "robustness" in the presence of outliers. To compute optical flow robustly we must reduce the sensitivity of the recovered optical flow to violations of the single motion assumption by detecting and rejecting outliers. This paper shows how a *robust statistical* formulation of these estimation problems makes the recovered flow field less sensitive to assumption violations and allows us to accurately recover multiple motions. The main idea is to reformulate the least-squares estimation problems using *robust estimation* techniques which are less sensitive to outliers. 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 framework to a variety of standard techniques for recovering optical flow; these include correlation [7], area-based regression [6], correlation with regularization [9], and gradient-based approaches with regularization [6, 10].

Previous work in optical flow estimation has focused on the violation of the spatial coherence assumption at motion boundaries [14, 23, 38, 40, 48, 50, 53] while ignoring violations of the data conservation, or 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 [6, 11, 19]. By generalizing the notion of a binary line process to that of an analog *outlier process* which can account for violations of both the brightness and smoothness assumptions the outlier-process formulation can be converted into a robust-estimation problem and *vice versa* for a particular class of robust estimation problems [11].

The following section provides a review of optical flow and multiple motions. Examples are provided to illustrate the application of a robust approach. Section 3 provides a brief introduction to the topic of robust estimation. In Section 4, a robust estimation framework is developed and applied to a number of common optical flow formulations. For purposes of exposition, we will develop two algorithms in detail by applying the robust estimation framework. The first regression-based approach is described in Section 5 and applied to problems involving motion boundaries, fragmented occlusion, transparency, and specular reflections. Section 6 describes a regularization technique which uses a robust version of the standard optical flow constraint equation and a robust first-order smoothness term. The approach is applied to a number of synthetic and natural image sequences with piecewise-smooth flow fields to illustrate the effects of the robust data and smoothness terms.

2. MULTIPLE MOTIONS AND THE ESTIMATION OF 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. In many commonly occurring situations, this assumption is violated for some subset of the points within the image region; for example, it is violated at motion boundaries and when specular reflections are present.

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. It has long been realized that such a constraint is violated at surface boundaries and much of the recent work in motion estimation has focused on where it is violated and how to reformulate it.

Below we review the standard formulations of the two constraints, illustrate the problems caused by the single motion assumption, and review previous approaches for dealing with multiple motions.

2.1. Data Conservation

Let $I(x, y, t)$ be the image brightness, or a filtered version of the image brightness, at a point (x, y) at time t. The data conservation constraint can be expressed in terms of the standard *brightness constancy assumption* as follows:

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

Here (u, v) is the horizontal and vertical image velocity at a point and δt is small. This simply states that the image value at time t, at a point (x, y), is the same as the value in a later image at a location offset by the optical flow.

Correlation Methods. The most direct way to use the brightness constancy assumption is to formulate the data conservation error measure using sum-of-squared difference (SSD) correlation [2]. In this formulation, the image velocity is assumed to be approximately constant within a local neighborhood, $\mathscr{R}$, and the error associated with a given displacement is formulated as

$$E_{\mathrm{D}}(u, v) = \sum_{(x,y)\in\mathscr{R}} [I(x, y, t) - I(x + u\delta t, y + v\delta t, t + \delta t)]^2. \tag{2}$$

Gradient Methods—Locally Constant Flow. We can rewrite (2) by taking the Taylor Series approximation of $I(x + u\delta t, y + v\delta t, t + \delta t)$. Dropping the terms above first order and simplifying gives the following gradient-based formulation:

$$E_{\mathrm{D}}(u, v) = \sum_{(x,y)\in\mathcal{R}} (I_x(x, y, t)u + I_y(x, y, t)v + I_t(x, y, t))^2. \tag{3}$$

Here the subscripts indicate partial derivatives of the brightness function with respect to x, y, and t.

We find it convenient to write this data conservation constraint as

$$E_{\mathrm{D}}(\mathbf{u}) = \sum_{\mathcal{R}} ((\nabla I)^{\mathrm{T}}\,\mathbf{u} + I_t)^2, \tag{4}$$

where ∇I denotes the local brightness gradient vector, and $\mathbf{u} = [u, v]^{\mathrm{T}}$ denotes the flow vector [34]. As the region size tends to zero (i.e., if the estimation is restricted to information at a point in the image) this error measure becomes the gradient-based constraint used in the Horn and Schunck algorithm [27].

Regression Methods—Affine Flow. In order to allow larger spatial neighborhoods to be used while still retaining the advantage of a closed form estimation of the local flow vectors, the flow field can be modeled as a parametric function of the image coordinates. Assuming a model $\mathbf{u}(x, y; \mathbf{a})$ of the flow within a region, where $\mathbf{a}$ are the parameters of the model, we can combine information from neighboring gradient constraint equations to find the parameters $\mathbf{a}$ that minimize the sum of the constraints over a neighborhood $\mathcal{R}$:

$$E_{\mathrm{D}}(\mathbf{a}) = \sum_{\mathcal{R}} ((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t)^2. \tag{5}$$

Common models of image flow in a region $\mathcal{R}$ include constant, affine, and quadratic. For an affine flow model, we have

$$\mathbf{u}(x, y; \mathbf{a}) = \begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} a_0 + a_1x + a_2y \\ a_3 + a_4x + a_5y \end{bmatrix}. \tag{6}$$

Parametric models such as this have been used for estimating image motion over the entire image or a preselected region [4, 13, 16, 17, 31, 34, 58].

2.1.1. *The Generalized Aperture Problem*

The above formulations share a common problem: the choice of the region size for $\mathcal{R}$. A *large* region $\mathcal{R}$ is needed to sufficiently constrain the solution and provide insensitivity to noise; this is commonly referred to as the *aperture problem*. The larger the region however, the less likely it is that our assumptions about the motion will be valid over the entire region. For example, as the region size grows so does the likelihood that the region will contain multiple motions. When this occurs, there will be no single flow $\mathbf{u}(\mathbf{a})$ which produces small residuals $((\nabla I)^{\mathrm{T}}\,\mathbf{u}(\mathbf{a}) + I_t)^2$ over the entire region. When solving for one motion the large residuals corresponding to competing motions can be viewed as outliers. Additionally, the approaches above make assumptions about the spatial variation of the motion within the region; for example, that it is constant or affine. These motion models are only good approximations locally. For these reasons the region should be *small*. We refer to the dilemma surrounding the appropriate size of aperture, $\mathcal{R}$, as the *generalized aperture problem*[1] (or GAP) [30]:

1. $\mathcal{R}$ must be *large* to sufficiently constrain the solution,
2. $\mathcal{R}$ must be *small* to avoid violating the assumptions.

2.1.2. *Multiple Motions: An Example*

To illustrate the problems posed by multiple motions we will briefly consider an example in which we want to estimate the motion in Fig. 1a. In this image sequence, the camera is stationary and a person is moving behind a plant resulting in "fragmented occlusion." The motion of the plant and the person can each be well approximated by a constant global translation (the horizontal motion of the plant and background is zero and the person is translating approximately −0.79 pixels). Consider what happens when we apply the least-squares regression approach to this sequence (where $\mathcal{R}$ is the entire image); that is, we minimize

$$E_{\mathrm{D}}(u, v) = \sum_{\mathcal{R}} (I_xu + I_yv + I_t)^2. \tag{7}$$

Each constraint $(I_xu + I_yv + I_t)$ forms a line in u-v-space. Figure 1b shows a small fraction of the constraint lines obtained for the image sequence. In Fig. 1b, two distinct populations of constraint lines are visible corresponding to the two motions present in the region. The solution obtained by least-squares regression attempts to satisfy all these constraints simultaneously. This results in an erroneous motion estimate $((u, v) = (-0.19, 0.02))$ which lies between the two motions. The white "cross" marks the location of the least-squares estimate, while the gray crosses mark the two "correct" motions.

The robust estimation framework proposed here produces a different result. The approach allows the dominant

[1] Negahdaripour and Yu [41] use the same term to refer to the ill-conditioning produced by their generalized brightness-change model.

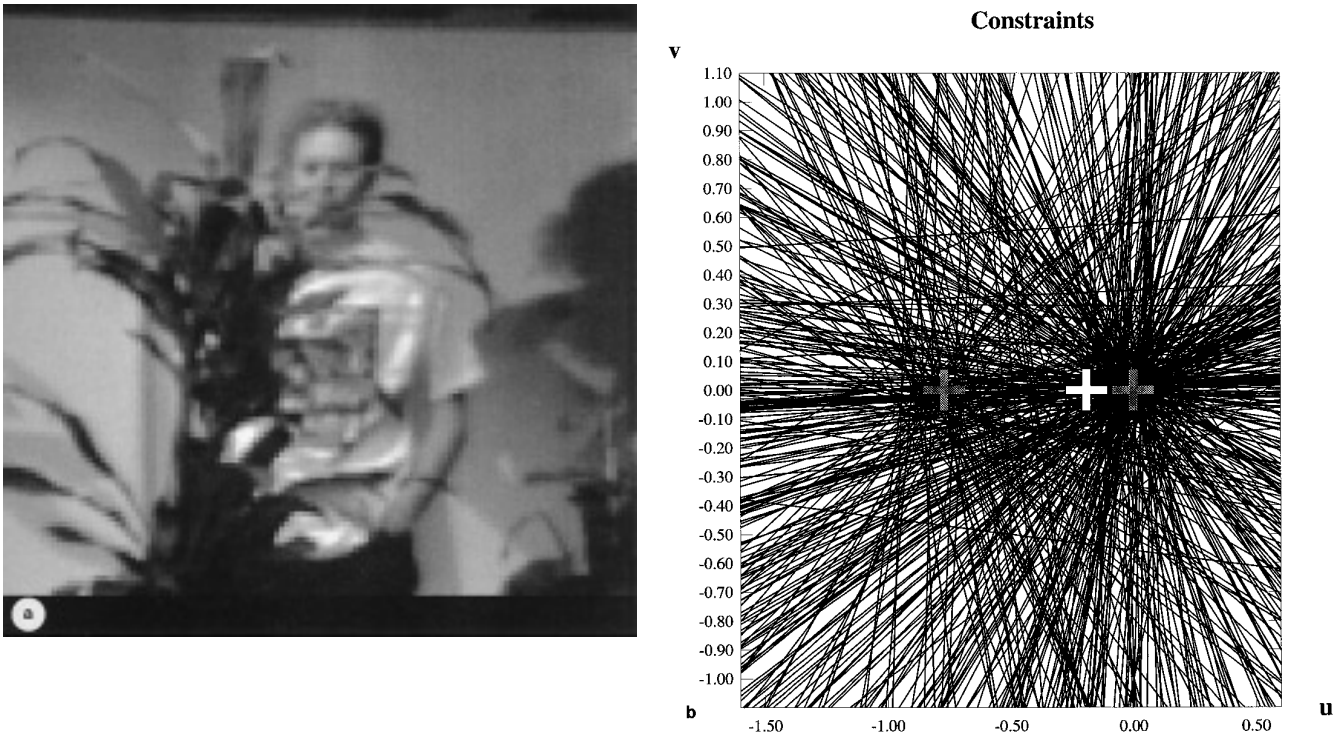

FIG. 1. Image sequence with fragmented occlusion. (a) First image. (b) Constraint lines (see text).

motion (of the plant and the background) to be accurately estimated while "ignoring" the motion of the person. Additionally the approach detects where the single motion assumption is violated (i.e., where $|((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t)|$ is large). These violations (or "outliers") appear as dark regions in Fig. 2a and correspond to the moving person and his shadow. The constraint lines associated with the dominant motion are shown in Fig. 2b with the recovered motion ($(u, v) = (-0.03, 0.00)$) marked with a white cross. These constraint lines are derived from the white areas in Fig. 2a.

We can now examine the constraints that were rejected as outliers to see if they correspond to a consistent motion. The regions consistent with a second motion and the associated constraint lines are shown in Fig. 3, and the recovered motion for the person ($(u, v) = (-0.77, -0.03)$) is marked by a white cross.

2.1.3. *Previous Approaches*

There have been a number of previous attempts to improve accuracy and robustness of flow estimates in regions containing multiple motions by relaxing the single motion assumption. Shulman and Hervé [50] point out that the standard brightness constancy assumption is commonly violated. While they use robust statistical techniques to cope with spatial discontinuities, they do not apply robust techniques to the data term. Instead, they propose a more general model of image brightness variation.

Bergen *et al.* [5] propose an area-regression approach for estimating two motions from three frames. The approach uses an iterative algorithm to estimate one motion, performs a nulling operation to remove the intensity pattern giving rise to the motion, and then solves for the second motion. The process is repeated and the motion estimates are refined. In contrast, the robust approach can theoretically recover n motions from *two* frames, by directly fitting multiple parametric models to the constraint equations.

Using standard least-squares regression, it is sometimes possible to detect multiple motions by examining the residual of the least-squares solution [29]. After obtaining an initial least-squares estimate, "outliers" are detected and removed. The least-squares solution is then recomputed and the process iterates. This type of robust process is sensitive to the initial quadratic estimate which may be arbitrarily bad.

In considering a correlation-based approach, Okutomi and Kanade [42] develop an "adaptive window" technique that adjusts the size of the correlation region to minimize the uncertainty in the estimate. Their implementation of the approach is limited by the use of a fixed shape (rectangular) window that cannot adapt to irregular surface boundaries. The approach also cannot cope with fragmented occlusion (for example, trees or fences) where, regardless of window size or shape, multiple motions are present.

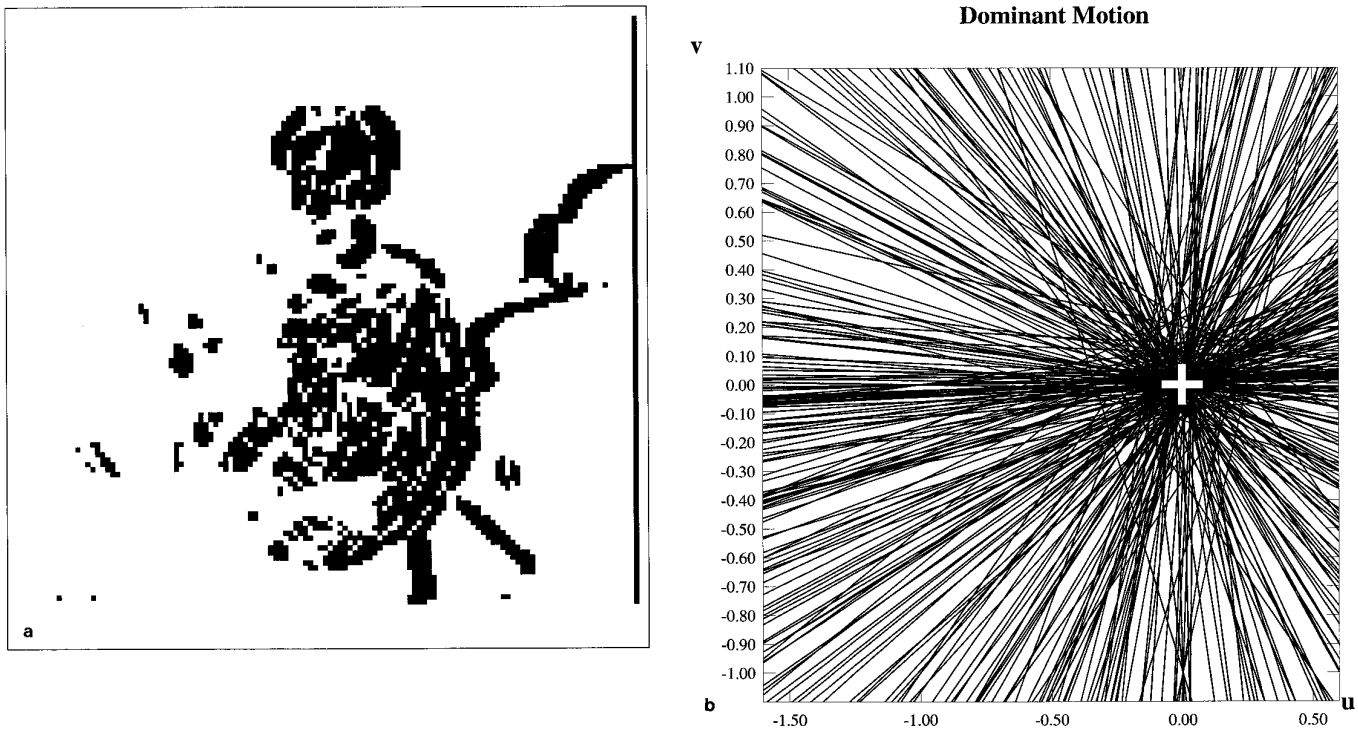

FIG. 2. Robustly determining the dominant motion. (a) Violations of the single motion assumption shown as dark regions. (b) Constraint lines corresponding to the dominant motion.

