Estimating Optical Flow in Segmented Images using Variable-order Parametric Models with Local Deformations

Michael J. Black* and **Allan Jepson**†

*Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304
Phone: (415)812–4745, FAX: (415)812–4334, email: black@parc.xerox.com

†Department of Computer Science, University of Toronto, Toronto, Ontario M5S 1A4, Canada
Phone: (416)978–6488, FAX: (416)978–1455, email: jepson@vis.toronto.edu

Submitted: July 1994
Revised: June 1996

Abstract

This paper presents a new model for estimating optical flow based on the motion of planar regions plus local deformations. The approach exploits brightness information to organize and constrain the interpretation of the motion by using segmented regions of piecewise smooth brightness to hypothesize planar regions in the scene. Parametric flow models are estimated in these regions in a two step process which first computes a coarse fit and estimates the appropriate parameterization of the motion of the region (two, six, or eight parameters). The initial fit is refined using a generalization of the standard area-based regression approaches. Since the assumption of planarity is likely to be violated, we allow local deformations from the planar assumption in the same spirit as physically-based approaches which model shape using coarse parametric models plus local deformations. This parametric+deformation model exploits the strong constraints of parametric approaches while retaining the adaptive nature of regularization approaches. Experimental results on a variety of images indicate that the parametric+deformation model produces accurate flow estimates while the incorporation of brightness segmentation provides precise localization of motion boundaries.

To appear IEEE PAMI.

1 Introduction

Estimating the optical flow in scenes containing significant depth variation, independent motion, or articulate objects necessitates the segmentation of the scene into regions of coherent motion. If the scene were segmented into roughly planar surface patches then the motion of each surface patch could be estimated using a parametric flow model. Given large numbers of constraints computed within the patch and a small number of parameters to be estimated, these parametric models provide strong constraints on the motion within a region resulting in accurate flow estimates. In contrast to recent parametric approaches which assume that an arbitrary image region can be modeled by a single motion, we independently model the motion of segmented planar surface regions. But segmentation is a hard problem in its own right and, in particular, the recovery of segmented, or piecewise smooth, flow fields is notoriously difficult. Instead, this paper makes the simple hypothesis that *image regions of piecewise smooth brightness are likely to correspond to surfaces in the world.* These brightness regions are assumed to be planar surfaces in the scene and their motion is estimated using a variable-order parametric flow model containing two, six, or eight parameters. In this way, information about image brightness is used to organize and constrain the interpretation of the optical flow. Since the assumption of planarity may be violated, we allow local deformation from the planar assumption in the same spirit as physically-based approaches which model shape using coarse parametric models plus deformations. The resulting model, in which optical flow is represented by the motion of planar image patches with local deformations, exploits the strong constraints of parametric approaches while retaining the adaptive nature of regularization approaches. Experiments with natural and synthetic image sequences indicate that the parametric+deformation model produces accurate flow estimates while the incorporation of brightness segmentation provides precise localization of motion boundaries.

The algorithm can be thought of as having low- and medium-level processing. At the low level there is a process which is always smoothing the image brightness while accounting for brightness discontinuities. There is another low-level process that is always providing coarse estimates of image motion. The medium level tries to organize and make sense of the low-level data by first finding connected regions of piecewise smooth brightness and then by estimating the motion of these regions. This process is illustrated in Figure 1. This medium-level motion-estimation process has

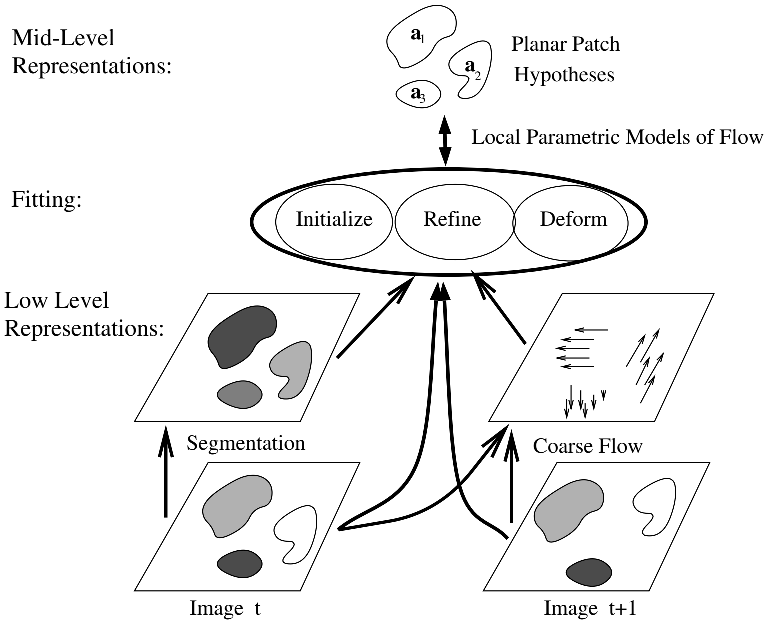


Figure 1: Medium-level processes exploit structure in the image brightness to interprate the coarse optical flow estimates in terms of a small number of parameters.

three steps. The first fits a parametric model to the coarse motion estimates in each region to provide an initial estimate of the image motion. A variable-order fitting procedure is used to estimate the appropriate model (translational, affine, or planar) which best captures the image motion in each region. In the second step, the parametric fit from the initial estimate is used to warp the image regions into alignment. Gradient-based optical flow constraints are computed from these registered regions and are used to refine the initial parametric fit by performing regression over each region. Robust regression techniques [12, 19] are used to compute both the initial and refined estimates if the motion parameters. Finally, the planar patches are allowed to deform at the low-level subject to weak constraints from the optical flow constraints, the spatial coherence of the neighboring flow estimates, and the motion estimate for the planar patch.

The following section reviews previous work on parametric flow models, segmented flow fields, and combining brightness and motion information. Section 3 briefly describes the low-level segmentation and motion estimation processes. The medium-level processing, including the initial fitting and refinement of the parametric motion models within the segmented regions, is described in Section 4. Section 5 describes the full model with planar patches plus local deformations. Examples are provided throughout the text and additional experiments with synthetic and natural images are described in Section 6. Open issues and future directions are addressed in Section 7.

2 Previous Work

2.1 Parametric Models of Image Motion

Parametric models of optical flow within an image region provide both a concise representation and enforce strong constraints on the interpretation of the motion. These techniques use regression or a Hough transform to estimate a few parameters (eg. two, six, or eight) given hundreds or thousands of constraints computed over the entire image or some pre-selected region [6, 16, 17, 25, 31, 49]; when the image motion conforms to the model assumptions this produces accurate flow estimates. The problem with this approach is that parametric motion models applied over the entire image or arbitrary, pre-selected, regions are rarely valid in real scenes due to surfaces at varying depths, transparency, or the independent motion of objects.

Approaches have been devised which ameliorate some of the problems of global parametric models. Bergen *et al.* [7] use an iterative registration algorithm to account for multiple global motions in the scene. Jepson and Black [23] assume that the motion in the scene can be represented by a mixture of distributions and they use the EM algorithm to decompose the motion into a fixed number of layers. In similar work, Darrell and Pentland [14] use a stochastic approach to segment the motion into a set of layers with support maps which assign pixels to layers. Additionally they use a minimum description length encoding principle to automatically choose the appropriate number of layers. Black and Anandan [12] use robust statistics to estimate a dominant motion in the scene and then fit additional motions to outlying measurements. All of these approaches are formulated as global techniques which can cope with a small number of global motions but not with general flow fields. As global approaches, they do not address how to select appropriate image regions in which to apply the parametric models.

In related work, Irani and Peleg [22] fit a dominant motion to the scene using a least squares method and they detect outlying measurements which are grouped together and segmented. These groups hopefully correspond to independently moving objects and their motion is estimated independently. This approach begins to deal with the issue of estimating the motion of segmented regions but, like the other global parametric approaches it assumes a single dominant motion and a small number of "outlying" motions.

Meyer and Bouthemy [35] use a motion segmentation technique to extract regions corresponding to independently moving objects. The optical flow within these regions is then modeled and estimated using a parametric flow model (eg. affine). The regions and their boundaries are tracked and updated over time. Like our approach, they use parameterized motion models within segmented regions but unlike our method they use motion rather than brightness information to extract these regions.

Another set of approaches apply parametric models to coarse flow fields by grouping the flow vectors into consistent regions. Adiv [1] uses a Hough technique to group flow measurements into regions consistent with the motion of planar surfaces. The approach of Wang and Adelson [46] is similar but uses a k-means clustering algorithm to group the flow vectors into layers of consistent affine motion. These approaches, like the regression approaches, are essentially global techniques in that they assume the image motion can be represented by a small number of global layers. Additionally they fail to exploit information present in the image brightness about the nature of surfaces in the scene.

2.2 Exploiting Image Brightness

To improve motion segmentation a number of researchers have attempted to combine intensity and motion information. Thompson [44] describes a region merging technique which uses similarity constraints on brightness and motion for segmentation. Heitz and Bouthemy [20] combine gradient-based and edge-based motion estimation and realize improved motion estimates and the localization of motion discontinuities. Black [9] jointly estimates piecewise smooth motion and brightness over an image sequence. Discontinuities are detected using motion and brightness simultaneously and are classified as either structural boundaries or surface markings. Recently, motion segmentation and color segmentation have been combined to improve the localization of moving object contours [15]. In focusing on motion boundaries these approaches use weak models of optical flow (eg. regularization) and hence neglect one of the benefits of having a segmentation in the first place; that is, that the motion of a segmented region can often be described using a simple parametric model which allows many constraints to be integrated across the region.

There are numerous feature-based schemes which estimate motion by tracking points, edges, or region contours computed from the brightness image (eg. [47]). Sull and Ahuja [42] estimate the motion of region boundaries and follow this with a Hough technique that groups the regions into planar surfaces. These approaches use information about image brightness to constrain the motion estimation, but brightness contours alone are an impoverished representation. The motion information available over an entire region, particularly if it is reasonably textured, provides additional constraints which can improve the accuracy of the recovered motion.