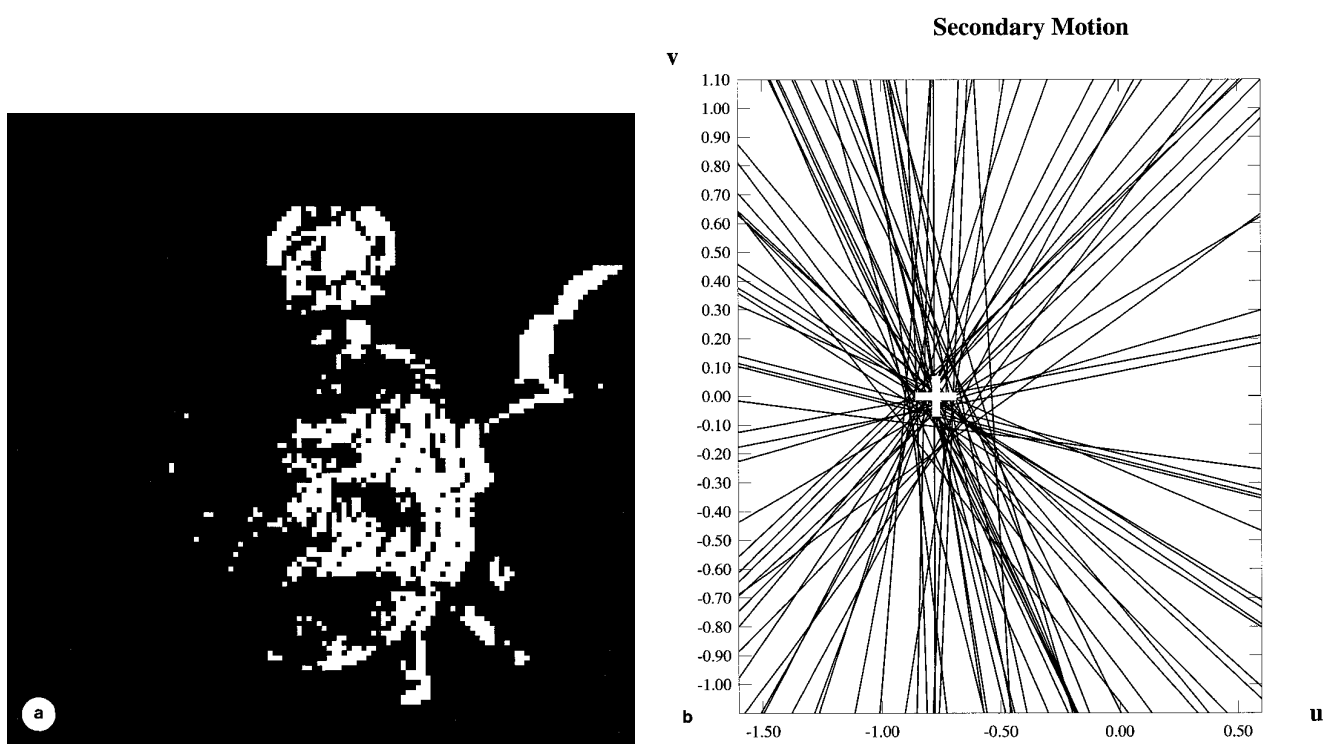

FIG. 3. Estimating multiple motions. (a) White regions correspond to a second consistent motion. (b) Constraint lines corresponding to the secondary motion.

Shizawa and Mase [49] model multiple motions within a region as producing an additive superposition of brightness distributions. Their "principle of superposition" allows them to derive a multiple motion version of the brightness constancy assumption in terms of a composition of multiple operators applied to the brightness distribution. This is essentially the same model used by Bergen *et al.* [5].

Unlike the regression techniques that try to find the best flow given the local intensity constraint equations, Schunck [48] proposes a method of *constraint line clustering* for computing flow estimates near motion discontinuities. The constraint line at the center of a region is intersected with all the other constraint lines within the region. A robust one-dimensional clustering operation is performed to detect the dominant cluster of constraint line intersections. The approach has two serious limitations. First, the accuracy of the result depends on the accuracy of the center constraint line, and at motion boundaries this will likely be inaccurate. Unfortunately, this is exactly the place where we want a robust approach. Second, the one-dimensional clustering implicitly assumes a constant flow model; we would like to be able to use more realistic models of the local flow variation.

Wang and Adelson [57] assume that an image region is modeled by a set of overlapping "layers." They compute initial motion estimates using a least-squares approach within image patches [34]. They then use K-means clustering to group motion estimates into regions of consistent affine motion. Jepson and Black [30] also assume a layered representation and model the constraint lines within a region as a mixture of distributions corresponding to the different layers. They then use an iterative algorithm that assigns constraints to the different layers and estimates the affine motion of each layer. In similar work, Adiv [1] used a generalized Hough technique to estimate the parameters of multiple motions. While the approach is robust, it is discrete and computationally expensive. Each of these approaches is designed to estimate multiple parametric motion models within some fixed image region. These approaches do not appear to offer a general framework which can also cope with piecewise-smooth flow fields.

In [7], we replaced the quadratic error norm used in correlation with a robust error norm. We observed that, when multiple image motions are present, the robust formulation results in better defined "peaks" corresponding to the multiple motions. Robust correlation was later used to provide a robust data term for estimating dense flow fields [9]. In [6, 10], a robust error norm is used to reject data outliers which occur with gradient-based approaches and robust regression was proposed in [6] to recover multiple motions within an image region.

2.2. Spatial Coherence

The various formulations of the data conservation constraint described above are attempts to provide an estimate of the image motion within a region of interest. However, due to the generalized aperture problem, these area-based techniques are not globally applicable and, in general, the best we can do is compute flow estimates in local regions over the image. However, there may be regions which have insignificant spatial intensity variation, leading to poor estimates. Also, as noted earlier, in order to allow the greatest spatial variation of the flow field, it may be desirable to keep the window sizes small, leading to potentially poor local estimates in many regions. The extreme case of this is seen as the standard aperture problem in the case of the gradient-based formulation, when the window size tends to zero and the local flow is only partially constrained by the data conservation constraint

$$I_x u + I_y v + I_t = 0.$$

The standard approach for handling the ill-conditioning that occurs from lack of sufficient intensity variation within local regions is to add a spatial coherence assumption in the form of a regularizing term E_S

$$\begin{aligned} E(\mathbf{u}) &= E_D(\mathbf{u}) + \lambda E_S(\mathbf{u}) \\ &= (I_x u + I_y v + I_t)^2 + \lambda E_S(\mathbf{u}), \end{aligned}$$

where λ controls the relative importance of the data conservation and spatial coherence terms. The introduction of a spatial coherence constraint restricts the class of admissible solutions, making the problem well-posed. Such regularization techniques have received a great deal of attention [36, 44]. The most common formulation of E_S is the *first-order,* or *membrane,* model [27]

$$E_S(u, v) = u_x^2 + u_y^2 + v_x^2 + v_y^2, \tag{8}$$

where the subscripts indicate partial derivatives in the x or y direction. For an image of size $n \times n$ pixels, we define a grid of *sites,*

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

where $(i(s), j(s))$ denotes the pixel coordinates of site s. The first-order smoothness constraint can then be discretized as

$$E_S(\mathbf{u}) = \sum_{s \in S} \left[\frac{1}{8} \sum_{n \in \mathcal{G}_s} [(u_s - u_n)^2 + (v_s - v_n)^2] \right], \tag{9}$$

where the subscripts s and n indicate sites in S and where $\mathcal{G}_s$ represents the set of north, south, east, and west neighbors of s in the grid.

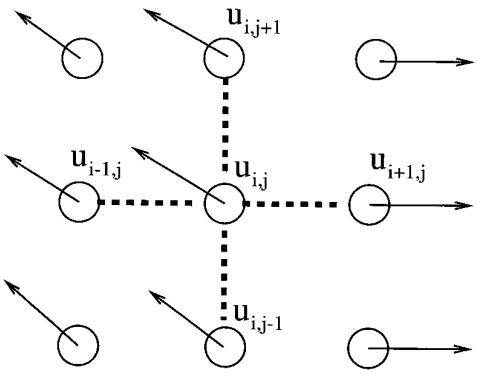

FIG. 4. Smoothing across a flow discontinuity.

2.2.1. *Problems Posed by Motion Boundaries*

As in the case of the data conservation term, the spatial coherence term above is formulated in terms of least-squares estimation. This formulation assumes that the optical flow corresponds to a single (continuously varying) motion locally; that is, the variation in $(u_s - u_n)$ is assumed to be Gaussian. As noted by Shulman and Hervé [50], this assumption is incorrect at motion boundaries and, moreover, we do not know what the true distribution should be.

Consider what happens if the flow field is discontinuous; that is, there are multiple image motions present in the neighborhood. Figure 4 illustrates the situation which occurs at a motion boundary. With the least-squares formulation the local flow vector $\mathbf{u}_{i,j}$ 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.

This oversmoothing is illustrated with an example. Consider the synthetic image sequence in Fig. 5 in which the left half of the image (Fig. 5a) is stationary and the right half is moving one pixel to the left between frames. The horizontal and vertical components of the flow are shown with the magnitude of the flow represented by intensity, where black indicates motion to the left and up and gray indicates no motion. The true horizontal and vertical motions are shown in Figs. 5b and 5c, respectively.

Figure 6 shows a plot of the horizontal motion where the height of the plot corresponds to the recovered image velocity. Figure 6a shows that the application of the least-squares formulation (Horn and Schunck [27]) results in motion estimates which vary smoothly thus obscuring the motion boundary. When we recast the problem in the robust estimation framework, the problems of oversmoothing are reduced (Fig. 6b). Moreover, the robust approach allows violations of the data conservation and spatial coherence assumptions to be detected. These violations, or outliers, are shown in black in Fig. 7. Notice in Fig. 7a that the spatial outliers correspond to the motion boundary. Of particular interest are the data outliers; even in this synthetic example, with no added noise, violations of the data conservation constraint, $(I_x u + I_y v + I_t) = 0$, occur at the motion boundary. The estimate of the derivatives I_x, I_y, and I_t require the pooling of information over some finite spatial neighborhood. When this neighborhood spans a motion boundary, the derivatives may be incorrect and the resulting constraint equation cannot be trusted.

2.2.2. *Previous Approaches*

A large number of researchers have focused on reformulating the regularization problem to allow discontinuities. An important class of techniques for coping with spatial discontinuities are the Markov random field (MRF) formulations [19] which a number of authors have applied to the optical flow problem [8, 9, 32, 38, 54]. These approaches represent discontinuities either explicitly with the use of a "line process" [19] or by using *weak continuity constraints* [12, 23].

The use of the gray-level intensity image to control the behavior of the smoothness constraint in optical flow has also been explored. These approaches disable the smoothness term at intensity edges [14] or constrain it to be enforced only along the directions for which the gray-level variation is not sufficient to determine the flow vector [40].

A related class of approaches uses confidence measures computed from the data to propagate flow measurements from areas of high confidence to areas of low confidence [2]. Singh [51], for example, uses covariance matrices computed from the SSD surfaces and the distribution of flow vectors in small neighborhoods to determine an optimal flow estimate.

Schunck [48] interleaves discontinuity detection and regularization. Given an estimate of the optical flow, motion discontinuities are detected in the flow field [53] and then a smoothness operator is applied that prevents smoothing across the boundary. This gives a new flow estimate and the process is repeated.

Darrell and Pentland [15] have noted the limitations of edge-based schemes when recovering multiple image motions in cases of fragmented occlusion. Instead, they propose a scheme in which different motions are assigned to multiple *layers* using a Markov random field approach. Madarasmi *et al.* [35] take a similar approach in the recovery of multiple transparent surfaces from depth data. They detect and reject outlying data measurements by confining them to a separate layer.

Shulman and Hervé [50] first pointed out that spatial discontinuities in optical flow can be treated as outliers and they proposed a robust regularization approach based on Huber's minimax estimator [28]. Their formulation re-

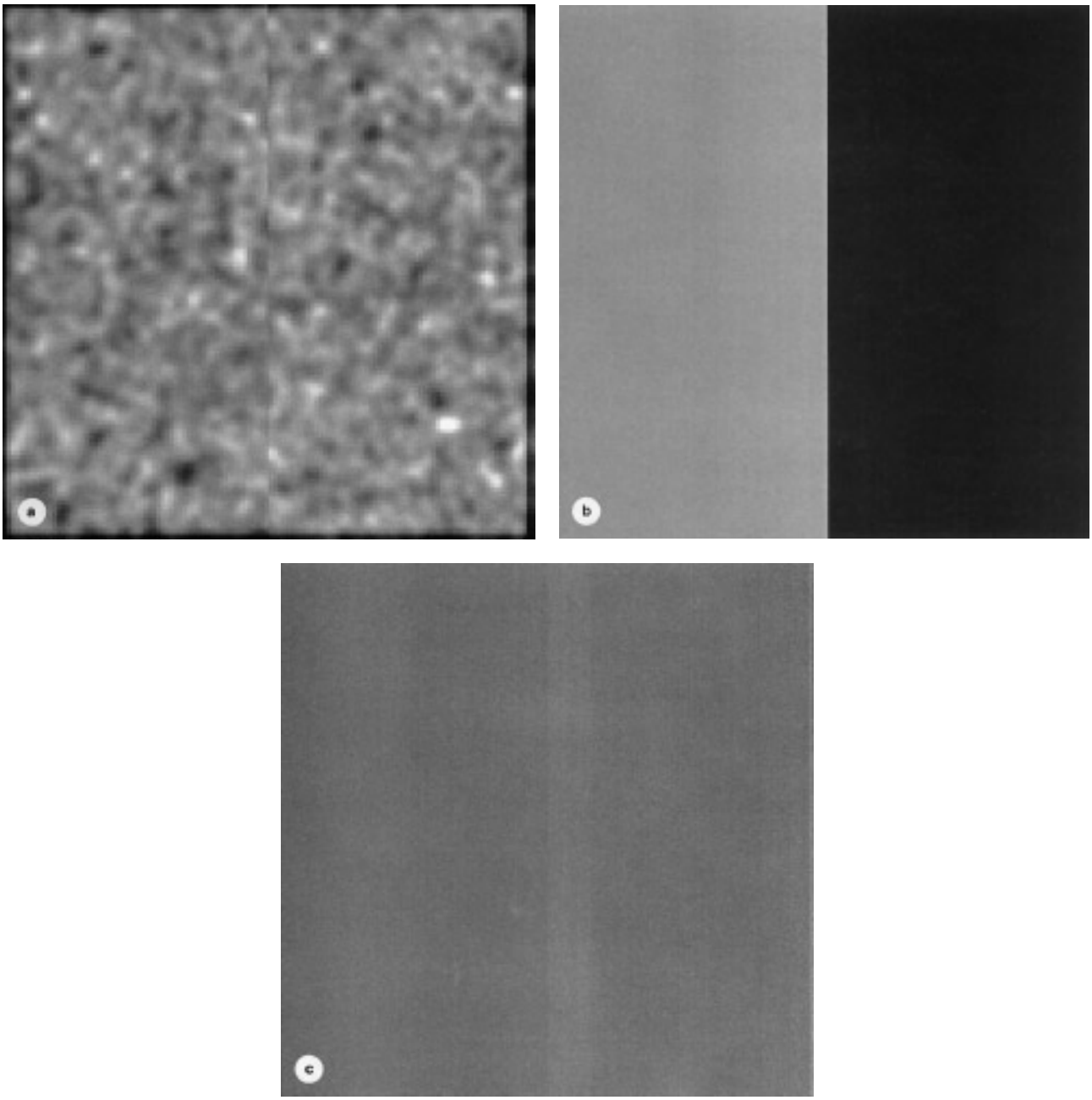

FIG. 5. Random noise example. (a) First random noise image in the sequence. (b) True horizontal motion (black = −1 pixel, white = 1 pixel, gray = 0 pixels). (c) True vertical motion.

sulted in a convex optimization problem which avoided the problem of local minima. Black and Anandan [9] formulated the regularization problem using a robust error norm with a redescending influence function for better outlier rejection. They also introduced a robust data term which was missing in previous approaches. Finally, Black and Rangarajan [11] have shown that there is a deep relationship between these robust approaches and the traditional line-process approaches.

3. ROBUST STATISTICS

The previous section illustrated the generality of the problem posed by motion discontinuities; measurements are corrupted whenever information is pooled from a spatial neighborhood that contains multiple image motions. Model violations such as these result in measurements that can be viewed in a statistical context as *outliers.* We appeal to the field of *robust statistics* [22, 28] to solve the problem

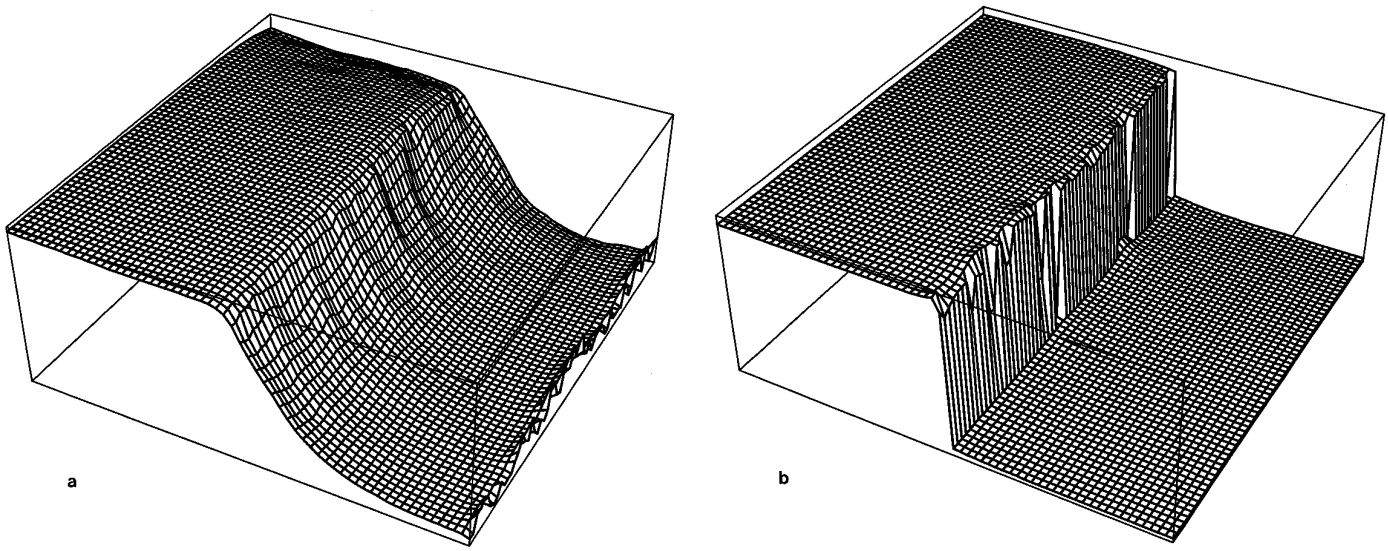

FIG. 6. Horizontal displacement. The horizontal component of motion is interpreted as height and plotted. Plotting the results illustrates the oversmoothing of the least-squares solution (a), and the sharp discontinuity which is preserved by robust estimation (b).

of estimation when the assumptions about the world are idealized and one expects that the assumptions will occasionally be violated. This section introduces the robust statistical techniques relevent to the proposed framework; for a mathematical details, see [22, 28, 47].

As identified by Hampel *et al.* [22], the main goals of robust statistics are to recover the structure that best fits the majority of the data while identifying and rejecting "outliers" or "deviating substructures." While most 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 [37] for a discussion). As mentioned in the previous section, robust statistical techniques have been applied to the problem of image velocity estimation [48, 50], but these previous formulations lack a coherent, unified, framework for addressing the effects of multiple image motions on both the data conservation and spatial coherence assumptions.

Robust estimation [22, 28] 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,

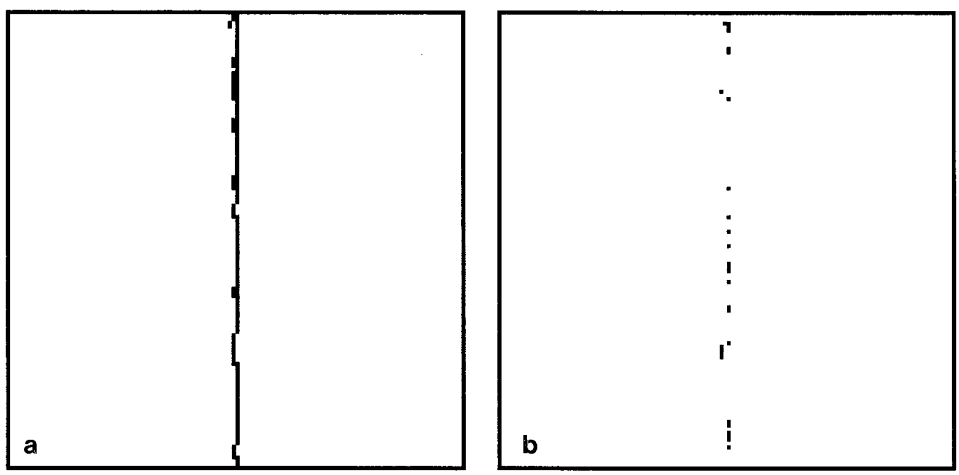

FIG. 7. Random noise sequence outliers. (a) Motion discontinuities where the smoothness constraint is violated. (b) Data outliers.

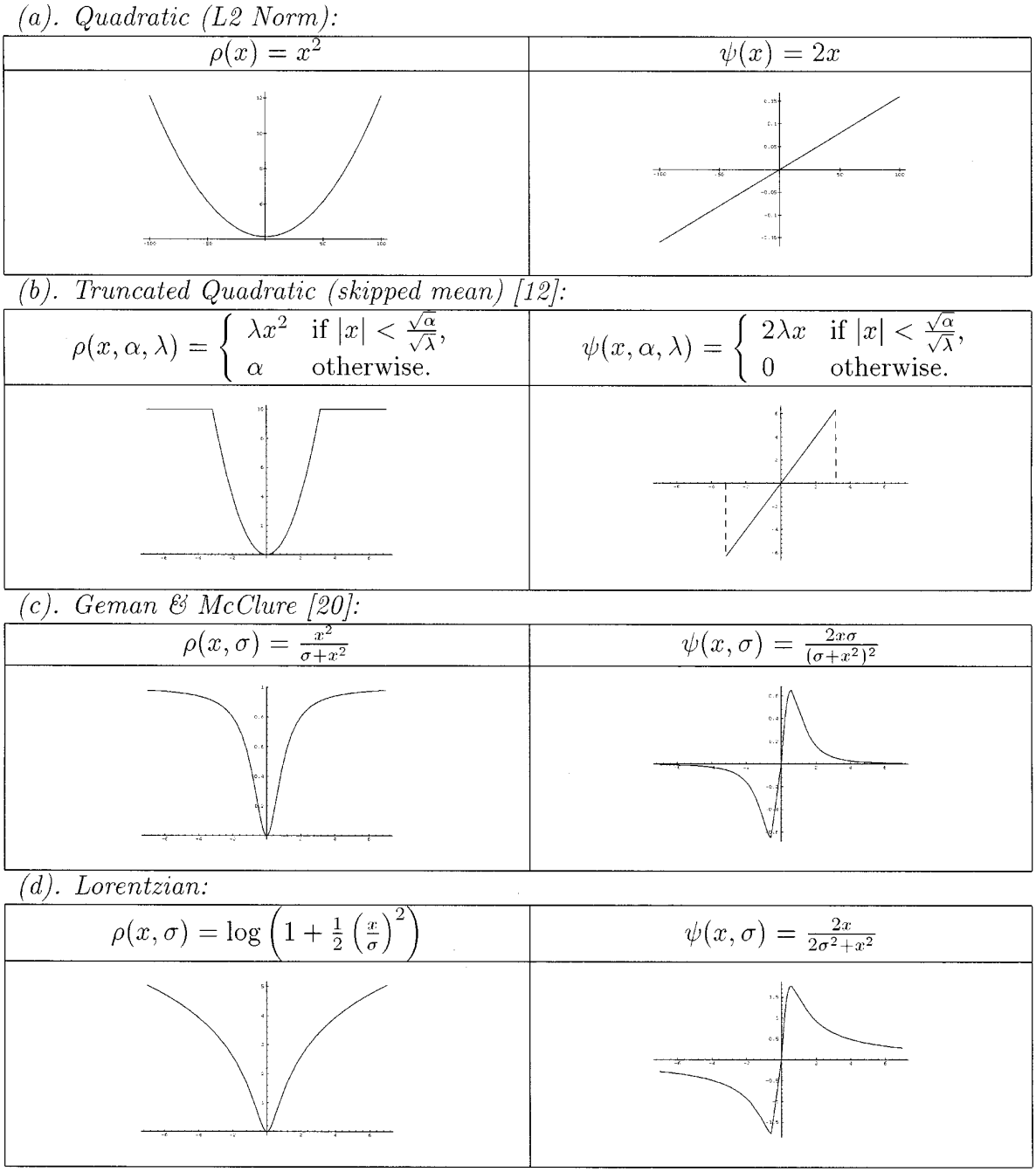

FIG. 8. Common ρ and ψ Functions.

the goal is to find the values for the parameters, $\mathbf{a}$, that minimize the size of the residual errors $(d_s - \mathbf{u}(s; \mathbf{a}))$,

$$\min_{\mathbf{a}} \sum_{s \in S} \rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s), \tag{10}$$

where σ_s is a scale parameter, which may or may not be present, and ρ is our error norm. When the errors in the measurements are normally distributed, the optimal ρ-function is the quadratic

$$\rho(d_s - \mathbf{u}(s; \mathbf{a}), \sigma_s) = \frac{(d_s - \mathbf{u}(s; \mathbf{a}))^2}{2\sigma_s^2}, \tag{11}$$

which gives rise to the standard *least-squares* estimation

problem. Minimizing (10) results in an *M-estimate* since this corresponds to *Maximum-likelihood* estimation. The choice of different ρ-functions results in different robust estimators and the *robustness* of a particular estimator refers to its insensitivity to outliers or gross errors.

3.1. Robust Estimation

As demonstrated in the previous section, least-squares estimation is not appropriate when multiple image motions are present. The application of least squares in these situations caused the recovered flow to be incorrect. The problem with the least-squares solution is that outlying measurements are assigned a high "weight" by the quadratic ρ-function (Fig. 8a, left). One way to see this is by considering the *influence function* associated with a particular ρ-function. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative, ψ, of the ρ-function [22]. In the least-squares case, the influence of data points increases linearly and without bound (Fig. 8a, right).

To increase robustness, a ρ-function must be more forgiving about outlying measurements; in particular, we will consider ρ-functions with *redescending* ψ-functions for which the influence of outliers tends to zero. One of the most common robust ρ-functions in computer vision is the truncated quadratic (Fig. 8b). 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.

There are numerous other ρ-functions that have been used in the computer vision literature, each with different motivations and strengths (see [6] for a review), but their common property is their ability to reduce the effect of outliers. Two examples are the Geman–McClure and Lorentzian ρ-functions (Figs. 8c and 8d) which we will consider in the remainder of the paper. These functions have differentiable ψ-functions which provide a more gradual transition between inliers and outliers than does the truncated quadratic.

3.2. Related Approaches

There are many avenues by which to approach the problem of robustness. One common approach is to detect outliers by examining the least-squares residuals. Unfortunately, even a small number of outliers can arbitrarily corrupt the initial estimate, making the residuals meaningless. Another approach involves the use of robust iterative procedures like least-median-of-squares (LMedS) [37, 47]. While this approach can tolerate a large percentage of outliers, it is most suited to regression or piecewise regression formulations. It is not clear how to estimate continuously varying flow fields involving explicit spatial smoothness constraints using this technique.

We prefer to work in the context of optimization, and formulate our minimization problems to account for outliers by using the robust ρ-functions described above. Although this approach is less robust than the LMedS method, it provides a natural framework for "robustifying" problems in optical flow. It decouples the problem of formulating an optimization problem robustly from the problem of recovering the solution.

4. ROBUST ESTIMATION FRAMEWORK

Section 2 introduced three standard approaches for estimating optical flow which were all posed in terms of least-squares estimation. To improve the robustness, without sacrificing the simple models of image brightness and spatial coherence, the minimization problems are reformulated to account for outliers by using the robust ρ-functions described in the previous section.

The regression approach is simply reformulated as

$$\min_{\mathbf{a}} E_{\mathrm{D}}(\mathbf{a}) \quad \text{where} \quad E_{\mathrm{D}}(\mathbf{a}) = \sum_{\mathscr{R}} \rho((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma), \tag{12}$$

where ρ is a robust ρ-function. Similarly, correlation can be reformulated as the minimization of

$$E_{\mathrm{D}}(\mathbf{u}) = \sum_{(x,y)\in\mathscr{R}} \rho(I(x, y, t) - I(x + u\,\delta t, y + v\,\delta t, t + \delta t), \sigma). \tag{13}$$

The objective function for the regularization approach, with a gradient-based data term, becomes

$$\begin{aligned} E(\mathbf{u}_s) = {} & \rho_{\mathrm{D}}(I_x u_s + I_y v_s + I_t, \sigma_{\mathrm{D}}) \\ & + \lambda \sum_{n\in\mathscr{I}_s} [\rho_{\mathrm{S}}(u_s - u_n, \sigma_{\mathrm{S}}) + \rho_{\mathrm{S}}(v_s - v_n, \sigma_{\mathrm{S}})], \end{aligned} \tag{14}$$

where ρ_{D} and ρ_{S} may be different ρ-functions. Such an approach can likewise be taken with many other early vision problems that are formulated in terms of least-squares optimization.

Notice that we have simply taken these standard formulations of optical flow and made the observation that they correspond to least-squares estimation. Because each approach involves pooling information over a spatial neighborhood these least-squares formulations are inappropriate when multiple image motions are present. By treating the formulations in terms of robust estimation, the problems of oversmoothing and noise sensitivity typically associated with these approaches are alleviated.

This approach has a number of advantages. The formula-

tion of minimization problems in the robust estimation framework is very similar to the familiar least-squares formulations and this gives the robust formulations an intuitive appeal. Robust ρ-functions and their influence functions are also powerful formal and qualitative tools that are useful for analyzing robustness. In particular, examination of the influence functions provides a means of comparing ρ-functions and their effects on outliers. Additionally, as shown in [11], robust ρ-functions are closely related to the traditional line-process approaches for coping with discontinuities. For many ρ-functions it is possible to recover an equivalent formulation in terms of analog line processes. Thus, one can adopt the robust formulation presented here and, if desirable, convert the optimization problem into one with explicit line (or outlier) processes.

4.1. Minimization

Given a robust formulation, there are numerous optimization techniques that can be employed to recover the motion estimates and the most appropriate technique will depend on the particular formulation and choice of ρ-function. In general, the robust formulations do not admit closed form solutions, and often result in an objective function that is nonconvex. We have explored the use of stochastic minimization techniques such as simulated annealing [9] but have found deterministic continuation methods [10] to be more efficient and practical.

4.1.1. *Continuation Methods*

If we choose a robust ρ-function that is twice differentiable then local minima of (12), (13), and (14) can be found using any number of descent methods (e.g., simultaneous over-relaxation (SOR)). Also, many robust ρ-functions have "scale" parameters which allow the shape of the function to be changed. These parameters allow the robust estimation problems to be solved using a continuation method.

Simultaneous Over-Relaxation. Simultaneous over-relaxation (SOR) belongs to a family of relaxation techniques which include *Jacobi's* method and the *Gauss–Seidel* method [45, 52, 56]. For illustration, we consider minimizing the objective function $E(u, v)$ with respect to u, but the same treatment applies for v, or for the a_i in the case of a parametric motion model. The iterative update equation for minimizing E at step $n+1$ and at site s is simply [12]

$$u_s^{(n+1)} = u_s^{(n)} - \omega \frac{1}{T(u_s)} \frac{\partial E}{\partial u_s}, \tag{15}$$

where $0 < \omega < 2$ is an *overrelaxation parameter* which is used to *overcorrect* the estimate of $u^{(n+1)}$ at stage $n+1$. When $0 < \omega < 2$ the method can be shown to converge [56], but the rate of convergence is sensitive to the exact value of ω. The term $T(u_s)$ is an upper bound on the second partial derivative of E,

$$T(u_s) \geq \frac{\partial^2 E}{\partial u_s^2}, \quad \forall s \in S. \tag{16}$$

Details of the robust parametric regression and robust regularization algorithms are provided in the following two sections along with experimental results.

Graduated Non-Convexity. To find a globally optimal solution when the objective function is nonconvex we choose a robust ρ-function with a control parameter and we solve the minimization problem using a continuation method [46]. The general idea is to take the nonconvex 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.

Blake and Zisserman [12] developed a continuation method called Graduated Non-Convexity by constructing a parameterized piecewise polynomial approximation to the truncated quadratic. Various other ρ-functions have been proposed for use in continuation methods [18, 33] and these approaches are closely related to "scale-space" techniques [43].

Formally, the objective function E is convex when the *Hessian matrix, H*, of E is positive definite. 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 locally convex when $\rho''(x, \sigma) > 0$, $\forall x$; that is, σ is chosen so that there are no outliers.

Beyond some threshold, τ, redescending ρ-functions begin to reduce the influence of measurements. Measurements greater than this threshold can be considered outliers. Consider, for example, the Lorentzian ρ-function:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \quad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}.$$

To construct a convex approximation of $E(\mathbf{u})$, we want all measurements to be treated as inliers. The point where the influence of outliers first begins to decrease as the magnitude of the residuals increases from zero occurs when the second derivative of the ρ-function is zero. For the Lorentzian, the second derivative

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

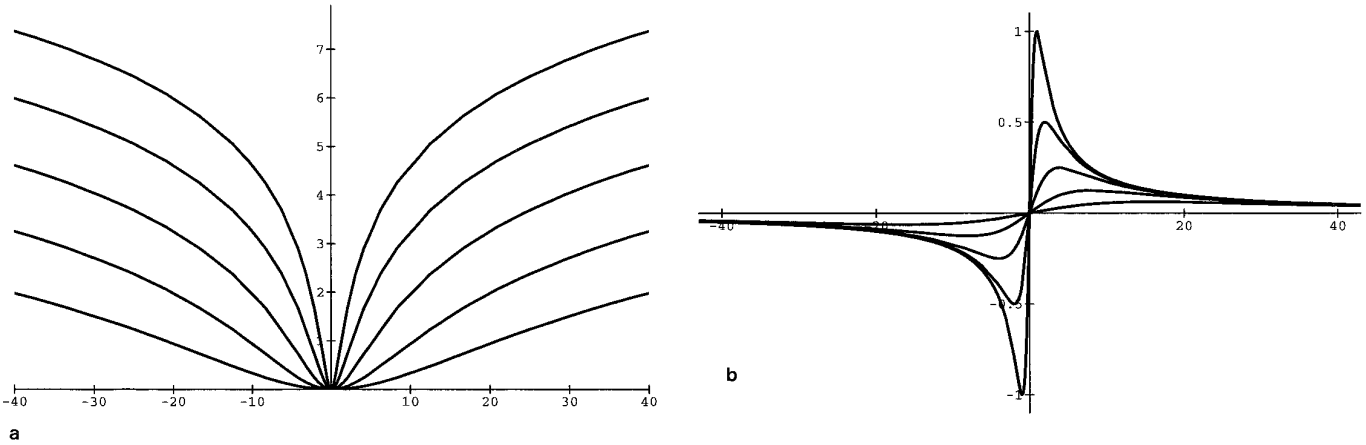

FIG. 9. Graduated Non-Convexity. $\rho(x, \sigma)$ and $\psi(x, \sigma)$ plotted for thresholds $\tau \in \{16, 8, 4, 2, 1\}$. (a) Error measure $\rho(x, \sigma)$. (b) Influence, $\psi(x, \sigma)$.

equals zero when $\tau = \pm\sqrt{2}\sigma$. If the maximum expected residual is τ, then choosing $\sigma = \tau/\sqrt{2}$ will result in a convex optimization problem. A similar treatment applies to other robust ρ-functions. Notice that this also gives a simple test of whether or not a particular residual is treated as an outlier. In the case of the Lorentzian, a residual x is an outlier if $|x| \geq \sqrt{2}\sigma$.

The minimization can begin with the convex approximation and the resulting estimate will contain no outliers. In this sense it will be very much like the least-squares estimate. Outliers can be gradually introduced by lowering the value of σ and repeating the minimization. While this approach works well in practice, it is not guaranteed to converge to the global minimum since, as with least squares, the solution to the initial convex approximation may be arbitrarily bad. Figure 9 shows the Lorentzian error function (Fig. 9a) and its ψ-function (Fig. 9b) for various values of $\tau = \sqrt{2}\sigma$.