In the context of stereo reconstruction, Luo and Maître [32] use a segmented intensity image to correct and improve disparity estimates by fitting a plane to the disparities within a region of uniform brightness. The accuracy of this approach is affected by the accuracy of the initial disparity estimates. Koch [26] segments regions using disparity and brightness and then regularizes depth estimates within the regions. While this approach preserves depth boundaries it uses a weak model within regions instead of fitting a model with a small number of parameters.

Ayer *et al.* [4] describe a method with similar motivations to the one presented here in that they combine static segmentation with motion information. They first robustly estimate a global parametric motion for the scene and detect regions which do not match this motion. The parametric motions of these outlying regions are then estimated. Motion estimation is performed using a multi-frame approach. Then, given a static segmentation, the computed motion information is used to label the static regions. For a given static region, there may be multiple possible motion estimates obtained using the robust motion segmentation and estimation procedure. For each static region the error of using each of these parametric motions is evaluated. The regions are finally labeled with the motion parameters that give the lowest error.

2.3 The Proposed Method

The approach described here is similar in its first stage to that of [32] in that coarse flow estimates are computed and then parametric models are fit to the estimates within the segmented brightness regions. This process significantly improves the coarse motion estimates but we use this only as an initialization step. The motion of the regions is refined directly using brightness constraints from the images in a generalization of the standard global regression approaches [6]. Unlike the approach of Ayer *et al.* [4] we estimate the motion directly in these segmented regions and do not attempt to perform segmentation based on motion. Finally, we treat the assumption of planar patches as a coarse approximation and allow local deformations to the motion estimates using an energy minimizing approach. This is similar to work which uses superquadrics to compute a coarse parametric description of 3D shape and then allows local deformations to account for fine structure [38, 43]. It is also related to work on decomposing rigid image motion into the motion of a plane plus residual motion parallax [40].

3 Early Processing

At the low level there are two processes which examine the input images: segmentation and coarse motion estimation. The exact methods used for these early processes are not crucial to the optical flow model described in this paper, so the algorithms are described only briefly and the reader is referred to [13] for a complete description of the segmentation approach and to [11] for the coarse flow estimation. The static image segmentation method described below is just one of many possibilities. Any other method that gives connected regions could be employed and the better the static segmentation results, the better the motion estimates will be. We choose this method to provide examples which illustrate the interplay between brightness segmentation and motion estimation.

3.1 Segmentation

For the experiments described here we have used a weak-membrane model of image brightness described in [13]. The goal is to reconstruct a piecewise smooth brightness image $\mathbf{i}$ given noisy data $\mathbf{d}$ by minimizing an objective function using a continuation method. Both spatial discontinuities and texture are treated as outlying measurements and rejected using analog outlier processes.

Assume that the data is an $n \times n$ image of sites S, and each site (or pixel), $s \in S$, has a set of neighbors $t \in \mathcal{G}_s$. For a first-order neighborhood system, $\mathcal{G}_s$, these are just the sites to the North, South, East, and West of site s. We also define a dual lattice, $\mathbf{l}$, of all nearest neighbor pairs (s, t) in S. This lattice is coupled to the original in such a way that the best interpretation of the data will be one in which the data is piecewise smooth. An analog spatial outlier process $l_{s,t} \in \mathbf{l}$ takes on values $0 \leq l_{s,t} \leq C$, for some positive constant C (for the remainder of the paper we take $C = 1$). The outlier process indicates the presence $(l_{s,t} \to 0)$ or absence $(l_{s,t} \to 1)$ of a discontinuity between neighboring sites s and t. We also define a penalty $0 \leq \Psi(l_{s,t}) \leq \infty$ which is paid for introducing a discontinuity. The penalty function goes to infinity as $l_{s,t}$ goes to 0 (that is, we pay an infinite penalty for introducing a complete discontinuity) and $\Psi(l_{s,t}) \to 0$ when there is no discontinuity ($l_{s,t} \to 1$).[1] For these experiments we take

$$\Psi(z) = z - 1 - \log z,$$

which is derived from the Lorentzian error norm [13]. Additionally, we introduce a measurement outlier process $m_s \in \mathbf{m}$ on the data term which treats image texture as outliers with respect to the piecewise smooth reconstruction.

The approach taken is to minimize the following objective function composed of a data term and a spatial coherence term

$$\begin{aligned} E_I(\mathbf{i}, \mathbf{d}, \mathbf{l}, \mathbf{m}) \quad = \quad & \sum_{s \in S} \left(\frac{1}{2\sigma_D^2}(i_s - d_s)^2 \, m_s + \Psi_D(m_s) \right. \\ & \left. + \frac{1}{|\mathcal{G}_s|} \sum_{t \in \mathcal{G}_s} [\frac{1}{2\sigma_S^2}(i_s - i_t)^2 \, l_{s,t} + \Psi_S(l_{s,t})] \right), \end{aligned} \tag{1}$$

where the σ_* are scale (or control) parameters, and $|\mathcal{G}_s|$ is the size of the neighborhood.

The objective function is minimized using an algorithm similar to the EM-algorithm [34] in which at each iteration we solve for the m_s and $l_{s,t}$ in closed form and then update the i_s using one step of Newton's method. The initial estimate for $\imath$ is just taken to be **d**. The minimization process is embedded in a continuation method in which the value of the σ_* are lowered according to a scale factor; this has the effect of tracking the solution as the function becomes increasingly non-convex.

The approach is applied to a pair of images in synthetic Yosemite sequence[2], the first of which is shown in Figure 2*a*. For this experiment σ_D started at $25.0/\sqrt{2}$ and was lowered to $10.0/\sqrt{2}$ while σ_S started at $10.0/\sqrt{2}$ and was lowered to $2.0/\sqrt{2}$. In practice we have found that a simple two stage continuation method produces adequate results with 30 iterations of Newton's method at each stage. Figure 2*b* shows the piecewise smooth reconstruction **i** while Figures 2*c* and 2*d* show the value of the data and spatial outlier processes respectively (black indicates an outlier). The spatial outliers will be used for region segmentation at the medium level.[3]

[1] We could also choose a penalty function such that $\Psi(l_{s,t}) \to 1$ as $l_{s,t} \to 0$ (see [13]).

[2] This sequence was generated by Lynn Quam and provided by David Heeger.

[3] The approach described here is equivalent to minimizing

$$E_I(\mathbf{i}, \mathbf{d}) = \sum_{s \in S} [\rho(i_s - d_s, \sigma_D) + \frac{1}{|\mathcal{G}_s|} \sum_{t \in \mathcal{G}_s} \rho(i_s - i_t, \sigma_S)]$$

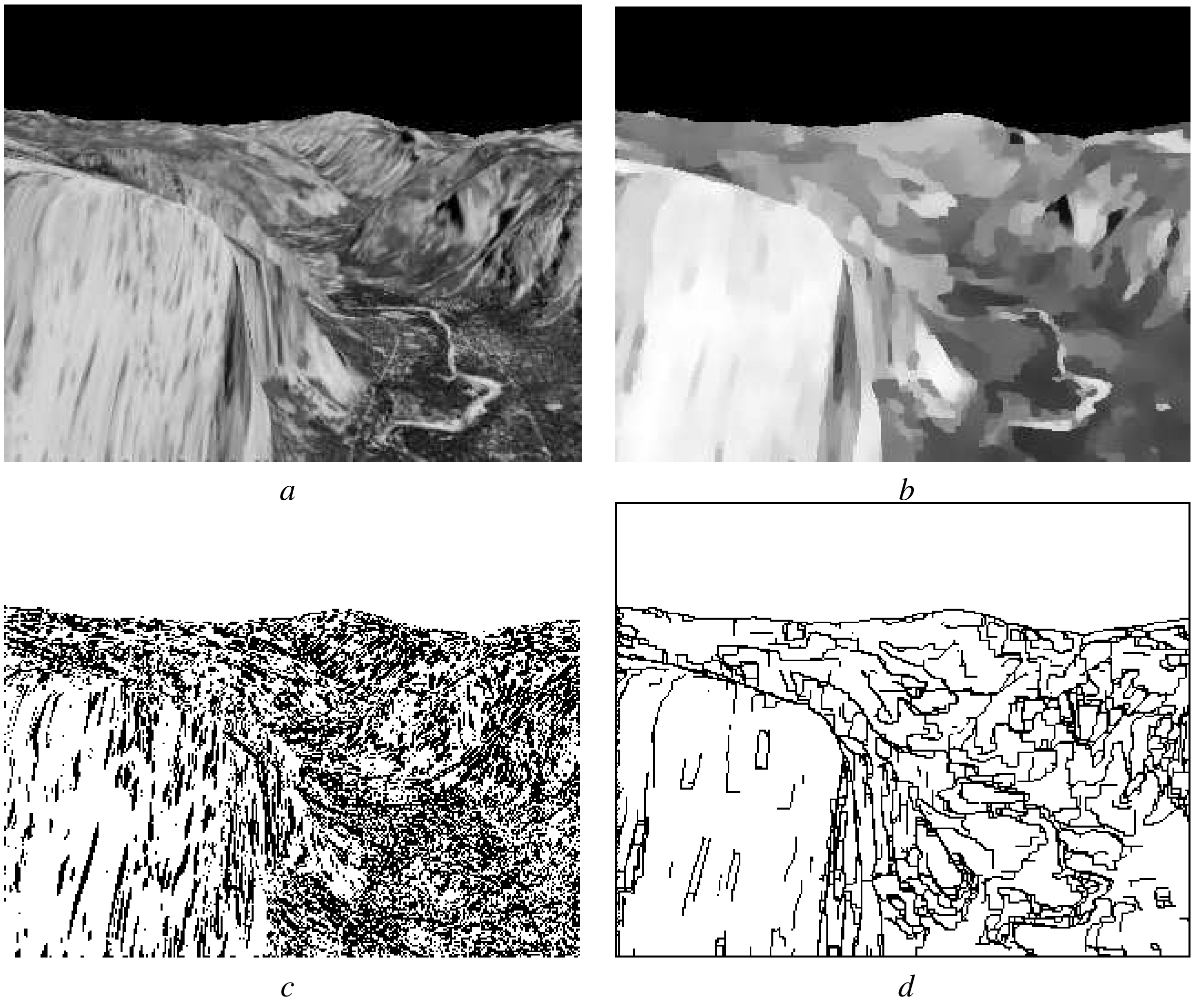


Figure 2: Yosemite Sequence. *(a)* Image 11 in the sequence ($\mathbf{d}$); *(b)* Piecewise smooth reconstruction ($\mathbf{i}$); *(c)* Data outliers (**m** thresholded at 0.5); *(d)* Spatial outliers (**l** thresholded at 0.5).