4.2. Large Motions

To cope with motions larger than a single pixel, a coarse-to-fine strategy [2, 21] is employed in which we construct a pyramid of spatially filtered and sub-sampled images. Beginning at the lowest spatial resolution with the flow $\mathbf{u}$ being zero, the change in the flow estimate $d\mathbf{u}$ is computed. The new flow field, $\mathbf{u} + d\mathbf{u}$, is then projected to the next level in the pyramid (scaled as appropriate) and the first image at that level is *warped* toward the later image using the flow information. The warped image is then used to compute the $d\mathbf{u}$ at this level. The process is repeated until the flow has been computed at the full resolution. Details of the coarse-to-fine strategy are given in the following sections and warping in the case of affine motion is described in Appendix B.

5. ROBUST REGRESSION EXPERIMENTS

To illustrate the robust estimation framework we first consider the regression approach where the goal is to recover the affine parameters $\mathbf{a}$ that minimize

$$E_{\mathrm{D}}(\mathbf{a}) = \sum_{\mathscr{R}} \rho((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma), \tag{17}$$

where σ is a control parameter, $\mathscr{R}$ is taken to be the entire image for these experiments, and ρ is the Geman–McClure norm [20]

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

Also, recall that the affine flow in a region is described by

$$\mathbf{u}(x, y; \mathbf{a}) = \begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x + a_2 y \\ a_3 + a_4 x + a_5 y \end{bmatrix}. \tag{18}$$

The iterative update equations from the previous section are

$$a_i^{(n+1)} = a_i^{(n)} - \omega \frac{1}{T_{a_i}} \frac{\partial E_{\mathrm{D}}}{\partial a_i}, \tag{19}$$

for each of the six motion parameters where $\omega = 1.995$ for all experiments. Details of the update equations are given in Appendix A. Finally, the spatial and temporal derivatives (I_x, I_y, I_t) were estimated using the simple technique described by Horn [26].

The algorithm begins by constructing a Gaussian pyramid. At the coarse level all affine parameters are initially

zero. The translational terms a_0 and a_3 are solved for first by performing n iterations of the update equations above (or until the change in the parameters is less than a threshold). For all experiments in this section n is 30 and the threshold is 10^{-6}. If the image at the current level is larger than 50×50 then the linear terms (a_1, a_2, a_4, a_5) are found in the same manner. For smaller images, the affine parameters cannot be determined with sufficient robustness.

The next level in the pyramid is then processed. The initial estimates at this level are ($2a_0$, a_1, a_2, $2a_3$, a_4, a_5). These parameters are used to warp the first image toward the second (see Appendix B for details). The change in the affine parameters is then computed using the iterative update scheme. The process is repeated until the full resolution is reached.

A continuation method is used during the iterative updating of each level. At each iteration, the control parameter σ is lowered according to the schedule $\sigma_{t+1} = 0.95\sigma_t$. The effect of this process is that initially all points in the region contribute to the solution and gradually the influence of outlying residuals is reduced.

Once the dominant motion has been determined the outlying measurements can be determined by checking whether $|(\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t)| \geq \tau$, where τ is determined by the error norm and the control parameter. For the Geman–McClure norm $\tau = \sigma/\sqrt{3}$. If a sufficient number of outliers are present then we perform the affine regression again using only the outlying constraints. This process of robustly fitting affine motions continues until the number of constraints remaining unaccounted for is below a threshold.

This section presents four different experiments which show the robust regression approach applied to motion boundaries, fragmented occlusion, transparency, and specular reflections. The parameters in all experiments are the same unless otherwise noted. The only parameters that vary are the starting and ending values of σ and the solutions are fairly stable for changes in σ of less than an order of magnitude.

5.1. Motion Boundaries: Estimating the Dominant Motion

The first experiment illustrates the robust estimation of a single dominant image motion within a region. For clarity we consider a simple case involving two constant motions. Figure 10 shows an experiment using a translational model where the goal is to estimate the horizontal component of the dominant motion within the region. There are two random noise patterns present in the window; one moving to the right, the other moving up. The best estimate of the horizontal component is computed as increasing amounts of the upward motion (the distractor) are added to the region.

Figure 11 shows the results of the experiment. The dominant horizontal motion is shown as a solid line; initially the horizontal motion is 1.0 and then, when 50% of the region is occupied by the distractor, the horizontal motion is 0.0. The robust formulation does a good job of recovering the dominant horizontal motion and ignoring the distracting motion until approximately 40% of the region is occupied by the distractor. Not surprisingly, the least-squares approach performs poorly, producing the mean horizontal motion rather than the dominant one.

5.2. Fragmented Occlusion: Multiple Affine Motions

The synthetic image sequence shown in Figs. 12a and 12b contains two affine motions and is a good example of fragmented occlusion. It is difficult to see in this figure, but there is a crowd of people in the background that is occluded by the trees in the foreground. A four-level pyramid was used and the value of the control parameter was lowered from $\sigma = 25.0\sqrt{3}$ to $\sigma = 15.0\sqrt{3}$. The processing took approximately 100 s on a SPARCstation 10 to recover the two motions for the 256×192 pixel image.

The dominant motion was recovered and the outliers were detected. Figure 12c shows these outliers as black regions. These regions were then used to estimate a second affine motion for the background; the constraints consistent with the background motion are shown in Fig. 12d in white.

The recovered affine parameters of the foreground and background are:

	a_0	a_1	a_2	a_3	a_4	a_5
Foreground	0.765940	0.010489	0.001311	−2.137969	0.001398	0.003769
Background	−3.539432	0.003995	0.002025	−3.527813	−0.000553	0.003694

To demonstrate the accuracy of the estimated motion we warp the first image toward the second using the affine motion. The pixel difference between the warped image and the second image is computed and shown in Figs. 12e and 12f. The nulling of the foreground and background in the figures attempts to illustrate that the correct motions have been recovered (see [5] for comparison).

5.3. Transparent Motion

Multiple image motions can also occur as the result of transparency and reflection. Figure 13a shows an example of transparent motion in which the head of a person appears reflected in the glass of a photograph. In addition to assuming a single motion within a region, the brightness constancy assumption assumes that the brightness of a pixel corresponds to a single surface in the scene. In the case of reflections this assumption is violated [5, 49].

The dominant motion for the entire image is estimated using an affine flow model and a three-level pyramid. The

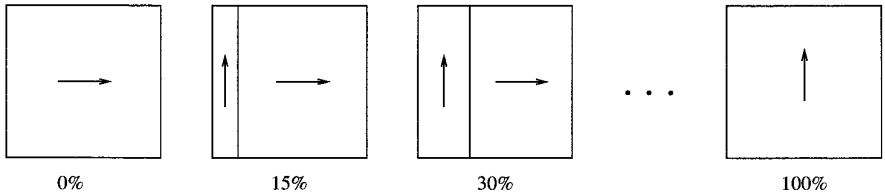

FIG. 10. Constant model experiment.

value of the control parameter was lowered from $\sigma = 20.0\sqrt{3}$ to $\sigma = 7.5\sqrt{3}$. The pixels that were unaccounted for by the dominant motion were used to compute a second motion using the same parameters. The two motions took 49 s to estimate on a 75-MHz SGI Indy workstation for the 320×420 pixel image.

The recovered parameters are:

	a_0	a_1	a_2	a_3	a_4	a_5
Dominant	0.284268	−0.005267	0.002021	−0.757518	−0.003393	−0.004080
Secondary	1.595042	−0.001662	−0.003794	0.033757	0.001282	0.001587

The dominant motion corresponds to the image of the woman in the photograph. Figure 13b shows in black the outliers where the brightness constancy assumption is violated for the dominant motion. These outliers correspond primarily to the figure of the face reflected in the photograph. The pixels used to compute the motion of the reflected face are shown in Figure 13c.

Figure 13d shows the result obtained by warping the first image by the dominant motion and computing the difference image between the warped and second images. This nulling of the dominant motion reveals the reflected face in the difference image. Similarly, Figure 13e shows the difference image obtained using the motion of the reflected face. In this case, the face motion is compensated for and, hence, the woman appears in the difference image.

For this experiment we had no "model" of transparent image formation (e.g., [5, 49]). Instead, we used the standard brightness constancy assumption which gave enough "good" constraints to estimate both motions. This approach, however, will not work for general transparent motion. Instead, we need to obtain our motion constraints using an approach which is more suited for transparency (e.g., the phase-based approach of Fleet and Jepson [17]).

5.4. Specular Reflections

Finally, we consider the problem of specular reflections within a region. In general, the motion of a specularity will not be consistent with the motion of the surfaces that give rise to it. For example, consider the image in Fig. 14a in which the camera is translating toward a Coke can. The

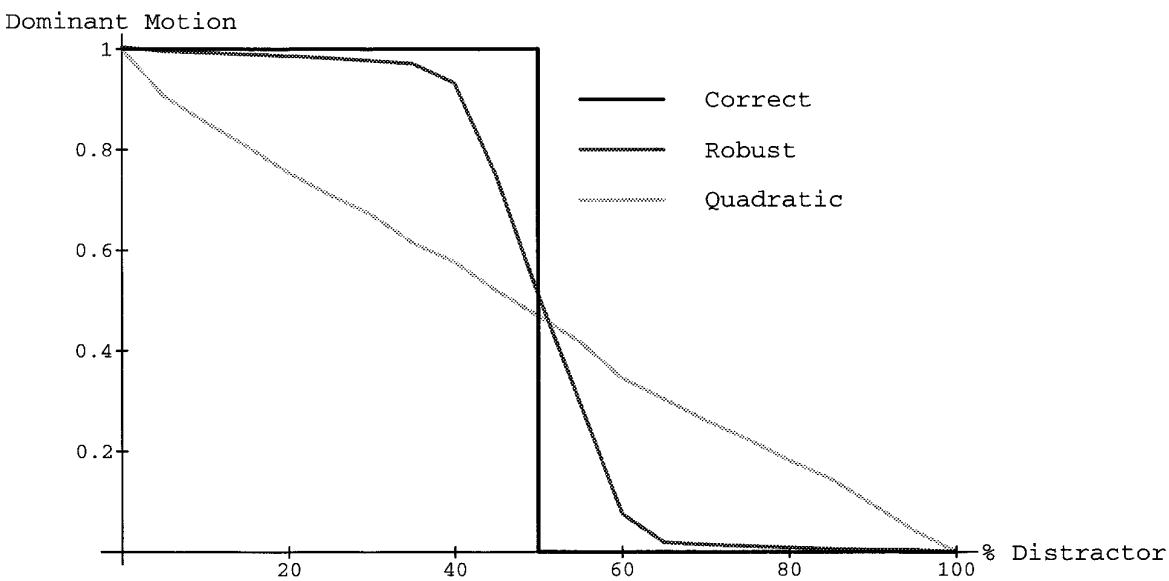

FIG. 11. Constant model experimental results. Error for quadratic and robust formulations as a function of the amount of distractor.

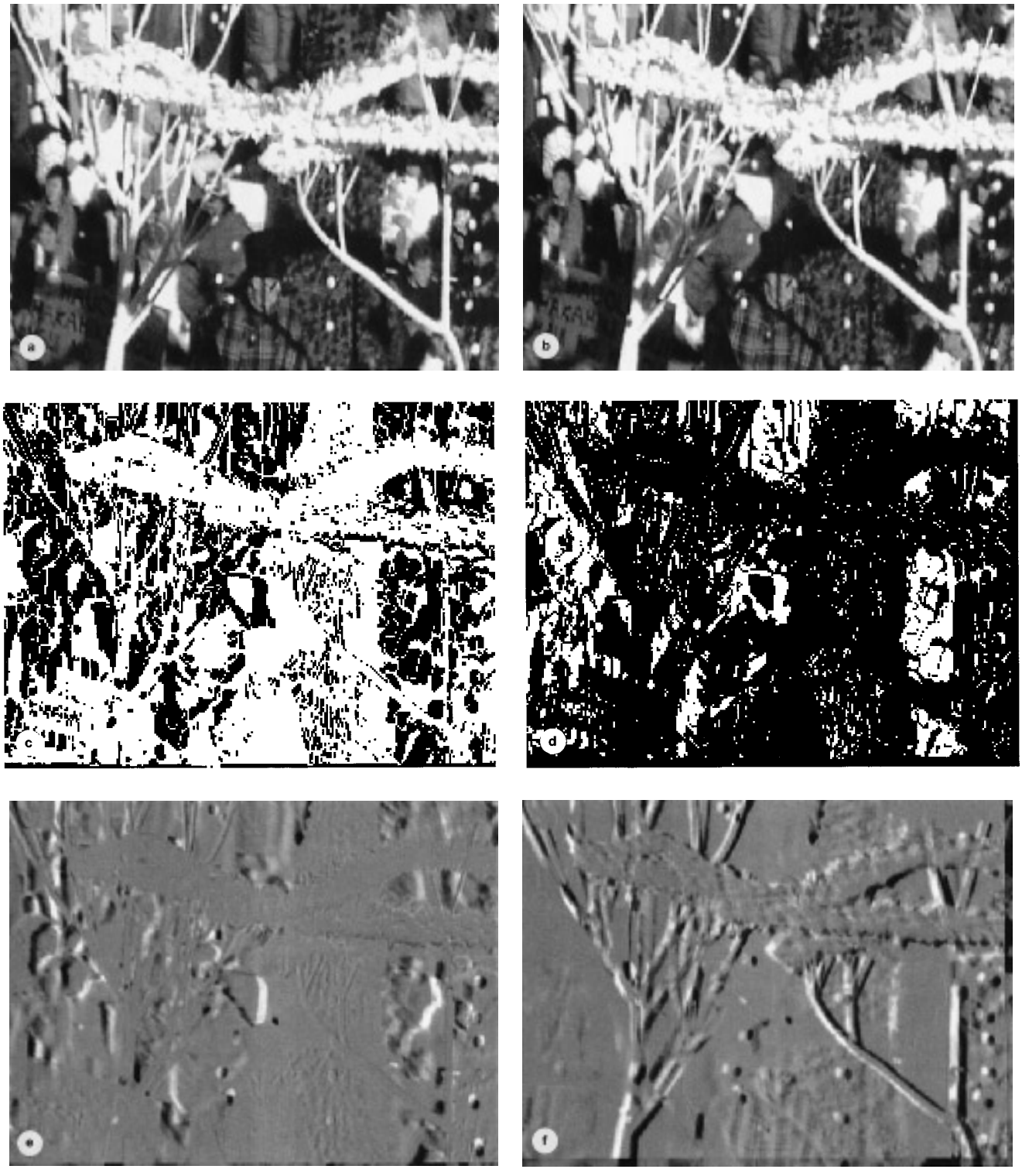

FIG. 12. Multiple affine motion experiment. (a, b). Images in the Skater sequence containing two affine motions. (c). Black regions correspond to outliers (i.e., the white regions are consistent with the dominant motion of the foreground). (d). White regions correspond to data consistent with the motion of the background. (e, f). Warped difference images corresponding to the foreground and background motions respectively (see text).

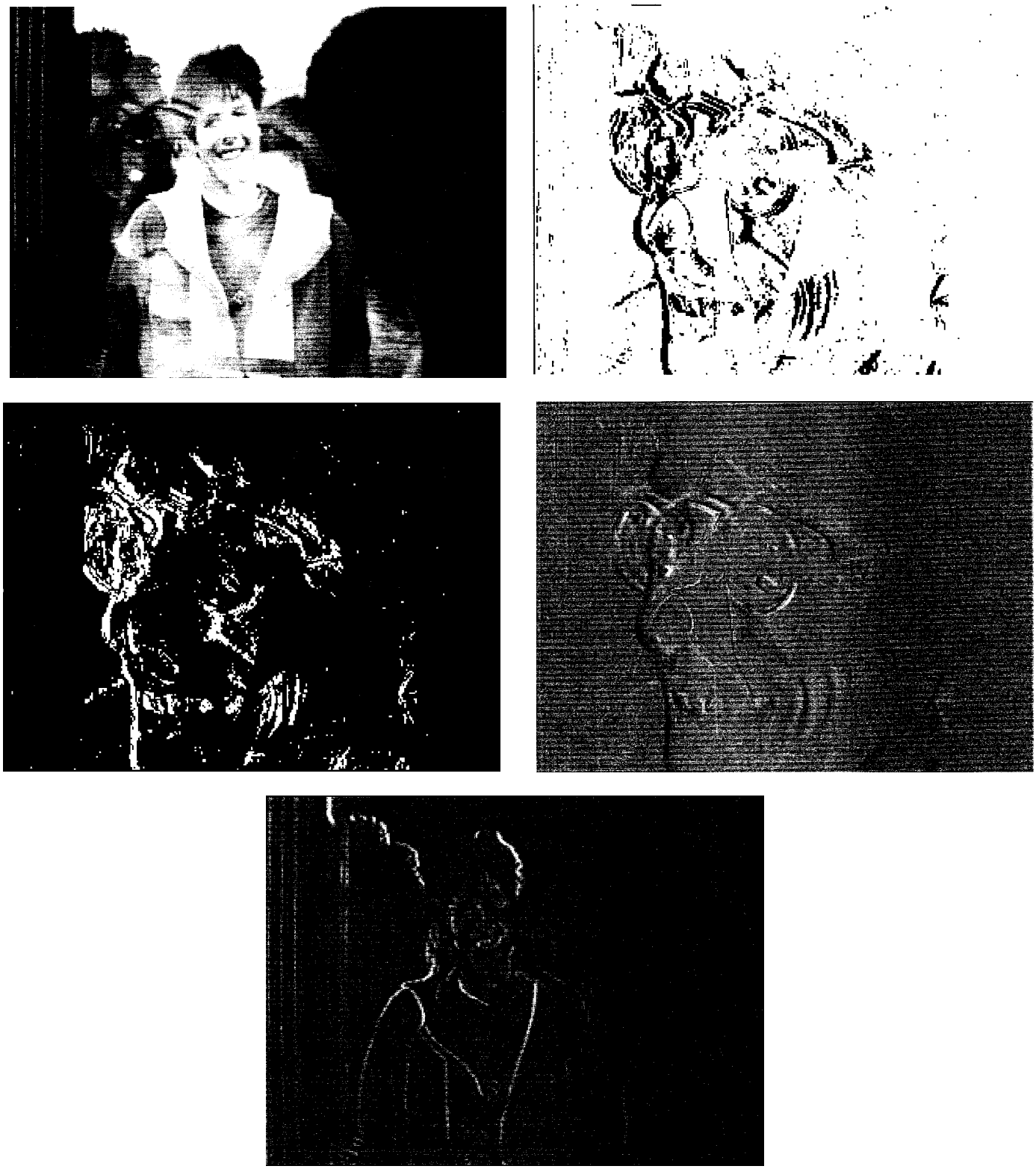

FIG. 13. Transparency sequence. (a) First image of a face reflected in a framed photograph. (b) Black pixels indicate where the brightness constancy assumption is violated. (c) White pixels correspond to points used for estimating the second motion. (d) Difference image obtained by warping the first image by the dominant velocity and computing the pixel differences between the warped and the second images. (e) Difference image obtained using the second estimated motion.

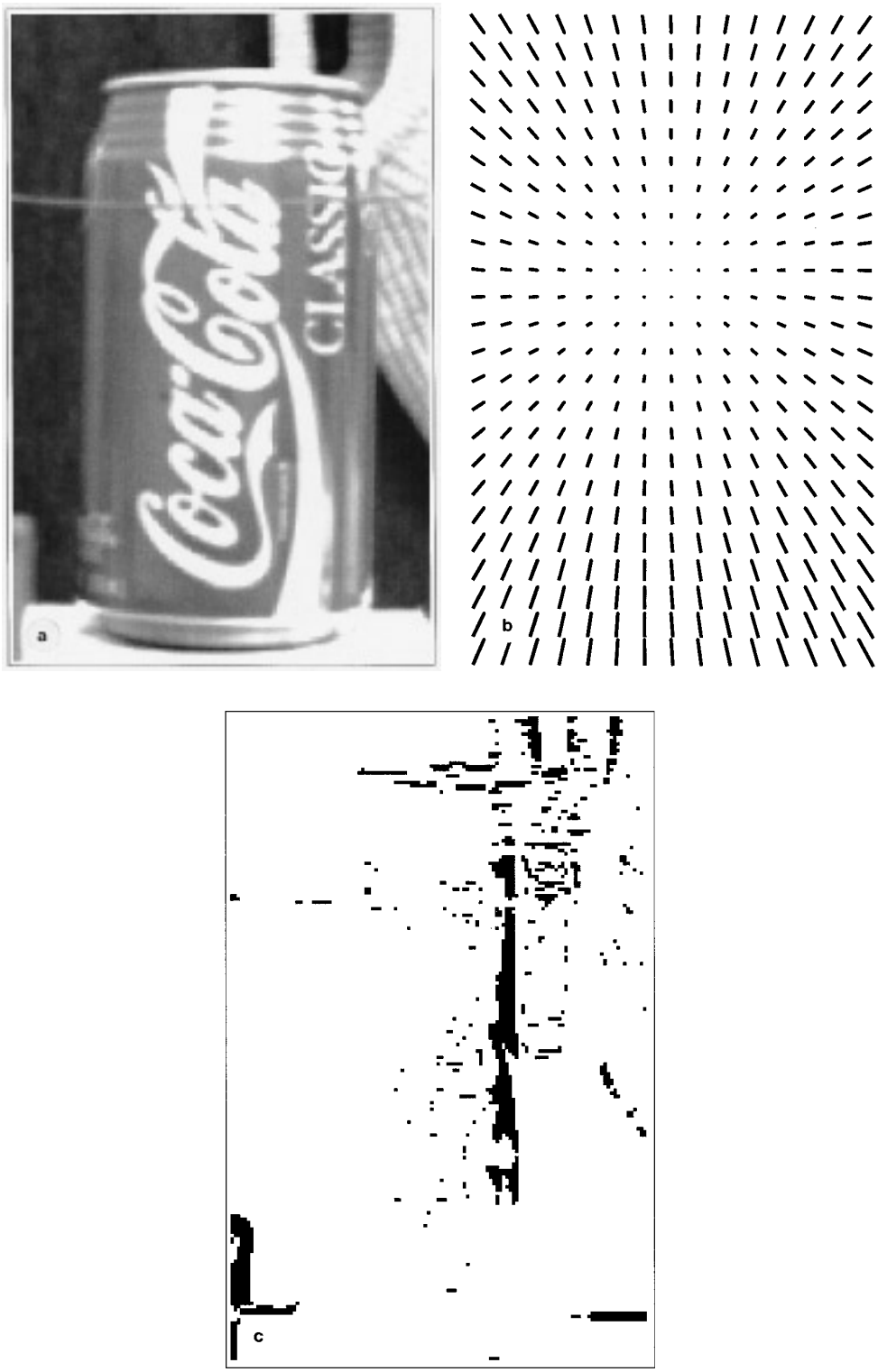

FIG. 14. Specular reflections. (a). Portion of the NASA Coke can sequence with the camera translating along the camera axis. (b). Recovered affine flow shows the image expansion. (c). Specular reflections on the can are not consistent with the affine expansion and are treated as data outliers (in black).

dominant motion of the image region is well modeled by affine flow, yet the background and foreground have slightly different motions and there are specular reflections on the surface of the can.

The affine flow recovered by the robust regression approach is seen in Fig. 14b. A three-level pyramid was used, and the initial and final values for σ were $20\sqrt{3}$ and $10\sqrt{3}$, respectively. The motion took approximately 24 s to compute on a SPARCstation 10 for the 128×200 pixel image.

The recovered affine parameters are:

a_0	a_1	a_2	a_3	a_4	a_5
0.050001	0.017162	0.001172	0.327598	0.000948	0.018397

Figure 14c shows the detected data outliers in black. These outliers correspond predominantly to the specular reflection on the Coke can. The background motion is very similar to that of the Coke can and hence only a few outliers are detected in the textured regions of the sweater.

6. ROBUST REGULARIZATION EXPERIMENTS

We now turn to see how the robust estimation framework can be applied to another common approach for estimating optical flow. We consider a robust gradient-based formulation,

$$E(u,v) = \sum_{s\in S}\Bigg[\lambda_{\mathrm{D}}\rho_{\mathrm{D}}(I_x u_s + I_y v_s + I_t, \sigma_{\mathrm{D}}) + \lambda_{\mathrm{S}}\Bigg[\sum_{n\in\mathscr{G}_s}\rho_{\mathrm{S}}(u_s - u_n, \sigma_{\mathrm{S}}) + \sum_{n\in\mathscr{G}_s}\rho_{\mathrm{S}}(v_s - v_n, \sigma_{\mathrm{S}})\Bigg]\Bigg], \tag{20}$$

where ρ is a robust error norm and where σ_{D} and σ_{S} are the control parameters. For historical reasons, in this set of experiments we take $\rho_{\mathrm{S}} = \rho_{\mathrm{D}}$ to be the Lorentzian error norm. The exact choice of error norm is less important than the qualitative shape of its influence function which, in the case of the Lorentzian, is similar to the Geman–McClure norm used in the previous section.

The iterative update equations for minimizing E at step $n + 1$ are simply [12]

$$u_s^{(n+1)} = u_s^{(n)} - \omega\frac{1}{T(u_s)}\frac{\partial E}{\partial u_s}, \tag{21}$$

$$v_s^{(n+1)} = v_s^{(n)} - \omega\frac{1}{T(v_s)}\frac{\partial E}{\partial v_s}, \tag{22}$$

where

$$\frac{\partial E}{\partial u_s} = \sum_{s\in S}[\lambda_{\mathrm{D}} I_x \psi(I_x u_s + I_y v_s + I_t, \sigma_{\mathrm{D}}) + \lambda_{\mathrm{S}}\sum_{n\in\mathscr{G}_s}\psi(u_s - u_n, \sigma_{\mathrm{S}})], \tag{23}$$

$$\frac{\partial E}{\partial v_s} = \sum_{s\in S}\Bigg[\lambda_{\mathrm{D}} I_y \psi(I_x u_s + I_y v_s + I_t, \sigma_{\mathrm{D}}) + \lambda_{\mathrm{S}}\sum_{n\in\mathscr{G}_s}\psi(v_s - v_n, \sigma_{\mathrm{S}})\Bigg]. \tag{24}$$

While determining the optimal value of the overrelaxation parameter, ω, is difficult in the case of a nonlinear problem, we can approximate it with the optimal value for Jacobi relaxation which is determined by the largest eigenvalue ($\mu_{\max}$) of the Jacobi iteration matrix

$$\mu_{\max} = \cos \pi h, \quad h = \frac{1}{(n+1)}, \tag{25}$$

for an $n \times n$ problem [52]. The approximation to the optimal overcorrection is then

$$\omega_{\mathrm{opt}} = \frac{2(1 - \sqrt{1 - \mu_{\max}^2})}{\mu_{\max}^2}. \tag{26}$$

In practice, this approximation works well and, for an $n \times n$ image, acceptable convergence is reached within only n iterations. Faster convergence can be achieved using *Chebyshev acceleration* [45].

Recall that the terms $T(u_s)$ and $T(v_s)$ are upper bounds on the second partial derivatives of E. The second derivative is maximized when both the data and smoothness errors are zero everywhere, which means we can take

$$T(u_s) = \frac{\lambda_{\mathrm{D}} I_x^2}{\sigma_{\mathrm{D}}^2} + \frac{4\lambda_{\mathrm{S}}}{\sigma_{\mathrm{S}}^2}, \quad T(v_s) = \frac{\lambda_{\mathrm{D}} I_y^2}{\sigma_{\mathrm{D}}^2} + \frac{4\lambda_{\mathrm{S}}}{\sigma_{\mathrm{S}}^2}. \tag{27}$$

Recall the outlier threshold for the Lorentzian, in terms of σ, is given by $\tau = \pm\sqrt{2}\sigma$. For the spatial coherence term, the value of τ_S is determined experimentally. Motion discontinuities $l_{s,t}$ between pixel sites s and t can be recovered from the computed flow field by examining where the threshold τ_S is exceeded

$$l_{s,t} = \begin{cases} 0, & |u_s - u_t| \geq \tau_{\mathrm{S}} \text{ or } |v_s - v_t| \geq \tau_{\mathrm{S}}, \\ 1, & \text{otherwise.} \end{cases} \tag{28}$$

A coarse-to-fine strategy is used with a fixed number of iterations of the SOR algorithm at each level in the pyramid where an iteration involves the updating of every site in

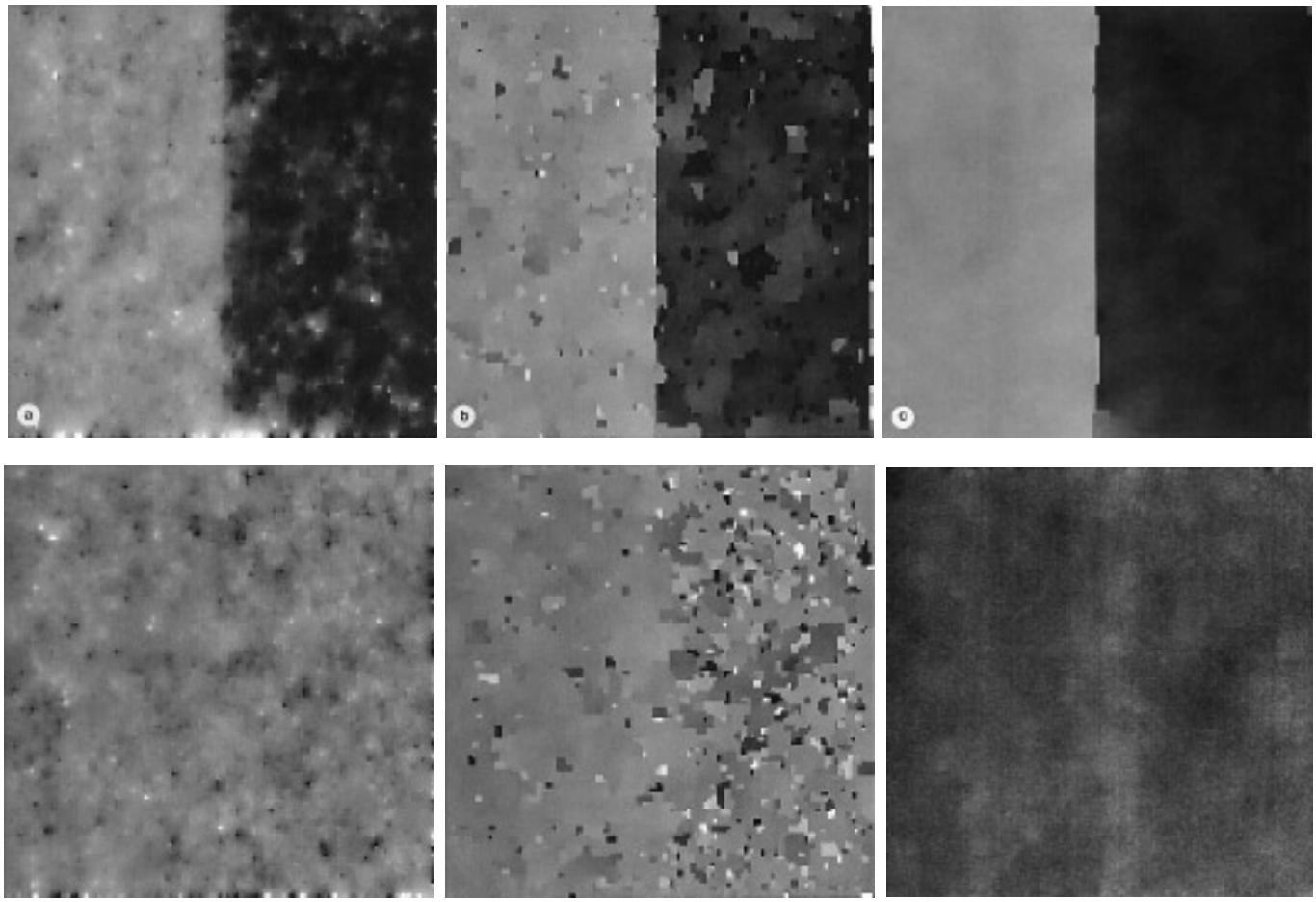

FIG. 15. Effect of robust data term (10% uniform noise). (a) Least-squares (quadratic) solution. (b) Quadratic data term and robust smoothness term. (c) Fully robust formulation.

the flow field. When a level is finished the flow field is projected to the next level in the pyramid and the first image is warped using this flow field:

$$I_{\text{warped}}(x, y) = I(x - u(x, y), y - v(x, y)). \tag{29}$$

The image derivatives are then computed at this level using the warped image as opposed to the original image. When the flow has been updated the process repeats until the flow has been computed at the finest level. After the coarse-to-fine strategy has been applied to the convex approximation, the values of σ_S and σ_D are lowered according to the annealing schedule and the entire coarse-to-fine strategy is repeated. As the coarse-to-fine process is repeated, a flow vector is projected to the next finer level only when it differs significantly from the current estimate at the fine level (see [6] for details).

The remainder of this section contains experiments with natural and synthetic images that illustrate the effects of the robust data and smoothness terms, the behavior of the continuation method, and the accuracy of the recovered flow. The experiments use a linear continuation schedule which lowers the control parameters more gradually than the geometric schedule used in the previous section. In the case of the parametric models a large number of constraints and a small number of parameters mean that the solution typically converges quickly and a rapid continuation method can be used. With regularization, the convergence is slower and benefits from a more gradual lowering of the control parameters.

All parameters were exactly the same in all the experiments illustrating the stability of the robust approach. A three-level pyramid was used with a six stage continuation method in which σ_D varied linearly from $18/\sqrt{2}$ to $5/\sqrt{2}$ and σ_S varied linearly from $3/\sqrt{2}$ to $0.03/\sqrt{2}$. The weights for the data and spatial terms were $\lambda_D = 5$ and $\lambda_S = 1$, respectively. Twenty iterations were performed at each level of the pyramid.

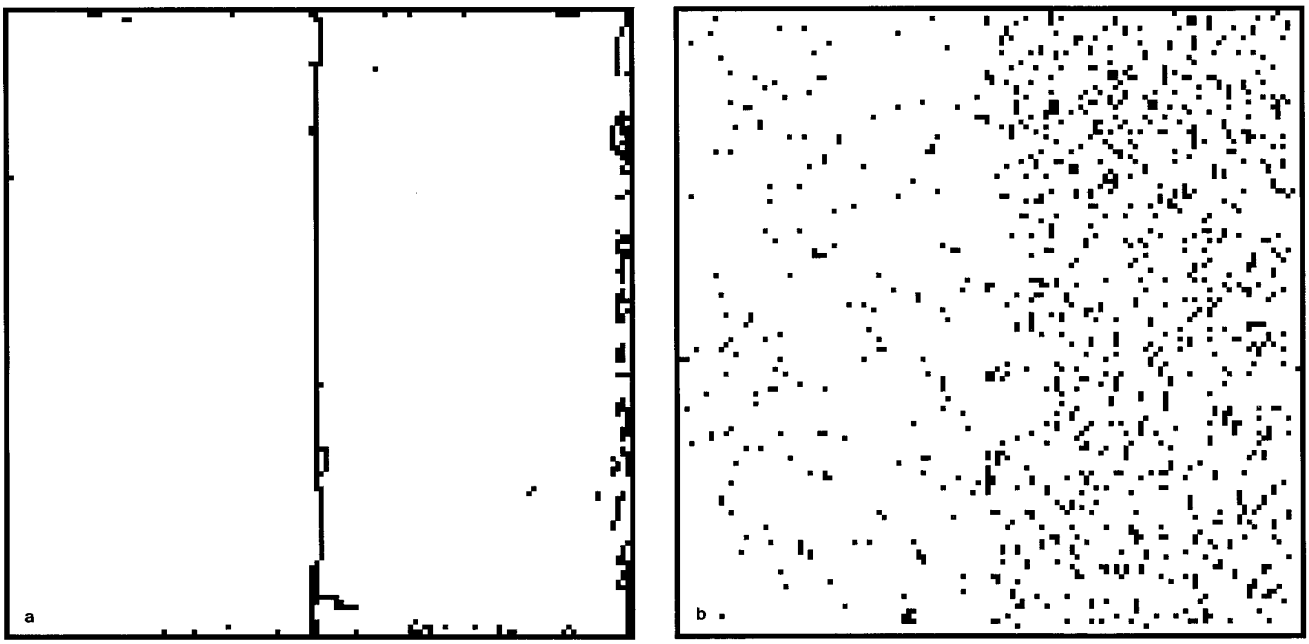

FIG. 16. Outliers in the smoothness and data terms (10% uniform noise). (a) Flow discontinuities. (b) Data outliers.