3.2 Coarse Optical Flow

Let $I(x, y, t)$ be the image brightness at a point (x, y) at time t and I_x, I_y, and I_t be the partial derivatives of I with respect to x, y, and t. To estimate the horizontal and vertical image velocity $\mathbf{u}(\mathbf{x}) = [u(\mathbf{x}), v(\mathbf{x})]^T$ at a point $\mathbf{x} = (x, y)$ we minimize an objective function composed of a data term and a spatial smoothness term [11]:

$$E_M(\mathbf{u}) = \sum_{\mathbf{x}} [\lambda_D \rho((\nabla I(\mathbf{x})\mathbf{u}(\mathbf{x}) + I_t(\mathbf{x})), \sigma_D) + \frac{\lambda_S}{|\mathcal{G}(\mathbf{x})|} \sum_{\mathbf{z} \in \mathcal{G}(\mathbf{x})} \rho(\|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{z})\|, \sigma_S)], \qquad (2)$$

where ρ is the Lorentzian error norm [13]. The more general version with explicit outlier processes is presented here since the formulation allows the addition of spatial coherence constraints on the analog outlier processes although such constraints are not used for the current experiments.

where $\nabla I = [I_x, I_y]$, $\mathcal{G}(\mathbf{x})$ are the four nearest neighbors of $\mathbf{x}$ on the grid, λ_D and λ_S control the relative importance of the data and spatial terms respectively, and where ρ is a robust error norm. For all the experiments presented here ρ is taken to be the Lorentzian

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x, \sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{3}$$

where ψ is the partial derivative of ρ with respect to x. This ψ-function characterizes the "influence" that a particular measurement has on the solution [19].

Error norms like the Lorentzian have the property that beyond a particular point (where the second derivative of the norm is zero) the influence of a measurement on the solution begins to decrease. Measurements with residual errors which fall beyond this point we refer to as outliers. In the case of the Lorentzian, if the absolute value of the residual error is greater than $\sqrt{2}\sigma$ the measurement is considered an outlier [12]. To derive the coarse estimate we choose σ to be sufficiently large that no measurements are treated as outliers. In general, using a robust error norm may cause E_M to be non-convex. To obtain the coarse estimate, the values of the σ_* are chosen so that the objective function is convex and the function is minimized using Newton's method [11]. A coarse to fine strategy, with warping between layers, is used to estimate large motions within the differential framework.

Consider the Yosemite image sequence whose first image is shown in Figure 2*a*. In this sequence the camera translates and rotates while "flying through" the synthetic Yosemite valley resulting in a diverging flow field. The the sequence is synthetic, the actual flow field is know and is shown in Figure 2*c*. For this sequence 20 iterations of the minimization method were used and the parameters were taken to be $\lambda_D = 10.0$, $\lambda_S = 1.0$, $\sigma_D = 10.0/\sqrt{2}$, $\sigma_S = 1.0/\sqrt{2}$ and 20 iterations of Newton's method were used; these values were used for all other experiments in this paper. For this sequence a three-level pyramid was used in the coarse-to-fine processing.

The horizontal and vertical components of the coarse flow are shown in Figure 3. This coarse flow estimate is very noisy and since the sequence is synthetic, we can compute the error in the flow using using the angular error measure of Barron *et al.* [5]. They represent image velocities as 3-D unit direction vectors $\mathbf{v} \equiv \frac{1}{\sqrt{u^2+v^2+1}}(u, v, 1)^T$. The error between the true velocity $\mathbf{v}_t$ and the estimated velocity $\mathbf{v}_e$ is given by $\arccos(\mathbf{v}_t \cdot \mathbf{v}_e)$. For an additional point of comparison we also compute the mean of the absolute vector difference in pixels between the estimated and the true

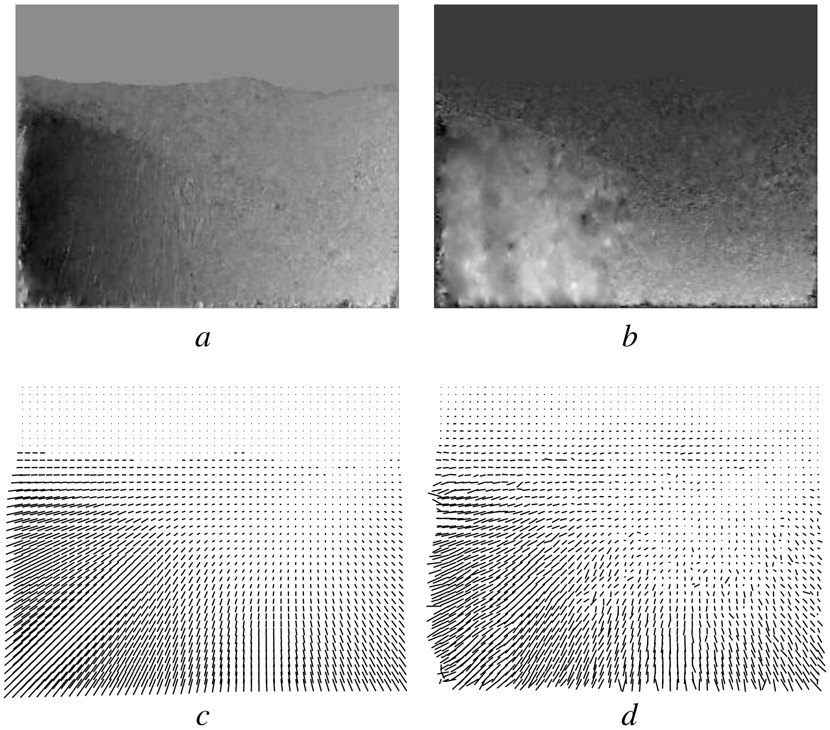


Figure 3: Yosemite Sequence, coarse optical flow. *(a)* Horizontal component of flow (leftward motion = black; rightward motion = white); *(b)* Vertical component of flow (upward motion = black; downward motion = white). *(c)* Actual vector field. *(d)* Computed coarse vector field.

flow vectors [37]. The performance of the algorithm can be quantified as follows[4]:

	Vector Difference	Angular Error	Standard Deviation	Percent of flow vectors with error less than:				
				$< 1°$	$< 2°$	$< 3°$	$< 5°$	$< 10°$
Coarse	0.479	$8.0°$	$7.0°$	3.6%	11.7%	21.4%	39.8%	72.6%

Of course, better initial flow estimates could be obtained with a more sophisticated coarse estimation process. These coarse results are presented as an initial baseline obtainable with a simple dense optical flow algorithm. The next section will illustrate how the medium-level processing significantly improves on these coarse estimates.

4 Medium-Level Processing

The low-level processes described in the previous section are characterized by local processing and weak models of the scene based on regularization. Medium-level processes can be seen as trying

[4]Flow errors were not computed in the sky area since the version of the Yosemite sequence used here does not contain clouds.

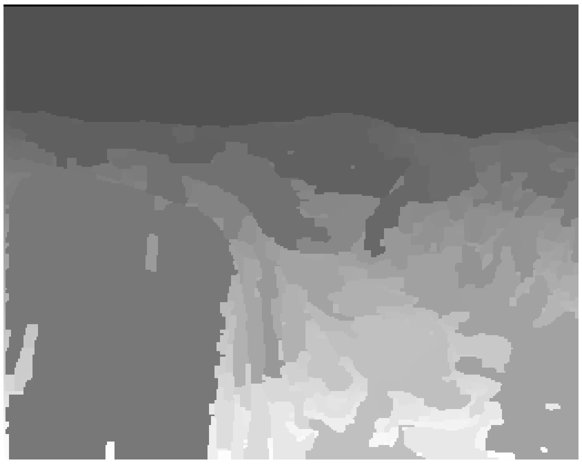


Figure 4: Yosemite Sequence. Connected Components.

to find order and structure in the low-level data and, in doing so, impose more powerful models for interpreting the data. For example, if we have a hypothesis that a region in the image corresponds to a planar surface in the scene, we can use that information to constrain the interpretation of the motion of that region.

We make a very simple hypothesis (which may be wrong) that regions of piecewise-smooth brightness in the image correspond to planar surfaces in the scene. The goal is to use information about image brightness to organize our interpretation of the motion in the scene. From the spatial outliers detected in the piecewise-smooth reconstruction of the image brightness (Figure 2*d*) we detect a set of connected regions R using a standard connected-components labeling algorithm. The connected components for the example image are shown in Figure 4; there are approximately 1000 regions, some of which are only a few pixels in area. These regions become our planar-surface hypotheses. Issues relating to under- and over-segmentation are addressed in Section 7.

4.1 Fitting Parametric Models to Flow Estimates

The image motion of a rigid planar region of the scene can be described by the following eight-parameter model [1, 48]:

$$u(x, y) = a_0 + a_1 x + a_2 y + a_6 x^2 + a_7 xy, \quad (4)$$

$$v(x, y) = a_3 + a_4 x + a_5 y + a_6 xy + a_7 y^2, \quad (5)$$

where the a_i are parameters to be estimated and where $u(x, y)$ and $v(x, y)$ are the horizontal and vertical components of the flow at the image point $\mathbf{x} = (x, y)$. The image points (x, y) are defined relative to some point (x_c, y_c) which can be taken to be a single point (for example, the center of the image) or it can depend on the image region (for example, it can be the centroid of each region). Using the notation from [6] let:

$$\mathbf{X}(\mathbf{x}) = \begin{bmatrix} 1 & x & y & x^2 & xy & 0 & 0 & 0 \\ 0 & 0 & 0 & xy & y^2 & 1 & x & y \end{bmatrix}, \tag{6}$$

$$\mathbf{a} = \begin{bmatrix} a_0 & a_1 & a_2 & a_6 & a_7 & a_3 & a_4 & a_5 \end{bmatrix}^T. \tag{7}$$

To robustly estimate the motion $\mathbf{a}_r$ of a region $r \in R$ we minimize

$$\min_{\mathbf{a}_r} \sum_{\mathbf{x} \in r} \rho(\|\mathbf{X}(\mathbf{x})\mathbf{a}_r - \mathbf{u}_m(\mathbf{x})\|, \sigma), \tag{8}$$

where $\mathbf{u}_m(\mathbf{x}) = [u_m(x, y), v_m(x, y)]^T$ is the coarse flow estimate and σ is a scale parameter. Since the coarse optical flow estimates are expected to have gross errors it is important that the estimation of the motion parameters be performed robustly. For this reason we take ρ to be an error norm with a redescending influence function [19] which has the property of reducing the influence of outlying measurements on the solution. For the experiments in this paper the error norm was taken to be[5]

$$\rho(x, \sigma) = \frac{x^2}{\sigma + x^2}. \tag{9}$$

For this norm a residual error is considered to be an outlier when its absolute value is greater than $\sigma/\sqrt{3}$.

Eqn. (8) is simply minimized using a continuation method in which σ starts at a high value and is gradually lowered. For each value of σ the objective function is minimized using one step of Newton's method. The effect of this is to track the solution while gradually reducing the influence of outlying measurements.

4.1.1 Variable-order Fitting

In many situations the full eight-parameter flow model is not necessary to represent the motion of a region. To avoid over-fitting the motion of a region we use a variable-order fitting approach [8, 29]

[5] We could have chosen any of a number of error norms with redescending influence functions; for example the Lorentzian norm of the previous section. What is more important than the particular function is the qualitative shape of the influence function.

which first assumes a purely translational model by fitting the parameters $\mathbf{a}^{(2)} = [a_0 \ a_3]^T$. The fit is then refined by fitting the six affine parameters $\mathbf{a}^{(6)} = [a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5]^T$. The motion estimates are used to register the regions in the two images by warping the second image towards the first. The resulting temporal error remaining after registration of region r is computed using the two estimates:

$$\epsilon^{(i)} = \sum_{\mathbf{x} \in r} \rho(I(\mathbf{x} - \mathbf{X}(\mathbf{x})\mathbf{a}_r^{(i)}, t+1) - I(\mathbf{x}, t), \sigma). \tag{10}$$

If $\epsilon^{(6)} < \epsilon^{(2)}$ then an affine flow model is adopted for the region otherwise a simple translational model is used. This process is repeated by computing the eight-parameter planar model $a^{(8)}$ and comparing the results with the affine estimates in exactly the same fashion.

To achieve an accurate fit there must be a sufficient number of constraints in the region. To try and ensure sufficient constraints to accurately estimate the motion we require that regions have an area of at least 25, 100, or 400 pixels to be fit by a 2, 6, or 8 parameter model respectively. These values were determined empirically and are conservative.

Note that we are using information about the image brightness to choose the appropriate model. There are two reasons for this. The first is that our goal is to find the motion parameters which register the two regions (ie. minimize the temporal derivative of the registered regions). The second is that the coarse estimates may be very noisy and may have been smoothed across motion boundaries. We do not want to choose higher-order models to fit noisy or outlying flow vectors when a simple model accounts for the spatio-temporal brightness variation.

4.1.2 An Example

The results of fitting the local parametric flow models to the coarse optical flow data for the Yosemite sequence are shown in Figure 5. For this experiment σ began at $4.0\sqrt{3}$ and was lowered by a factor of 0.85 at each iteration to a minimum of $\sqrt{3}$; these values remained fixed for all experiments in this paper. Forty iterations of the minimization method were used to estimate each of the three parametric fits. The recovered parametric flow is projected onto the image to produce the dense flow estimates in Figure 5. The results are a significant improvement over the coarse flow in Figure 3. We can quantify the improvement as follows:

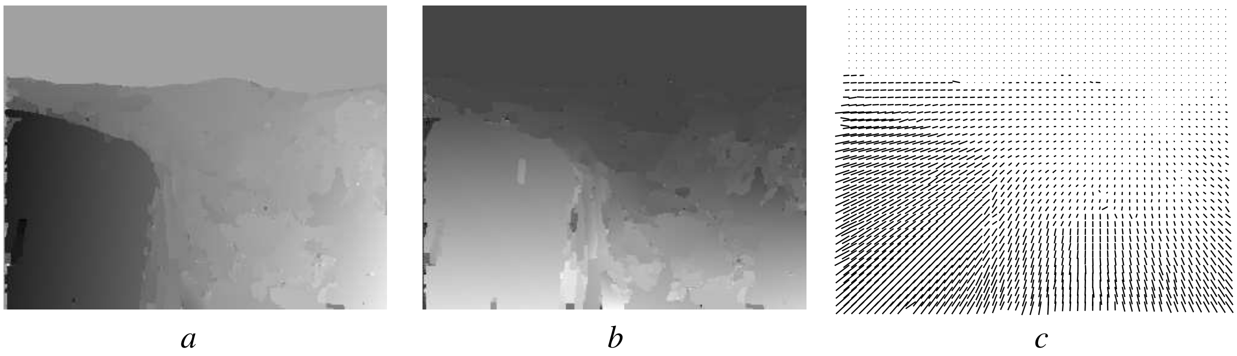


Figure 5: Yosemite Sequence. *(a)* Horizontal component of flow; *(b)* Vertical component of flow. *(c)* Vector field.

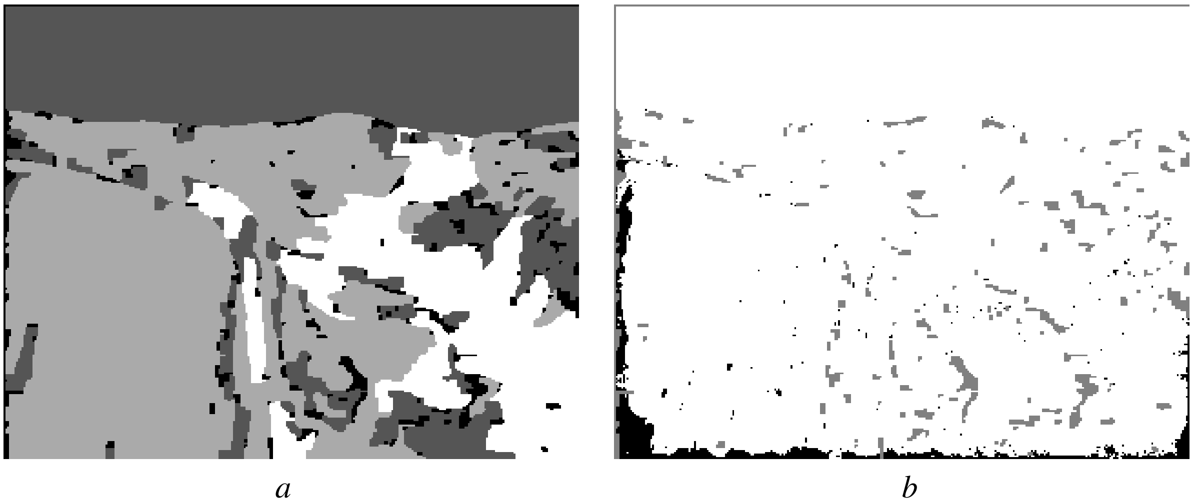


Figure 6: Yosemite Sequence. *a.* Order of the model used (black = none, dark gray = translation, light gray = affine, white = planar); *b.* Flow outliers (gray indicates regions where no parametric model was used, black indicates outliers).

	Vector Difference	Angular Error	Standard Deviation	Percent of flow vectors with error less than:				
				$< 1°$	$< 2°$	$< 3°$	$< 5°$	$< 10°$
Coarse	0.479	$8.0°$	$7.0°$	3.6%	11.7%	21.4%	39.8%	72.6%
Parametric	0.412	$5.2°$	$3.3°$	3.0%	13.6%	26.2%	51.0%	95.1%

The accuracy of the initial coarse flow is quite poor. By fitting local parametric models to the coarse data some of the noisy estimates are removed and the mean accuracy of the flow improves but, given inaccurate estimates to start with, only a small percentage of the flow vectors achieve high accuracy.

The order of the model used is shown in Figure 6*a*. Black indicates that the region was too small to fit a parametric model (ie. smaller than 25 pixels) and the coarse flow estimate was used. Dark gray indicates regions where a translational (2-parameter) model was used, light gray indicates an affine model, and white corresponds to the full 8-parameter model. For the majority of regions affine models produce good results. The regions requiring a higher order model fall in areas where the valley floor curves up to meet the hills on the right. Many of these regions are, in fact, not even planar.

Note that highly textured regions are more likely to be modeled by a high-order flow model than are untextured regions. If the image motion is actually planar, and the region is highly textured, then there will be high brightness errors if a lower-order model is used. If the same region is not textured, then a lower-order model can be used with little penalty. This is a variation of the "aperture problem" for regions.

Outlying coarse-flow vectors which are inconsistent with the parametric flow model are displayed in black in Figure 6*b*. The majority of flow vectors that were treated as outliers (ie. their influence was reduced) occur around the boundary of the image where the initial coarse estimates were poor.

4.2 Local Parametric Models of Image Motion

Fitting parametric models to the flow vectors in regions significantly improves the subjective quality of the flow field. Given the inaccuracy of the coarse flow estimates we would like to refine the motion estimates in each region by going back to the optical flow constraint equations at each pixel. The approach is a straightforward generalization of the approach described by Bergen *et al.*[6] for fitting a single global parametric motion to the entire image.