6.1. Robust Data and Smoothness Terms

A significant difference between the robust gradient method and previous approaches to computing optical flow with discontinuities is that here the data component of the objective function is made robust. A more traditional formulation of the problem would include a line process or weak continuity constraint for the smoothness term [12, 19, 23] and leave a quadratic data term.

The first experiment involves a synthetic sequence containing two textured surfaces, one which is stationary and one which is translating one pixel to the left. The second image in the sequence has been corrupted with 10% uniform random noise. We compare the performance of three common approaches: a purely quadratic formulation (Horn and Schunck), a version with a quadratic data term and robust smoothness term, and the fully robust formulation described here.

TABLE 1
Behavior of Data Term

Approach	RMS flow error	RMS brightness error
Both terms quadratic	0.1814	2.600
Quadratic data, robust smoothness	0.2208	1.889
Both terms robust	0.0986	2.653

Note. The table shows the effects of the robust data and smoothness terms. The root-mean-squared errors in the horizontal flow estimate and the data term, $(I_x u + I_y v + I_t)$, are shown for three common approaches.

Table 1 summarizes the results. The purely quadratic solution attempts to be faithful to both the smoothness model and the noisy data; the result is moderately high errors in both the flow estimate and the brightness constraint. Adding a robust smoothness term (for example by employing a line process) results in lower errors in the brightness constraint equation but with higher error in the flow estimate. With such a formulation, gross errors in the brightness data pull the solution away from the true flow, while the robust term compounds matters by allowing discontinuities to be introduced. The fully robust version appears to provide the best balance. The robust data term allows the brightness constraint to be violated and, consequently, this version has the highest brightness error and the lowest error in the recovered flow field.

The results are illustrated in Fig. 15. The top row shows the horizontal motion and the bottom row shows the vertical motion recovered by each of the approaches (black = -1 pixel, while = 1 pixel, gray = 0 pixels). Figure 15a shows the noisy, but smooth, results obtained by least-squares. Figure 15b 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 15c shows the improvement realized when both the data and spatial terms are robust.

Outliers are detected where the final values of the data coherence and spatial smoothness terms are greater than the outlier thresholds τ_D and τ_S ($\sqrt{2}\sigma_D$ and $\sqrt{2}\sigma_S$ for the Lorentzian). Motion discontinuities are simply outliers

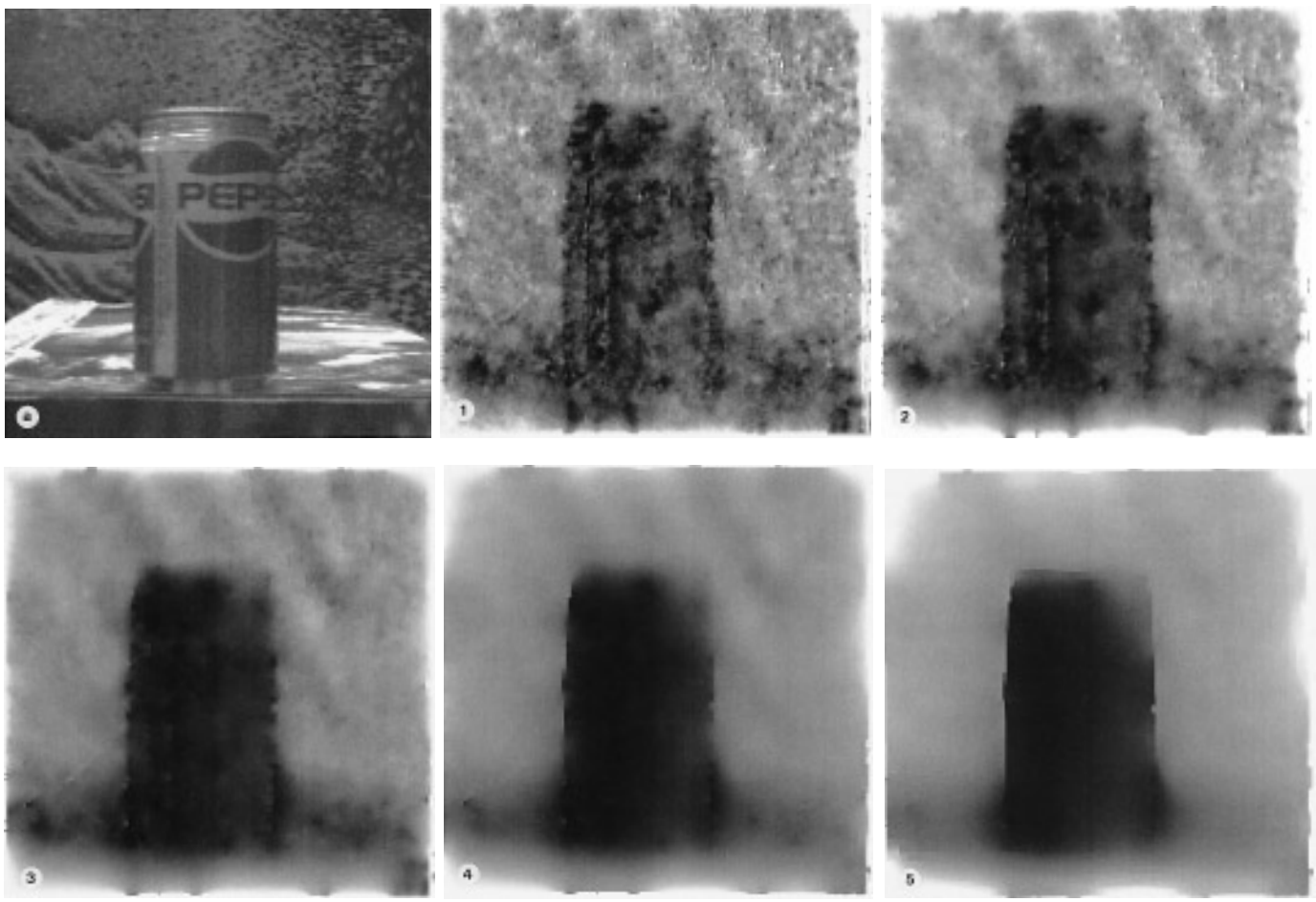

FIG. 17. The Pepsi sequence; horizontal flow. Results of the continuation method stages 1–5 (see text).

with respect to spatial smoothness (Fig. 16a). A large number of data points are treated as outliers by the data term; especially when the motion is large (Fig. 16b).

6.2. Pepsi Sequence

To illustrate the behavior of the continuation method we consider a natural image sequence containing a Pepsi can and a textured background in which the camera motion is parallel to the image plane. The first image is shown in the Fig. 17a. The can is displaced approximately 1.6 pixels to the left between frames, while the background moves approximately 0.7 pixels to the left. The images are 201 $\times$ 201 pixels and each of the six stages took approximately 49 s to compute on a SPARCstation 10.

Figure 17 (1–5) shows the horizontal component of the flow field at stages 1–5 of the continuation method. Initially the flow is smooth but noisy. As the control parameters are lowered the flow becomes smoother until discontinuities begin to appear; the flow then becomes piecewise-smooth. Note that the diagonal banding visible in the early stages is apparently noise in the image sequence and can also be seen in the results of Heel [25].

The effect of lowering the parameters is seen in Figs. 18 and 19 which show the data and spatial outliers at each stage in the continuation method. Figure 18 shows that the data outliers occur at motion boundaries and in the area of banded noise in the upper right portion of the image. Figure 19 illustrates how spatial outliers are introduced as the value of σ_S is lowered. The final outliers correspond well to the boundary of the Pepsi can.

6.3. The SRI Tree Sequence

The next experiment is more complex in that there are more spatial discontinuities, but the motion is still simple translation parallel to the image plane. The first image in the SRI tree sequence is seen in Fig. 20a. The images are 256 $\times$ 233 pixels, and each of the six stages took approximately 72 s to compute on a SPARCstation 10.

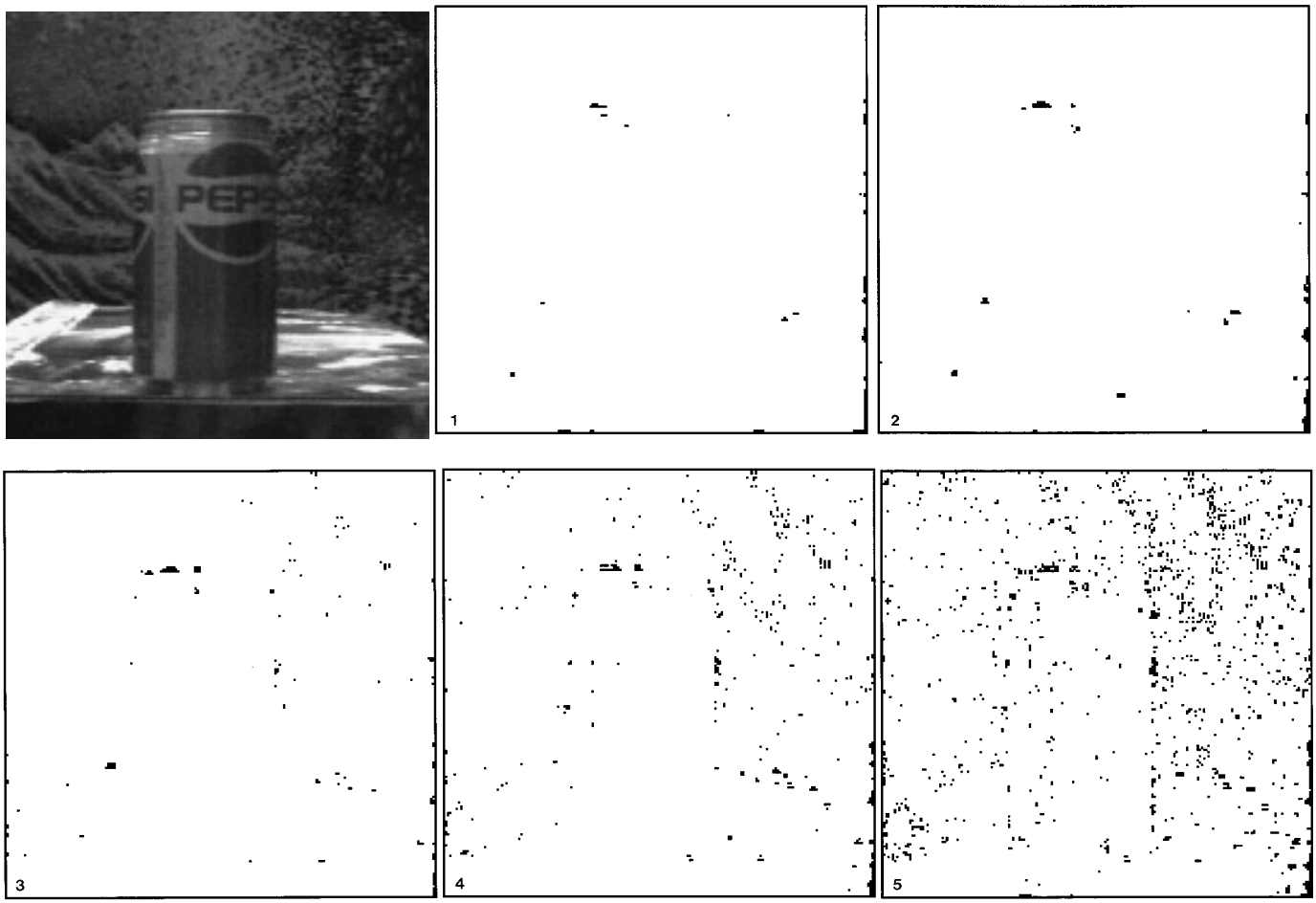

FIG. 18. The Pepsi sequence; data outliers. Results of the continuation method stages 1–5 (see text).

Figure 20 shows the evolution of the horizontal component of the flow field after stages 1–5 of the continuation method. The flow is poor in the early stages but is refined as the control parameters are lowered and discontinuities are introduced. The robust flow exhibits sharp motion boundaries yet still recovers the smoothly varying flow of the ground plane.

We compare this robust solution to the least-squares result in Fig. 21 and, as expected, the least-squares flow estimate (Fig. 21a) shows a good deal of over-smoothing. The least-squares flow was generated using the coarse-to-fine version of the Horn and Schunck algorithm [27].

Finally, Figs. 22 and 23 show the evolution of the data and spatial outliers, respectively, over the six-stage continuation method. The banding present in the data outliers is the result of noise in the image sequence.

6.4. Yosemite Fly-Through

The synthetic Yosemite fly-through sequence tests the accuracy of the robust algorithm on a realistic image sequence with more complex camera motion than the previous experiments. This is a challenging sequence due to aliasing, discontinuities, and the large range of motion (the largest displacement is approximately four pixels). Most published optical flow algorithms have great difficulty with this sequence (see [3] for a detailed analysis of the common techniques applied to this sequence). The images are 316×252 pixels, and each of the six stages took approximately 78 s to compute on a SPARCstation 10.

Figures 24a and 24b show frames 13 and 14 of the image sequence, respectively. The true flow field is shown in Fig. 24c, while Fig. 24d shows the flow field recovered by the robust gradient-based algorithm. Inspection reveals that the recovered flow is visually similar to the original with some errors near the boundaries.

The error in the estimated flow is computed using the angular error measure of Barron *et al.* [3]. They represent image velocities as 3D unit direction vectors

$$\mathbf{v} \equiv \frac{1}{\sqrt{u^2 + v^2 + 1}} (u, v, 1)^{\mathrm{T}}.$$

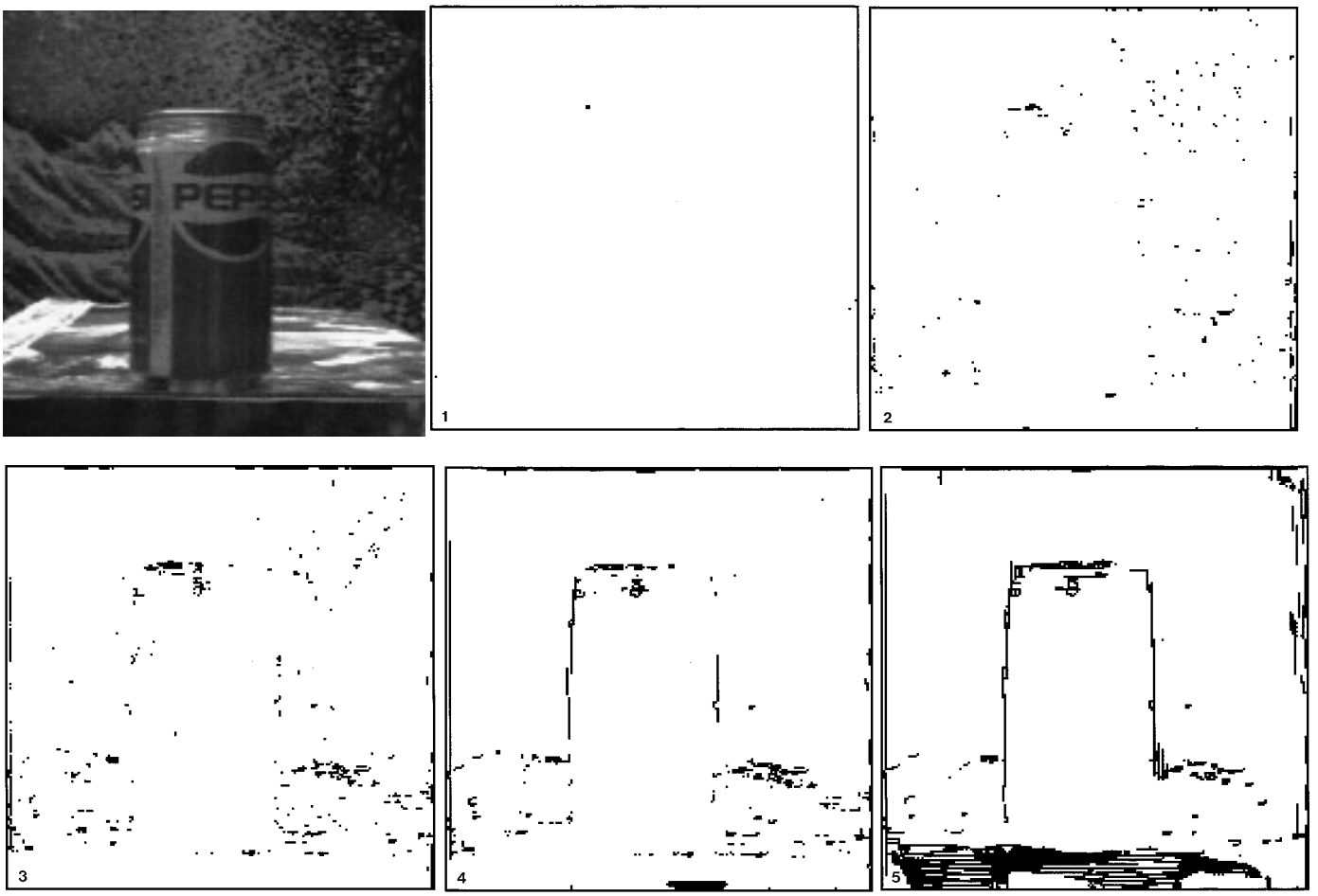

FIG. 19. The Pepsi sequence; spatial outliers. Results of the continuation method stages 1–5 (see text).