For each region $r \in R$ the brightness constancy assumption is

$$I(\mathbf{x}, t) = I(\mathbf{x} - \mathbf{u}(\mathbf{x}; \mathbf{a}_r^{(i)}), t + 1) \;\; \forall \mathbf{x} \in r$$

where $\mathbf{u}(\mathbf{x}; \mathbf{a}_r^{(i)}) = \mathbf{X}(\mathbf{x})\mathbf{a}_r^{(i)}$, $\mathbf{a}_r^{(i)}$ are the parameters for region r, and $i \in \{2, 6, 8\}$ indicates the parametric model to be used as determined by the initial fitting procedure. Given the current fit $\mathbf{a}_r^{(i)}$ for a region we warp the image at time $t + 1$ towards the image at time t. The original region at time t and this warped region are used to estimate the spatial and temporal derivaties I_x, I_y, and I_t. Let $\nabla I = [I_x, I_y]$, then to refine the current fit we minimize

$$\min_{\delta \mathbf{a}_r^{(i)}} \sum_{\mathbf{x} \in r} \rho((\nabla I(\mathbf{x}) X(\mathbf{x}) \delta \mathbf{a}_r^{(i)} + I_t(\mathbf{x})), \sigma), \tag{11}$$

and then the refined fit is $\mathbf{a}_r^{(i)} + \delta \mathbf{a}_r^{(i)}$.

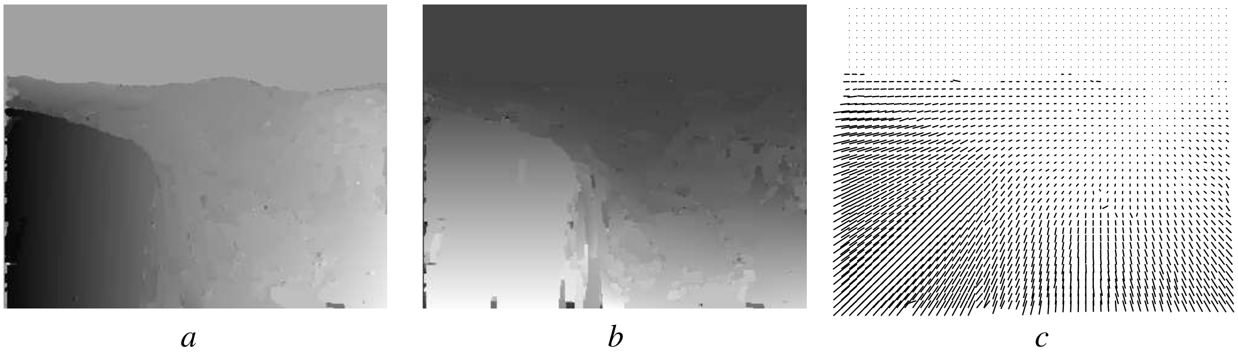


Figure 7: Yosemite Sequence. Refined motion estimates. *(a)* Horizontal component of flow; *(b)* Vertical component of flow. *(c)* Vector field.

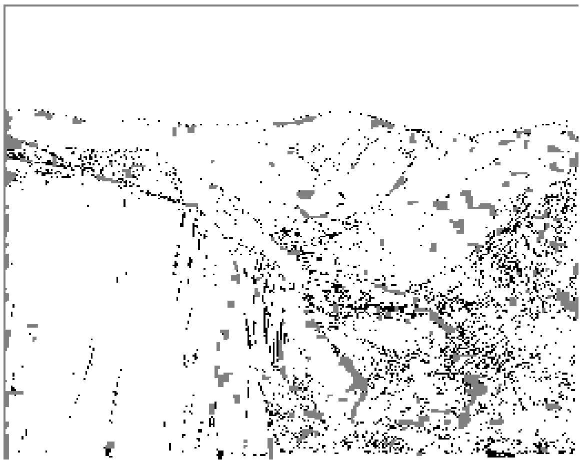


Figure 8: Yosemite Sequence. Black regions correspond to places where the brightness constancy assumption was violated. Gray indicates small regions where no parametric model was used.

To minimize Eqn. (11) we use exactly the same continuation method described above in which σ is gradually lowered and at each stage we apply one step in Newton's method. Since the initial flow estimates are fairly accurate we do not need to use a coarse-to-fine strategy as in [6].

The results of refining the flow for the Yosemite sequence are shown in Figure 7. For this experiment σ began at $20.0\sqrt{3}$ and was lowered by a factor of 0.85 at each iteration to a minimum of $10.0\sqrt{3}$. Once again, 40 iterations of the minimization method were used to refine the estimate, The results are visually similar to the initial fit though some improvement can be seen. Quantitatively, however, refining the motion estimates significantly improves the accuracy of the recovered flow field:

	Vector Difference	Angular Error	Standard Deviation	Percent of flow vectors with error less than:				
				$< 1°$	$< 2°$	$< 3°$	$< 5°$	$< 10°$
Coarse	0.479	$8.0°$	$7.0°$	3.6%	11.7%	21.4%	39.8%	72.6%
Parametric	0.412	$5.2°$	$3.3°$	3.0%	13.6%	26.2%	51.0%	95.1%
Refined	0.202	$2.9°$	$3.2°$	15.5%	49.4%	71.9%	87.1%	96.5%

Figure 8 shows where the brightness constancy assumption was violated. Outliers are shown in black and correspond to measurements where

$$\frac{\sigma}{\sqrt{3}} < |(\nabla I(\mathbf{x})X(\mathbf{x})\delta\mathbf{a}_r^{(i)} + I_t(\mathbf{x}))|.$$

One might ask "Why not start with this region-based regression approach and ignore the coarse flow computation?" This approach will work for large, slow moving, regions. The problem with such an approach becomes apparent when trying to estimate the motion of a small region which is moving quickly. To deal with large motions using a differential technique, it is necessary to use a coarse-to-fine approach. But small regions may have little support at the coarse levels making it impossible to recover their motion. The regularization present in the coarse stage typically provides a good initial estimate for small fast moving regions if they are part of a larger moving structure.

5 Local Deformations

Local models of planarity are likely to be violated often in practice, particularly in natural scenes. For this reason we would like to use local parametric models to provide a *coarse* description of the motion and allow *deformations* from the parametric model to account for errors in the assumption. This notion of modeling optical flow using a planar motion plus a general flow field has recently been used for recover 3D structure [40]. In a rigid scene the image motion of an arbitrary plane can be estimated and used to stabilize two images in the sequence effectively removing the rigid camera rotation. The residual motion in the scene, called *planar motion parallax*, is an epipolar field in which the magnitude of a residual motion vector is related to its depth relative to the planar surface used to stabilize the sequence. This simple relationship to 3D structure has been used for recovering structure from motion [27, 40]. Unlike these planar parallax approaches we do not stabilize the entire scene based on a single planar motion but, rather, stabilize an isolated patch based on its motion. The "deformations" we estimate from this planar motion are the result of planar motion parallax and are related to the patch's 3D variation from planarity. While we have not used this

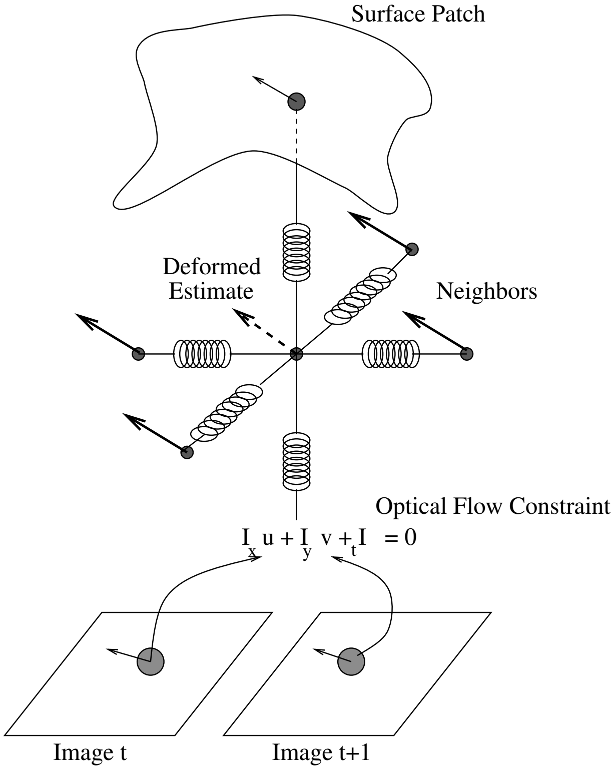


Figure 9: Deformation model. The flow at a point is connected, via non-linear springs, to its neighbors, the optical flow constraint at that point, and the estimated parametric model of the planar patch.

parallax to recover local structure, the application of plane+parallax methods to local image regions is an interesting area for further exploration.

We estimate local deformation, or parallax, using the robust optical flow estimation technique described in [11] with the addition of a new term now coupling the flow estimate to the parametric-prediction of the flow. The flow estimate at each point can be thought of as being connected, via non-linear springs, to its neighbors, the data (optical flow constraint equation), and the estimated motion of the planar-patch. This is illustrated in Figure 9. The estimate is pulled by all these forces and the strength of the force is determined by the robust error norm $\rho(x, \sigma)$. If the estimate gets pulled too far from its neighbors, the data, or the planar-patch estimate, the spring essentially goes "slack". This is equivalent to rejecting that measurement as an outlier.

Given the predicted flow in the planar patches, the image at time $t + 1$ is warped back towards the image at time t to register them. The deformation $\delta\mathbf{u}$ is estimated to account for the discrepancy between the warped and original images. This physical model is implemented as the minimization of the following objective function with respect to $\delta\mathbf{u}$:

$$
\begin{aligned}
E_D(\delta\mathbf{u}, \mathbf{u}, \mathbf{a}) \quad = \quad & \sum_{\mathbf{x}} [\rho((\nabla I(\mathbf{x})\delta\mathbf{u}(\mathbf{x}) + I_t(\mathbf{x})), \sigma_D) \\
& + \frac{1}{|\mathcal{G}(\mathbf{x})|} \sum_{\mathbf{z}\in\mathcal{G}(\mathbf{x})} [\rho(\|(\mathbf{u}(\mathbf{x};\mathbf{a}) + \delta\mathbf{u}(\mathbf{x})) - (\mathbf{u}(\mathbf{z};\mathbf{a}) + \delta\mathbf{u}(\mathbf{z}))\|, \sigma_S)] \\
& + \rho(\delta\mathbf{u}(\mathbf{x}), \sigma_M)] \,, \qquad (12)
\end{aligned}
$$

where $\mathcal{G}(\mathbf{x})$ are neighbors of $\mathbf{x}$, ρ is a robust error norm which reduces the influence of outlying measurements, $\mathbf{u}(\mathbf{x};\mathbf{a})$ is the refined flow from the medium-level processing, and where the spatial and temporal derivatives are computed with respect to the warped image pair.

The first term in E_D is a robust formulation of the standard optical flow constraint equation and enforces fidelity to the data. The second term pulls the deformation in a direction which minimizes the difference in the neighboring flow vectors. The final term forces the flow to be similar to the planar-patch estimate by penalizing for deformations.

Given the accurate initial estimate there is no need for a coarse-to-fine approach and in our experiments we simply minimize the objective function using Newton's method. A continuation method may be exploited using the scale parameters σ_* as was done in the previous section. We have not found this to be necessary since the estimates from the patches start the minimization near the global minimum.

The segmented image patches in the Yosemite sequence are only approximately planar. Allowing local deformations to the motion of the planar patches results in the deformations in Figures 10*a* and 10*b*. Figure 10*c* shows the vector field corresponding to these displacements with the flow vectors scaled by a factor of five to make them visible. The final parametric+deformation flow is shown in Figures 10*d-f*. For this experiment we took ρ to be the Lorentzian for consistency with the coarse motion estimation stage and used 40 iterations of the minimization scheme. The parameters were taken to be $\sigma_D = 3.0/\sqrt{2}$, $\sigma_S = 0.05/\sqrt{2}$, and $\sigma_M = 0.5/\sqrt{2}$ in Eqn. (12); these same values were used for all image sequences in this paper. Visually the flow after deformation varies more smoothly than the refined fit and quantitatively the deformation stage results in a significant improvement in the accuracy of the flow:

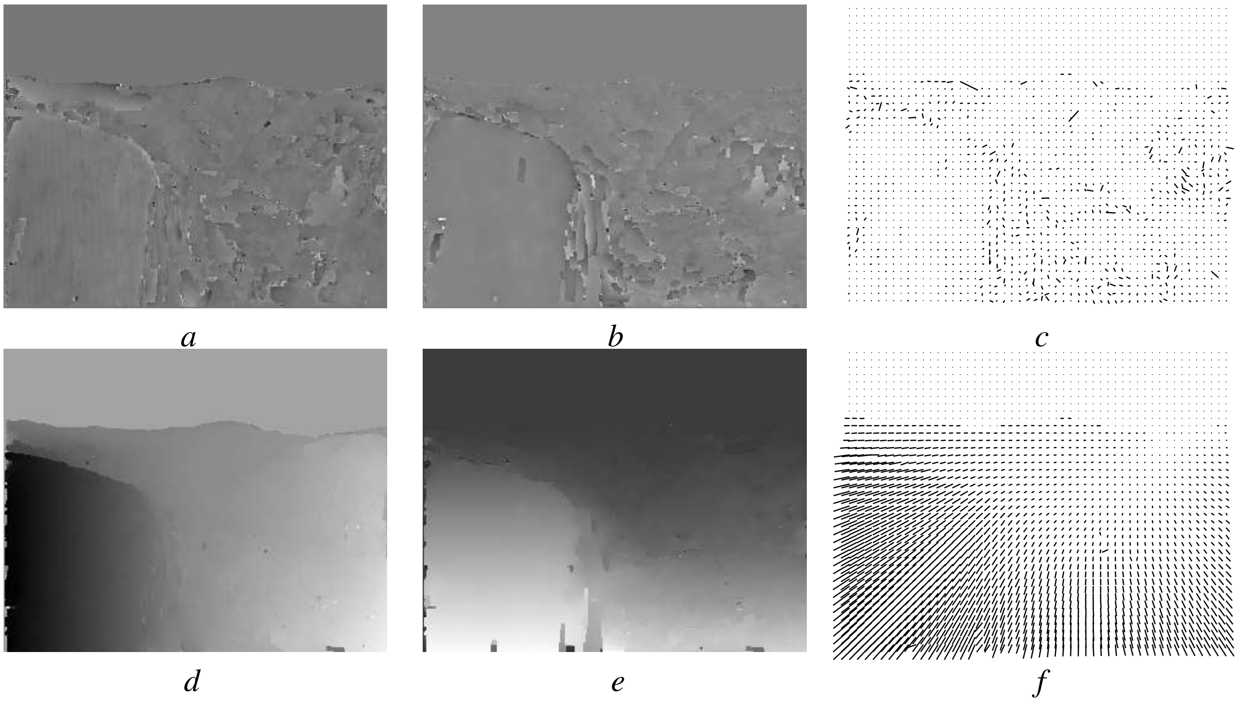


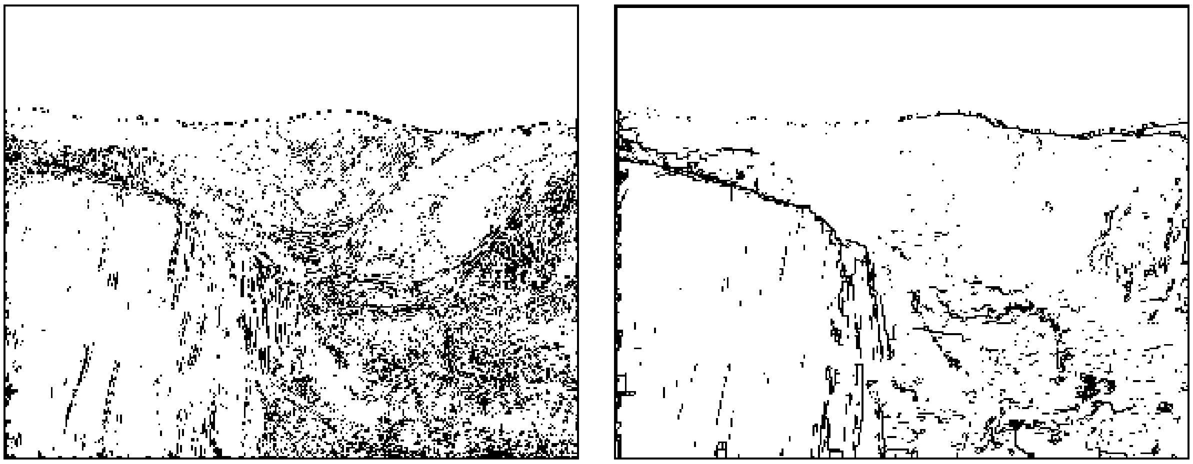
Figure 10: Yosemite Sequence. Parametric fit plus local deformations. *(a)* Horizontal flow deformation; *(b)* Vertical flow deformation. *(c)* Flow deformation (scaled by a factor of 5 to show the displacements). *(d)* Horizontal flow: Parametric plus deformations; *(e)* Vertical flow: Parametric plus deformations. *(f)* Vector field: Parametric plus deformations.

	Vector Difference	Angular Error	Standard Deviation	Percent of flow vectors with error less than:				
				$< 1°$	$< 2°$	$< 3°$	$< 5°$	$< 10°$
Coarse	0.479	$8.0°$	$7.0°$	3.6%	11.7%	21.4%	39.8%	72.6%
Parametric	0.412	$5.2°$	$3.3°$	3.0%	13.6%	26.2%	51.0%	95.1%
Refined	0.202	$2.9°$	$3.2°$	15.5%	49.4%	71.9%	87.1%	96.5%
Deformed	0.176	$2.3°$	$2.3°$	18.7%	57.0%	81.2%	93.8%	98.8%

Compare the recovered flow field in Figure 10*f* with the ground truth in Figure 3*c*. Figure 11 shows where the data term and the spatial term were treated as outliers.

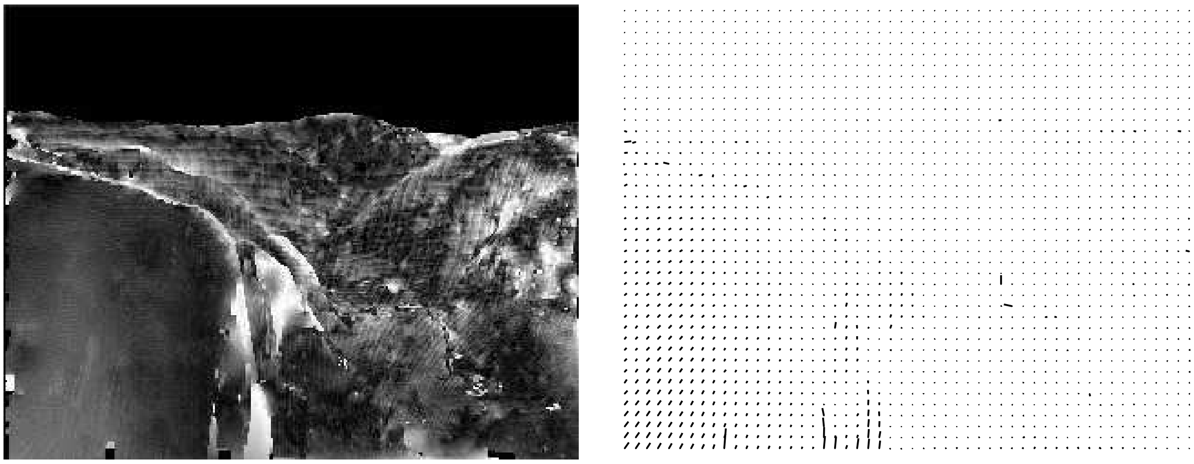
6 Experimental Results

To illustrate the performance of the approach we consider a wide variety of image sequences containing different types of camera motion, independent and articulate objects, and both indoor and outdoor scenes with varying amounts of texture. As mentioned in the text, most of the parameters used in the Yosemite sequence experiment remain unchanged for all the other experiments; where that is not the case it will be noted below. All experiments other than Yosemite used 30 iterations



a *b*

Figure 11: Yosemite Sequence. Planar+Deformation outliers. *(a)* Data outliers; *(b)* Spatial outliers.



a *b*

Figure 12: Yosemite sequence. *(a)* Angular error (scaled to show detail); *(b)* Vector difference field.

of both the coarse and refined parametric fitting processes, 20 iterations of the deformation process were used, and 40 iterations of the segmentation process.