The error between the true velocity $\mathbf{v}_t$ and the estimated velocity $\mathbf{v}_e$ is given by arccos $(\mathbf{v}_t \cdot \mathbf{v}_e)$. The performance of the algorithm can be quantified as follows:

Average error	Standard deviation	Percent of flow vectors with error less than:				
		1°	2°	3°	5°	10°
4.46°	4.21°	6%	22%	40%	75%	93%

Note that flow errors were not computed in the sky since unlike the Barron *et al.* images which contained clouds, our images were cloudless.

These results compare favorably with previous approaches. Table 2 lists the results of a number of algorithms applied to this sequence. The numbers reported from [3] are for a version of the Yosemite sequence containing clouds. Barron *et al.* report that if the sky is removed from consideration, the accuracy of the approaches improves by approximately 25%, but the density of the estimates remains effectively unchanged. The numbers are reproduced here to give a rough idea of how the robust approach performs with respect to other common approaches.

The first set of algorithms in the table produce dense flow fields and generally have large average angular errors. The second set of algorithms produce lower angular errors, but do not provide flow estimates everywhere; there is a trade-off of density for accuracy. The robust formulation described here produces errors in the range of the most accurate approaches, but still gives dense estimates.

7. CONCLUSION

If vision research is to make the journey from the carefully controlled environment of the research laboratory to the unconstrained and unpredictable world of humans and commercial applications then vision algorithms must be robust. In this paper, we have considered the issues of robustness related to the recovery of optical flow with multiple motions. In this regard, it is important to recognize the generality of the problems posed by multiple motions; measurements are corrupted whenever information is

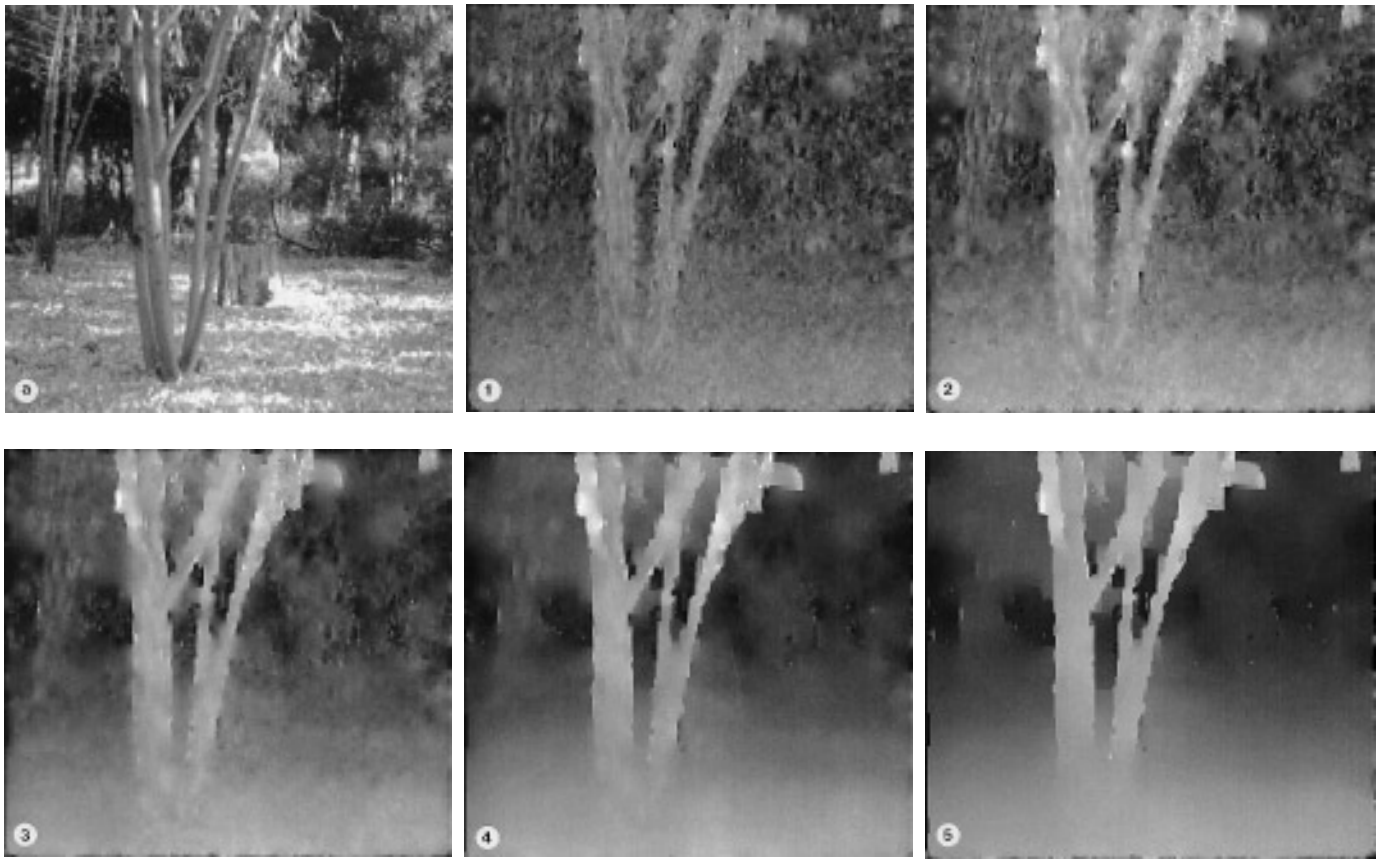

FIG. 20. SRI tree sequence; horizontal flow. Horizontal flow at each stage in the continuation method.

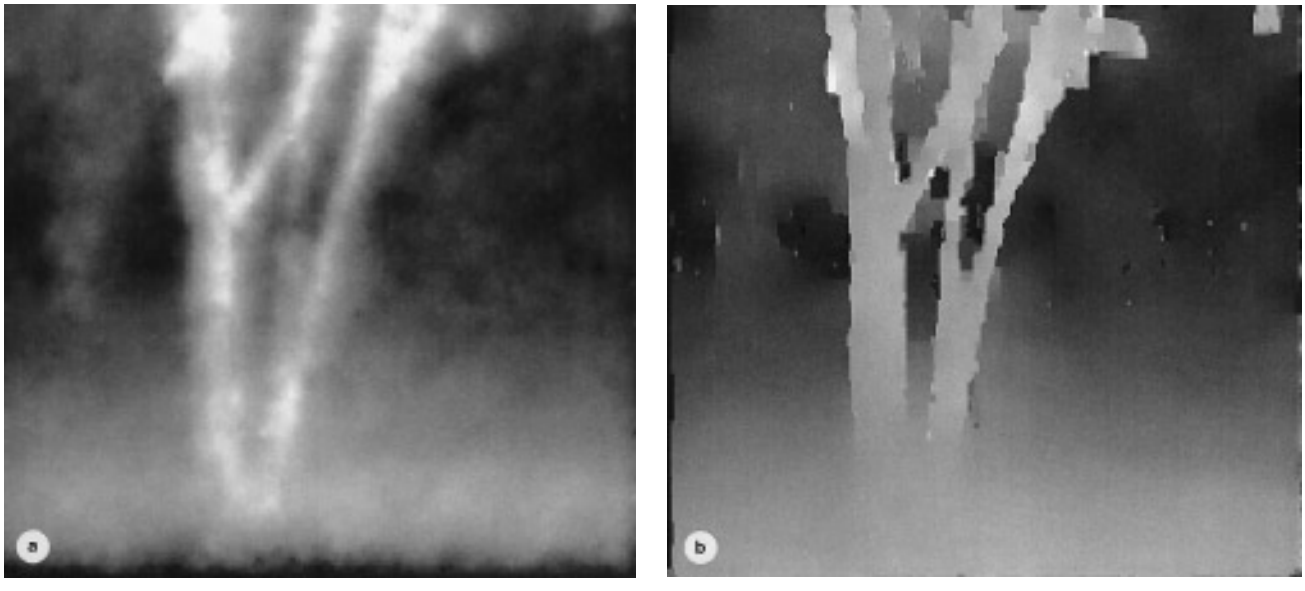

FIG. 21. The SRI tree sequence; horizontal flow. (a) Least-squares flow. (b) Robust gradient flow.

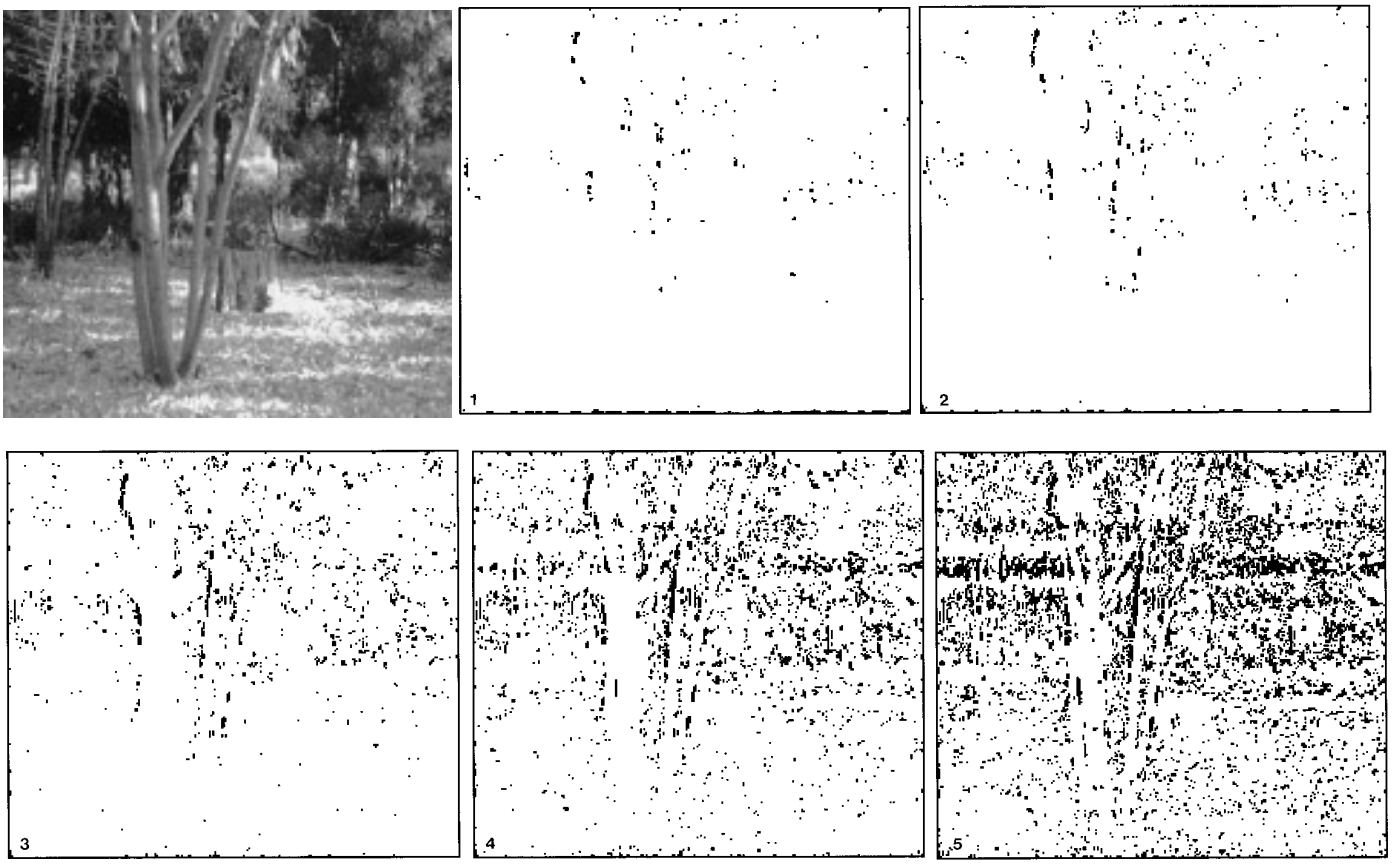

FIG. 22. SRI tree sequence; data outliers. Data outliers are shown at each stage in the continuation method.

pooled from a spatial neighborhood which spans a motion boundary. This applies to both the data conservation and spatial coherence assumptions and violations of these assumptions cause problems for the standard least-squares formulations of optical flow. By recasting these formulations within the robust estimation framework, erroneous measurements at motion boundaries are treated as outliers and their influence is reduced.

The paper demonstrates the application of the robust estimation framework to two common techniques for estimating optical flow. In particular, it shows how area-based regression techniques can be made robust to multiple motions resulting from occlusion, transparency, and specular reflections and how piecewise-smooth flow fields can be recovered using a robust gradient-based algorithm. The approach differs from previous work in that both the data and spatial terms are made robust. This results in the rejection of outlying data constraints which result from noise or multiple motions and prevents over-smoothing across motion boundaries.

The approach has a number of advantages. It allows us to treat the effects of multiple motions on the data conservation and spatial coherence constraints in a uniform manner. The approach provides an alternative interpretation of line processes and weak continuity constraints while generalizing their application to cope with nonspatial out-

TABLE 2
Comparison of Various Optical Flow Algorithms (Adapted from [3])

Technique	Average error	Standard deviation	Density (%)
Horn and Schunck [27]	32.43°	30.28°	100
Anandan [2]	15.84°	13.46°	100
Singh [51]	13.16°	12.07°	100
Nagel [39]	11.71°	10.59°	100
Uras *et al.* [55]	10.44°	15.00°	100
Heeger [24]	10.51°	12.11°	15.2
Fleet and Jepson [17]	4.29°	11.24°	34.1
Lucas and Kanade [34]	4.10°	9.58°	35.1
Weber and Malik [59]	3.42°	5.35°	45.2
Robust formulation	4.46°	4.21°	100

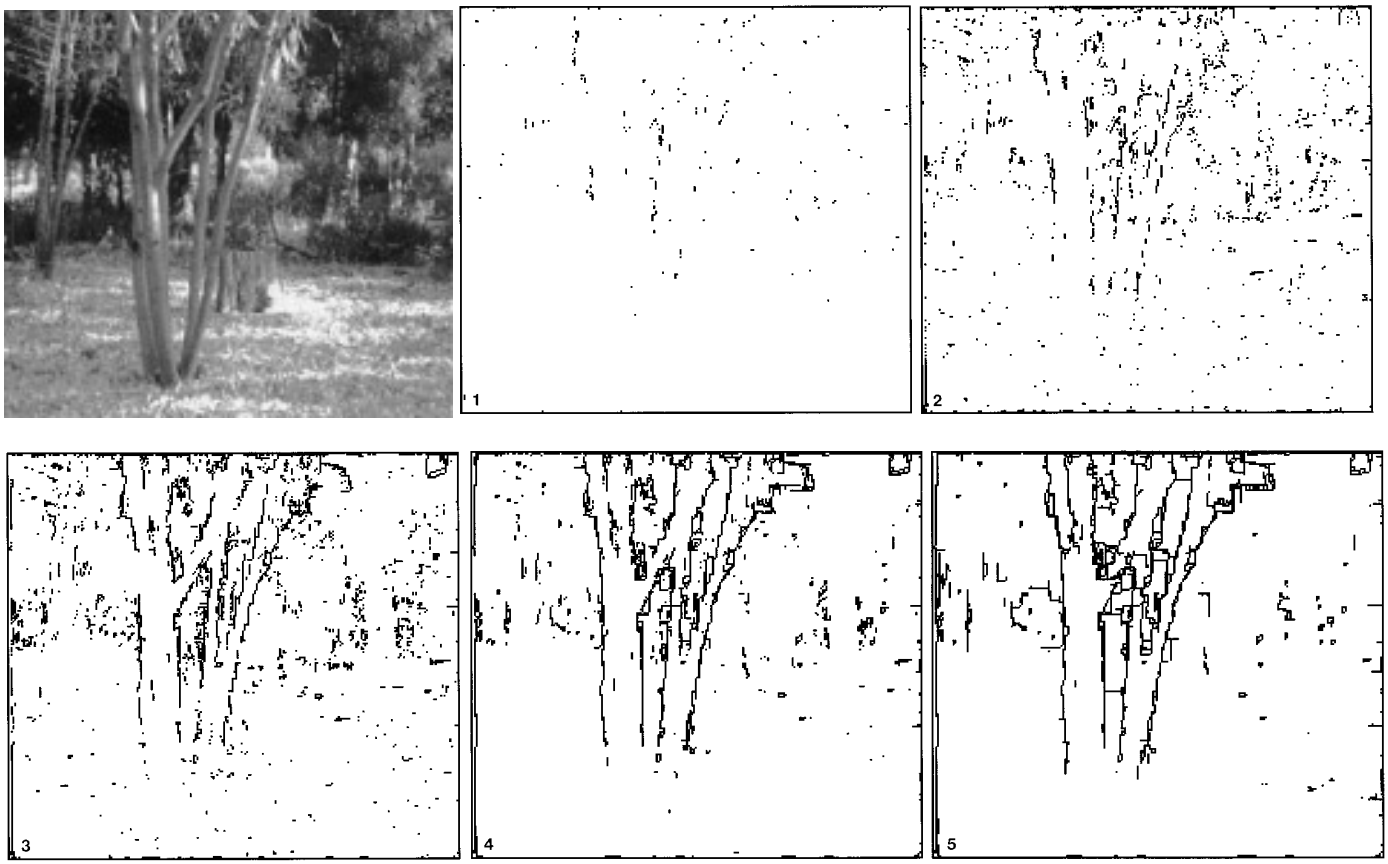

FIG. 23. SRI tree sequence; spatial outliers. Spatial outliers are shown at each stage in the continuation method. As the scale parameter decreases, more discontinuities appear.

liers. The approach allows us to detect model violations and hence to recover motion boundaries. Finally, 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 (see [11] for examples of robust image and surface reconstruction).