6.1 Yosemite Sequence (wrap-up)

The Yosemite sequence experiments presented throughout the text chronicle the quantitative improvement in the flow at each step in the processing. The recovered flow field in Figure 10*f* is visually similar to the ground truth in Figure 3*c*; for further comparison Figure 12 shows both the angular error at each pixel (*a*) and the vector difference field (*b*). The largest angular errors occur in regions which were in fact non-planar (most of the scene contains rolling hills). The vector differ-

Technique	Average Error	Standard Deviation	Density
Anandan [2]	$15.84°$	$13.46°$	100%
Singh [41]	$13.16°$	$12.07°$	100%
Nagel [36]	$11.71°$	$10.59°$	100%
Horn and Schunck (modified) [21]	$11.26°$	$16.41°$	100%
Uras *et al.* [45]	$10.44°$	$15.00°$	100%
Fleet and Jepson [17]	$4.29°$	$11.24°$	34.1%
Lucas and Kanade [31]	$4.10°$	$9.58°$	35.1%
Weber and Malik [50]	$3.42°$	$5.35°$	45.2%
Black and Anandan [11]$*$	$4.47°$	$3.90°$	100%
Black [10]$*$	$3.52°$	$3.25°$	100%
Parametric+Deformation$*$	$2.29°$	$2.25°$	100%

Table 1: Comparison of various optical flow algorithms.

ence image gives another look at the performance. Here we see a few outliers and then the largest errors occurring on the foreground rock face which is the fastest moving portion of the image. If the reader refers back to Figure 6*a* they will see that the face was modeled as an affine motion. The error in the estimate may be a result of choosing a model which is too simple.

The results of the planar+deformation approach are compared with other published results for the Yosemite sequence in Table 1 (cf. [5]). The accuracy of the approach is in the range of the most accurate approaches but with 100% density (not counting the sky). Methods followed by a "$*$" have errors computed without the sky region while the other methods include the sky. In [5] the error for Lucas and Kanade [31] and Fleet and Jepson [17] improves to $3.37°$ and $2.97°$ respectively when the sky is omitted though the density remains low. The accuracy of the other approaches might also be expected to improve in accuracy by approximately 25% if the sky is ignored (see [5]) which still remains well below the accuracy of the planar+deformation model.

6.2 Nap-Of-the-Earth Sequence

The next experiment considers a natural image sequence, similar to the Yosemite sequence, taken by a helicopter flying through a canyon (Figure 13*a*). The helicopter is translating forward and to the left while rotating to the right and the resulting flow field is strongly diverging. The sequence illustrates that there is nothing particularly special about the Yosemite sequence and that the approach will work for a similar natural sequence. The spatial discontinuities are shown in Figure 13*b* and the recovered horizontal an vertical motion is shown in Figures 13*c* and 13*d* respectively.

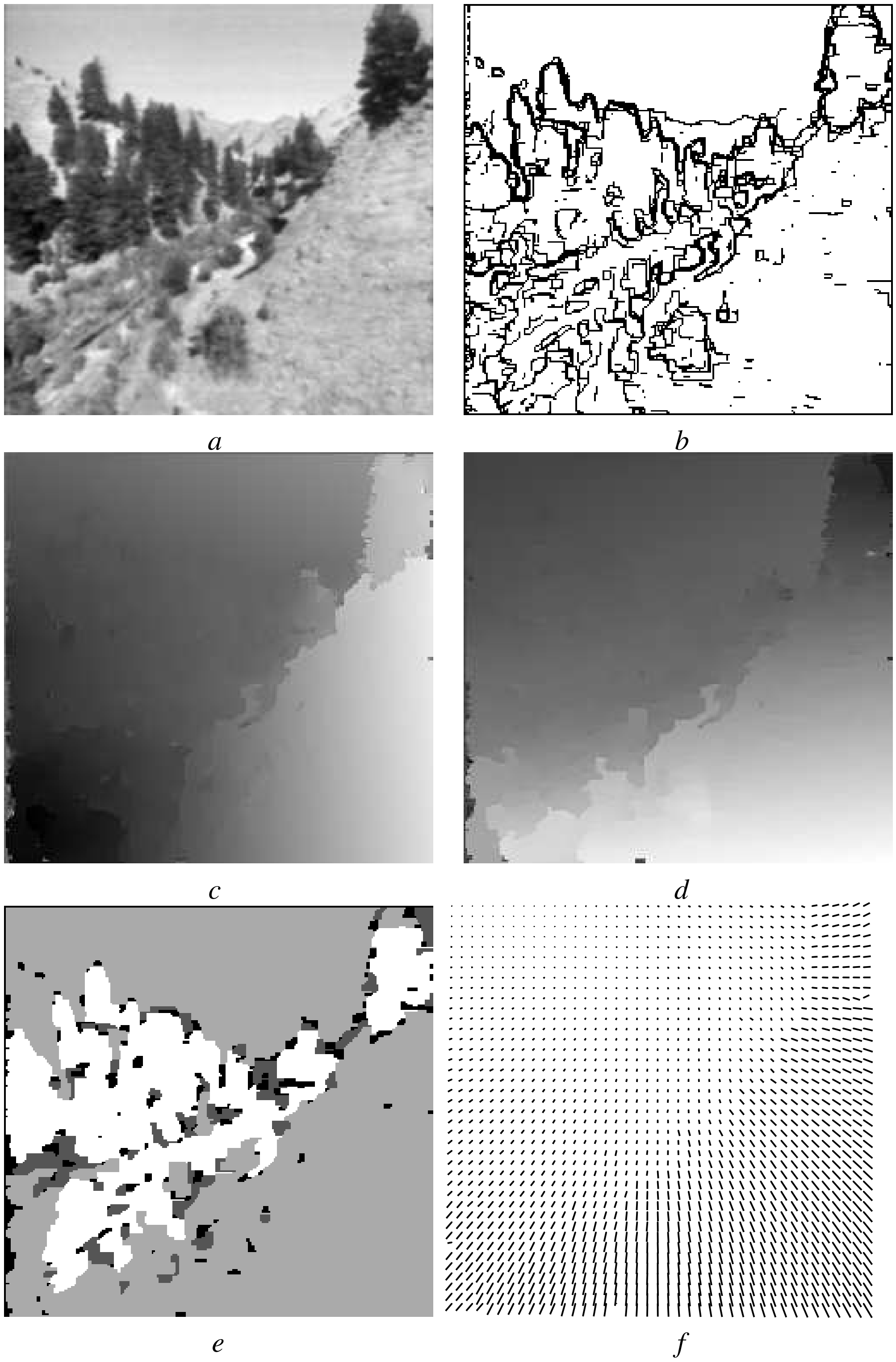


Figure 13: Nap-Of-the-Earth Helicopter Sequence. *(a)* First image; *(b)* Spatial discontinuities; *(c)* Planar+Deformations: horizontal flow; *(d)* Planar+Deformations: vertical flow; *(e)* Order of the model used; *(f)* Flow field.

The parameters were: for the segmentation process, $40 \geq \sigma_D\sqrt{2} \geq 20$ and $40 \geq \sigma_S\sqrt{2} \geq 4$ and the refined parametric fit, $15 \geq \sigma/\sqrt{3} \geq 5$. The vector field in Figure 13*f* gives a qualitative sense of the motion while Figure 13*e* shows the type of parametric model used for each region (translational (dark gray), affine (light gray), or planar (white)). Notice that there are problems with undersegmentation at the boundary between the land and sky. This results in a non-zero flow in the sky where there is no texture.

6.3 SRI Tree Sequence

A second natural outdoor sequence is provided to illustrate the effect of the algorithm at motion discontinuities. The first image in the SRI tree sequence is shown in Figure 14*a*. In this sequence the camera translates parallel to the image plane resulting in a horizontal optical flow field where the magnitude of the flow at a pixel is inversely proportional to the depth of the point in the scene. Despite the fact that the images are highly textured, the segmentation (Figure 14*b*) produces regions of adequate size to estimate the motion. The order of the models used within regions is shown in Figure 14*c*. Recall that black regions indicate that the region was too small for parameterized motion estimation and the coarse flow is used. There is no significant vertical displacement so only the horizontal component of the motion is shown in Figure 14. Figure 14*d* and *e* show the coarse horizontal displacement and flow respectively, while Figure 14*f* and *g* show the final planar+deformation displacement and flow. The data and spatial outliers detected during the deformation stage are shown in Figures 14*h* and 14*i* respectively. The horizontal bands in the data outlier image are due to noise in the image sequence. The spatial outliers correspond well to the branches of the trees. The parameters were: for the segmentation process, $25 \geq \sigma_D\sqrt{2} \geq 10$ and $20 \geq \sigma_S\sqrt{2} \geq 2$ and the refined parametric fit, $15 \geq \sigma/\sqrt{3} \geq 5$.