APPENDIX A: AFFINE REGRESSION DETAILS

The partial derivatives of the robust regression equation (17) are

$$\frac{\partial E_{\mathrm{D}}}{\partial a_0} = \sum_{\mathcal{R}} I_x \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma)$$

$$\frac{\partial E_{\mathrm{D}}}{\partial a_1} = \sum_{\mathcal{R}} I_x x \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma)$$

$$\frac{\partial E_{\mathrm{D}}}{\partial a_2} = \sum_{\mathcal{R}} I_x y \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma)$$

$$\frac{\partial E_{\mathrm{D}}}{\partial a_3} = \sum_{\mathcal{R}} I_y \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma)$$

$$\frac{\partial E_{\mathrm{D}}}{\partial a_4} = \sum_{\mathcal{R}} I_y x \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma)$$

$$\frac{\partial E_{\mathrm{D}}}{\partial a_5} = \sum_{\mathcal{R}} I_y y \psi((\nabla I)^{\mathrm{T}}\mathbf{u}(\mathbf{a}) + I_t, \sigma).$$

The scale parameters T_{a_i} used in the minimization are just

$$T_{a_0} = I_x^2 \max_x \rho''(x) \qquad T_{a_3} = I_y^2 \max_x \rho''(x)$$

$$T_{a_1} = I_x^2 x^2 \max_x \rho''(x) \qquad T_{a_2} = I_y^2 y^2 \max_x \rho''(x)$$

$$T_{a_4} = I_y^2 x^2 \max_x \rho''(x) \qquad T_{a_5} = I_y^2 y^2 \max_x \rho''(x).$$

For the Lorentzian estimator, we have

$$\psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2} \quad \text{and} \quad \max_x \rho''(x, \sigma) = \frac{1}{\sigma^2},$$

and for the Geman–McClure estimator,

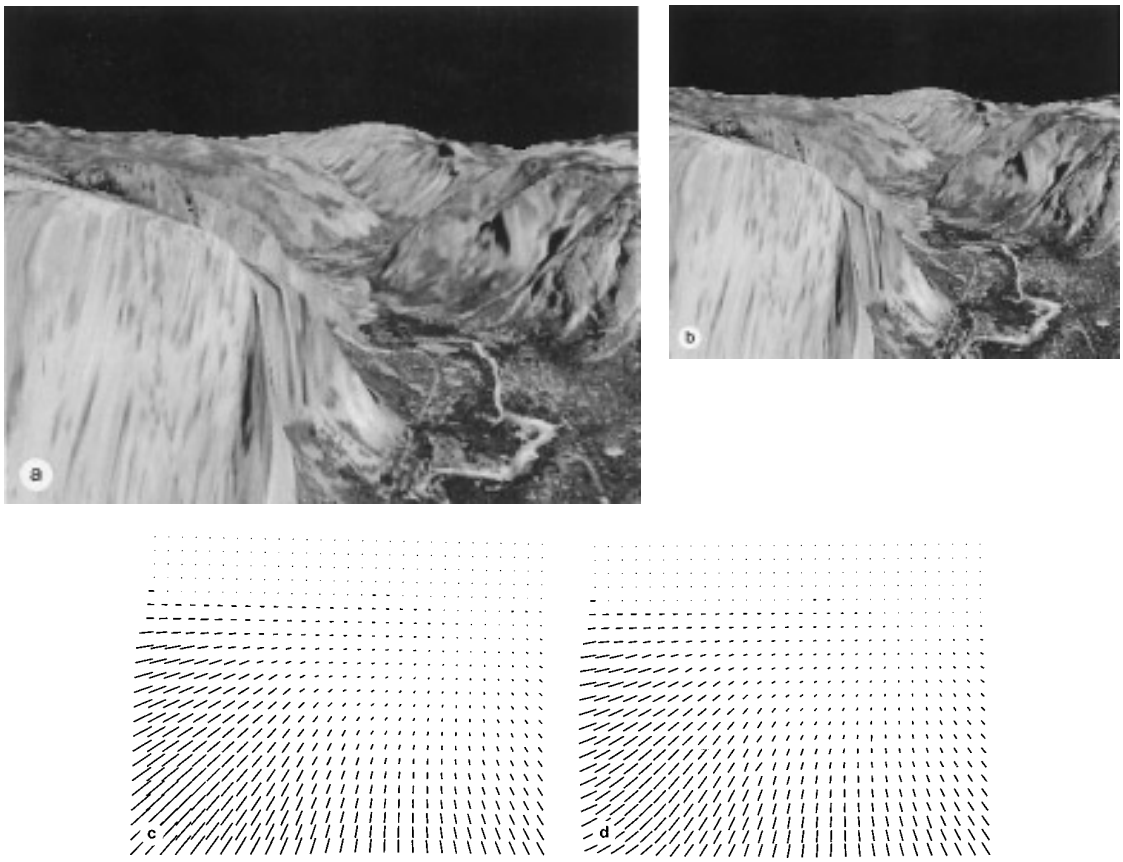

FIG. 24. Yosemite sequence results. (a, b) Iages 13 and 14, respectively. (c) Actual flow field. (d) Recovered flow field.

$$\psi(x, \sigma) = \frac{2x\sigma}{(\sigma + x^2)^2} \quad \text{and} \quad \max_x \rho''(x, \sigma) = \frac{2}{\sigma^2}.$$

APPENDIX B: AFFINE WARPING

The affine parameters estimated using the robust regression approach describe the motion from the first image toward the second. To warp the first image, we must take each pixel (x, y) in the warped image and ask "where did this pixel come from in the first image?". This "backward warp" gives us a location $(\hat{x}, \hat{y})$ in the first image. We then linearly interpolate the pixels in the neighborhood of $(\hat{x}, \hat{y})$ to find the brightness value at (x, y).

Let (c_x, c_y) be the coordinates of the center of the image region. Given the affine motion from $(\hat{x}, \hat{y})$ to (x, y),

$$\begin{aligned}(x, y)^{\mathrm{T}} &= (\hat{x}, \hat{y})^{\mathrm{T}} + (a_0, a_3)^{\mathrm{T}} + A * (\hat{x} - c_x, \hat{y} - c_y)^{\mathrm{T}} \\ &= (c_x, c_y)^{\mathrm{T}} + (a_0, a_3)^{\mathrm{T}} + (I + A) * (\hat{x} - c_x, \hat{y} - c_y)^{\mathrm{T}},\end{aligned} \tag{30}$$

where

$$A = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix}.$$

Inverting this affine transformation gives

$$(\hat{x} - c_x, \hat{y} - c_y)^{\mathrm{T}} = (I + A)^{-1}(x - c_x, y - c_y)^{\mathrm{T}} - (a_0, a_3)^{\mathrm{T}}. \tag{31}$$

Rearranging terms, we get

$$(\hat{x}, \hat{y})^{\mathrm{T}} = (x, y)^{\mathrm{T}} + (b_0, b_3)^{\mathrm{T}} + B * (x - c_x, y - c_y)^{\mathrm{T}}, \tag{32}$$

where

$$B = (I + A)^{-1} - I \tag{33}$$

$$(b_0, b_3)^{\mathrm{T}} = -(I + B) * (a_0, a_3)^{\mathrm{T}}. \tag{34}$$

ACKNOWLEDGMENTS

Portions of this work were performed by the first author at Yale University, the NASA Ames Research Center—Aerospace Human Factors

Research Division, and the University of Toronto with support from the Office of Naval Research (N00014-91-J-1577), the National Aeronautics and Space Administration (NGT-50749), the Natural Sciences and Engineering Research Council of Canada, and the Connaught Fund. We thank Trevor Darrell for providing the plant sequence, Jim Bergen for the skaters sequence, Bonivar Shridhar for the NASA sequence, Joe Heel for the Pepsi sequence, Bob Bolles and Harlyn Baker for the SRI tree sequence, and Lynn Quam, who generated the Yosemite fly-through sequence. We also thank the anonymous reviewers, Allan Jepson, and David Fleet for their comments.

REFERENCES

1. G. Adiv, Recovering motion parameters in scenes containing multiple moving objects, in *Proc. Int. Conf. on Pattern Recognition, Washington, DC, June 1983,* pp. 399–400.
2. P. Anandan, A computational framework and an algorithm for the measurement of visual motion, *Int. J. Comput. Vision* **2,** 1989, 283–310.
3. J. L. Barron, D. J. Fleet, and S. S. Beauchemin, *Performance of Optical Flow Techniques,* Technical Report RPL-TR-9107, Univ. of Western Ontario, July 1993.
4. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, Hierarchical model-based motion estimation, in *Proc. of Second European Conference on Computer Vision, ECCV-92* (G. Sandini, Ed.), Lecture Notes in Computer Science, Vol. 588 pp. 237–252, Springer-Verlag, Berlin/New York, 1992.
5. J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg, A three-frame algorithm for estimating two-component image motion, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(9), Sept. 1992, 886–896.
6. M. J. Black, *Robust Incremental Optical Flow,* Ph.D. thesis, Yale Univ., New Haven, CT, 1992; Research Report YALEU/DCS/RR-923.
7. M. J. Black and P. Anandan, Constraints for the early detection of discontinuity from motion, in *Proc. National Conf. on Artificial Intelligence, AAAI-90, Boston, MA, 1990,* pp. 1060–1066.
8. M. J. Black and P. Anandan, A model for the detection of motion over time, in *Proc. Int. Conf. on Computer Vision, ICCV-90, Osaka, Japan, Dec. 1990,* pp. 33–37.
9. M. J. Black and P. Anandan, Robust dynamic motion estimation over time, in *Proc. Computer Vision and Pattern Recognition, CVPR-91, Maui, HI, June 1991,* pp. 296–302.
10. M. J. Black and P. Anandan, A framework for the robust estimation of optical flow, in *Proc. Int. Conf. on Computer Vision, ICCV-93, Berlin, Germany, May 1993,* pp. 231–236.
11. M. J. Black and A. Rangarajan, The outlier process: Unifying line processes and robust statistics, in *Proc. Computer Vision and Pattern Recognition, CVPR-94, Seattle, WA, June 1994,* pp. 15–22.
12. A. Blake and A. Zisserman,*Visual Reconstruction.* The MIT Press, Cambridge, MA, 1987.
13. P. J. Burt, J. R. Bergen, R. Hingorani, R. Kolczynski, W. A. Lee, A. Leung, J. Lubin, and H. Shvaytser, Object tracking with a moving camera: An application of dynamic motion analysis, in *: Proceedings of the Workshop on Visual Motion, Irvine, CA, Mar. 1989,* pp. 2–12.
14. N. Cornelius and T. Kanada, Adapting optical flow to measure object motion in reflectance and X-ray image sequences, in *Proc. ACM Siggraph/Sigart Interdisciplinary Workshop on Motion: Representation and Perception, Toronto, Ont., Canada, Apr. 1983,* pp. 50–58.
15. T. Darrell and A. Pentland, *Discontinuity Models and Multi-Layer Description Networks.* Technical Report 162, MIT Media Lab Vision and Modeling Group, May 1991.
16. C. L. Fennema and W. B. Thompson, Velocity determination in scenes containing several moving objects, *Comput. Graphics Image Process.* **9,** 1979, 301–315.
17. D. J. Fleet and A. D. Jepson, Computation of component image velocity from local phase information, *Int. J. Comput. Vision* **5,** 1990, 77–104.
18. D. Geiger and F. Girosi, Parallel and deterministic algorithms from MRFs: Surface reconstruction, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(5), May 1991, 401–412.
19. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**(6), Nov. 1984, 721–741.
20. S. Geman and D. E. McClure, Statistical methods for tomographic image reconstruction, *Bull. Int. Statist. Inst.* **LII-4,** 1987, 5–21.
21. F. Glazer, *Hierarchical Motion Detection,* Ph.D. thesis, Univ. of Massachusetts, Amherst, MA, 1987; COINS TR 87-02.
22. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions,* Wiley, New York, 1986.
23. J. G. Harris, C. Koch, E. Staats, and J. Luo, Analog hardware for detecting discontinuities in early vision, *Int. J. Comput. Vision* **4**(3), June 1990, 211–223.
24. D. J. Heeger, Model for the extraction of image flow, *J. Opt. Soc. Am.* **4**(8), Aug. 1987, 1455–1471.
25. J. Heel, Temporally integrated surface reconstruction, in *Proc. IEEE Int. Conf. on Comp. Vision, ICCV-90, Osaka, Japan, Dec. 1990,* pp. 292–295.
26. B. K. P. Horn, *Robot Vision,* The MIT Press, Cambridge, MA, 1986.
27. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Artif. Intell.* **17**(1–3), Aug. 1981, 185–203.
28. P. J. Huber, *Robust Statistics,* Wiley, New York, 1981.
29. M. Irani, B. Rousso, and S. Peleg, Detecting and tracking multiple moving objects using temporal integration, in *Proc. of Second European Conference on Computer Vision, ECCV-92* (G. Sandini, Ed.), Lecture Notes in Computer Science, Vol. 588, pp. 282–287, Springer-Verlag, Berlin/New York, 1992.
30. A. Jepson and M. J. Black, Mixture models for optical flow computation, in *Proc. Computer Vision and Pattern Recognition, CVPR-93, New York, June 1993,* pp. 760–761.
31. J. K. Kearney, W. B. Thompson, and D. L. Boley, Optical flow estimation: An error analysis of gradient-based methods with local optimization, *IEEE Trans. Pattern Anal. Mach. Intell.* **9,** 1987, 229–244.
32. J. Konrad and E. Dubois, Multigrid Bayesian estimation of image motion fields using stochastic relaxation, in *Int. Conf. on Computer Vision, 1988,* pp. 354–362.
33. Y. G. Leclerc, Constructing simple stable descriptions for image partitioning, *Int. J. Comput. Vision* **3**(1), 1989, 73–102.
34. B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, in *Proc. 7th IJCAI, Vancouver, B. C., Canada, 1981,* pp.674–679.
35. S. Madarasmi, D. Kersten, and T. C. Pong, Multi-layer surface segmentation using energy minimization, in *Proc. Computer Vision and Pattern Recognition, CVPR-93, New York, June 1993,* pp. 774–775.
36. J. Marroquin S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *J. Am. Stat. Assoc.* **82**(397), Mar. 1987, 76–89.
37. P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, Robust regression methods for computer vision: A review, *Int. J. Comput. Vision* **6**(1), 1991, 59–70.

38. D. W. Murray and B. F. Buxton, Scene segmentation from visual motion using global optimization, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9**(2), Mar. 1987, 220–228.

39. H. H. Nagel, On the estimation of optical flow: Relations between different approaches and some new results, *Artif. Intell.* **33**(3), Nov. 1987, 299–324.

40. H. H. Nagel and W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**(5), Sept. 1986, 565–593.

41. S. Negahdaripour and C. Yu, A generalized brightness change model for computing optical flow, in *Proc. Int. Conf. on Computer Vision, ICCV-93, Berlin, Germany, May 1993,* pp. 2–11.

42. M. Okutomi and T. Kanada, A locally adaptive window for signal matching, *Int. J. Comput. Vision* **7**(2), Jan. 1992, 143–162.

43. P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), July 1990, 629–639.

44. T. Poggio, V. Torre, and C. Koch, Computational vision and regularization theory, *Nature* **317**(26), Sept. 1985, 314–319.

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

46. A. Rangarajan and R. Chellappa, A continuation method for image estimation using the adiabatic approximation, in *Markov Random Fields: Theory and Applications* (R. Chellappa and A. Jain, Eds.), Academic Press, San Diego, 1993.

47. P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection,* Wiley, New York, 1987.

48. B. G. Schunck, Image flow segmentation and estimation by constraint line clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(10), Oct. 1989, 1010–1027.

49. M. Shizawa and K. Mase, Principle of superposition: A common computational framework for analysis of multiple motion, in *Proc. IEEE Workshop on Visual Motion, Princeton, NJ, Oct. 1991,* pp. 164–172.

50. D. Shulman and J. Hervé, Regularization of discontinuous flow fields, in *Proc. Workshop on Visual Motion, Irvine, CA, Mar. 1989,* pp. 81–85, IEEE Comput. Soc. Press, Los Alamitos, CA, 1989.

51. A. Singh, *Optical Flow Computation: A Unified Perspective,* IEEE Comput. Soc. Press, Los Alamitos, CA, 1992.

52. G. Strang, *Linear Algebra and Its Applications,* Academic Press, New York, 1976.

53. W. B. Thompson, K. M. Mutch, and V. A. Berzins, Dynamic occlusion analysis in optical flow fields, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**(4), July 1985, 374–383.

54. Y. Tian and M. Shah, *MRF-Based Motion Estimation and Segmentation,* Technical Report CS-TR-92-13, Univ. of Central Florida, Orlando, FL, July 1992.

55. S. Uras, F. Girosi, A. Verri, and V. Torre, A computational approach to motion perception, *Biol. Cybernet.* **60,** 1989, 79–97.

56. R. S. Varga, *Matrix Iterative Analysis,* Prentice–Hall, Englewood Cliffs, NJ, 1962.

57. J. Y. A. Wang and E. H. Adelson, Layered representation for motion analysis, in *Proc. Computer Vision and Pattern Recognition, CVPR-93, New York, June 1993,* pp. 361–366.

58. A. M. Waxman and K. Wohn, Contour evolution, neighbourhood deformation and global image flow: Planar surface in motion, *Int. J. Robotics Res.* **4,** 1985, 95–108.

59. J. Weber and J. Malik, Robust computation of optical flow in a multiscale differential framework, in *Proc. Int. Conf. on Computer Vision, ICCV-93, Berlin, Germany, May 1993,* pp. 12–20.