6.4 Walking Sequence

The next experiment shows the application of the approach to a sequence containing both camera motion and an independently moving object in a cluttered indoor environment. In the sequence the camera pans to roughly track the walking figure. This results in a roughly uniform, and large, motion for the background while the motion of the person is small. The camera motion is not pure rotation resulting in some flow variation with depth. The parameters were: for the segmentation

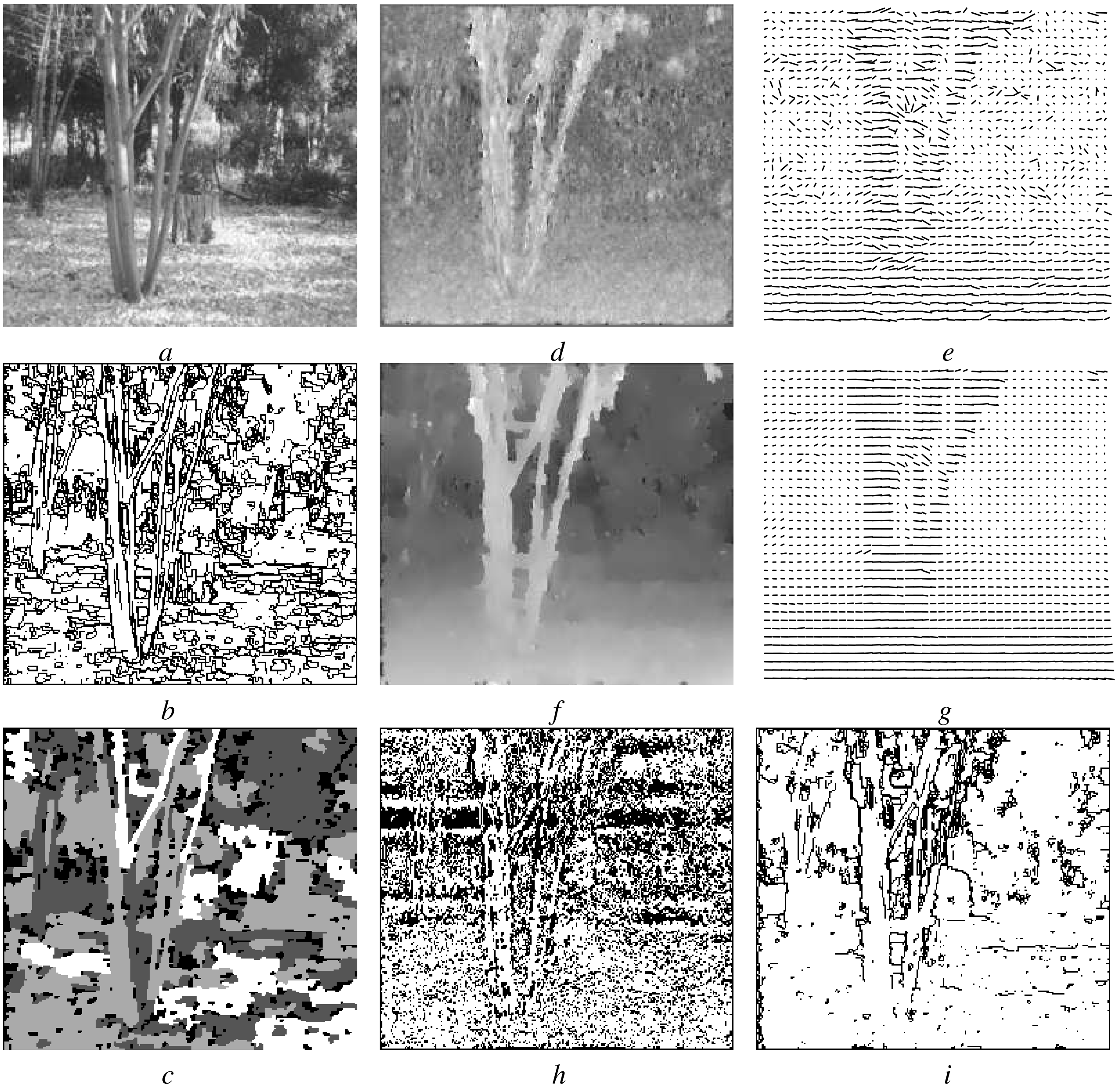
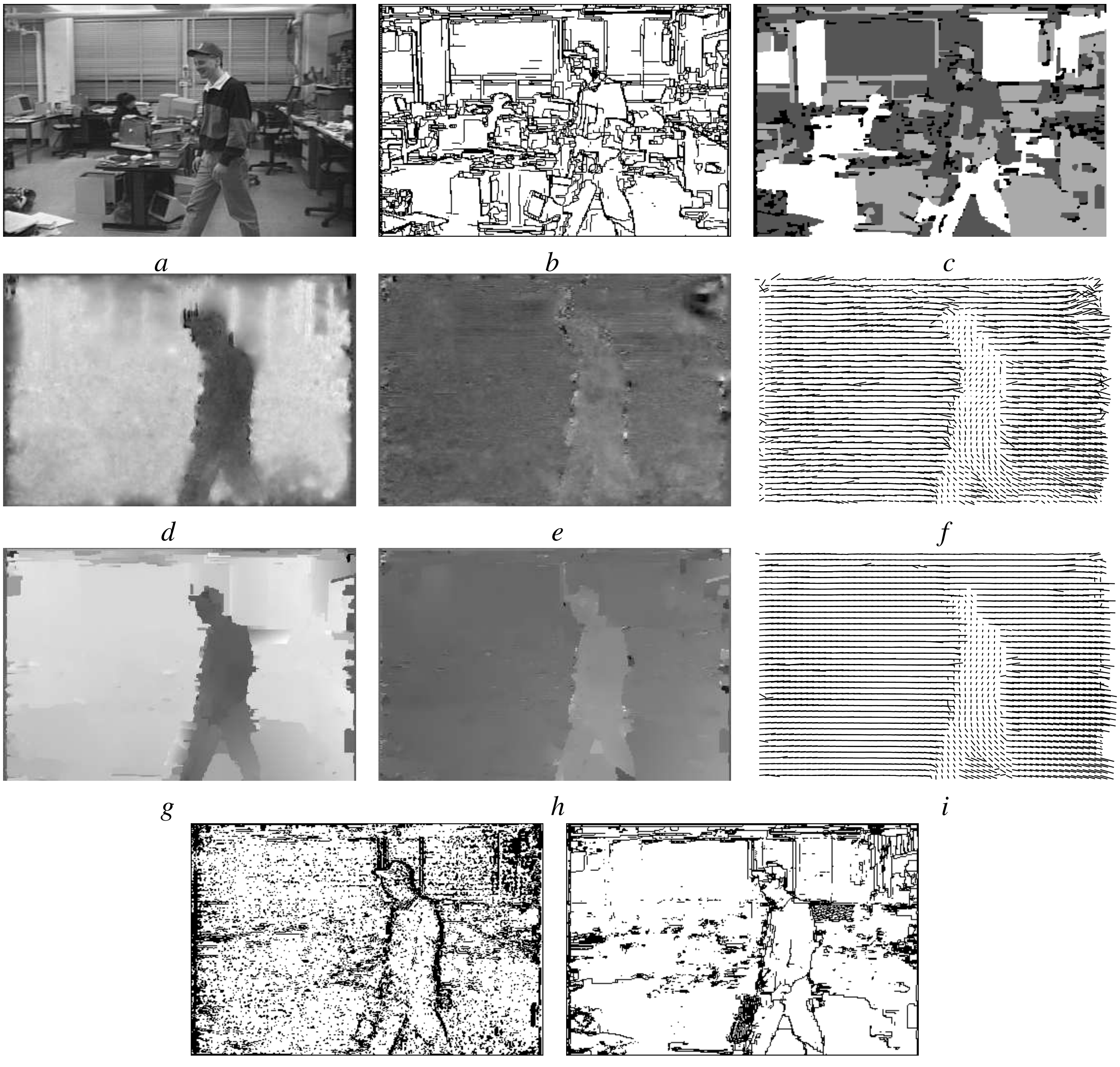


Figure 14: SRI Tree Sequence. *(a)* First image; *(b)* Spatial discontinuities; *(c)* Order of the model; *(d)* Coarse horizontal displacement; *(e)* Coarse flow field; *(f)* Planar+Deformations: horizontal displacement; *(g)* Planar+Deformations: flow field; *(h)* Data outliers; *(i)* Spatial outliers.



a *b* *c* *d* *e* *f* *g* *h* *i* *j* *k*

Figure 15: A cluttered scene in which the camera is panning and a person is walking. *(a)* First image; *(b)* Spatial discontinuities; *(c)* Order of the model; *(d/e)* Coarse horizontal/vertical displacement; *(f)* Coarse vertical flow field; *(g/h)* Planar+Deformations: horizontal/vertical displacement; *(i)* Planar+Deformations: flow field; *(j)* Data outliers; *(k)* Spatial outliers.

process, $15 \geq \sigma_D\sqrt{2} \geq 5$ and $100 \geq \sigma_S\sqrt{2} \geq 4$ and for the refined parametric fit, $15 \geq \sigma/\sqrt{3} \geq 5$. Figure 15*b* shows the brightness discontinuities found in the first image of the sequence (Figure 15*a*). The coarse flow estimates are shown in Figures 15*d*, 15*e*, and 15*f*. The final parametric+deformation results are shown in Figures 15*g*, 15*h*, and 15*i*. The bottom two images show the data and spatial outliers after deformation. Notice that the boundary of the moving person is well localized in Figure 15*k*. This sequence illustrates that the brightness segmentation helps in the accurate localization of motion boundaries.

7 Open Questions and Future Directions

Accurate and dense estimates of optical flow have a wide variety of applications to in diverse problems such as image coding, structure from motion, and the recognition of human activities. The goal of this work has been to explore two aspects of the optical flow problem. First we are interested in how to choose the appropriate area of integration within which to employ parameterized models of optical flow. This is an important problem since large areas of integration result in accurate motion estimates. But large, arbitrarily shaped, regions may have multiple motions within them or may not satisfy the assumptions of the parameterized model (eg. planarity). We have referred to this problem of choosing the appropriate region for integration as the *generalized aperture problem* [23].

The second aspect of the work involves the use of static brightness information to improve the estimation of image motion. Previous attempts to integrate motion and brightness have often focused on using brightness discontinuities to improve the localization of motion discontinuities. Here we have exploited brightness segmentation to help us address the first problem of choosing the region of integration.

Our simple assumption of piecewise constant brightness is clearly not satisfied in general. For example consider image sequences consisting of random dots. Humans have little trouble estimating motion in such sparse sequences but brightness segmentation will be of no help in organizing the moving dots. While the interaction between motion and brightness is clearly more complex than that presented here, our results suggest that the integration of these cues can significantly improve optical flow estimation. The integration of multiple cues however is a hard problem and we have presented only one simple approach.

In addition to the general issues of cue integration, this work leaves a number of unanswered questions and suggests interesting future research directions. For example, given a collection of planar patches in the scene and their motion, we would like to estimate the 3D motion of the camera. A rigid-body assumption could be incorporated into the flow estimation to constrain the motion of patches to be consistent with a rigid scene. An immediate application of this would be the detection of independently moving regions in the scene whose motion is inconsistent with a rigid 3D interpretation [33].

Due to over-segmentation based on brightness, the localization of objects, as opposed to surfaces patches, may require grouping patches together based on common motion. The approaches of Ayer *et al.* [4] and Wang and Adelson [46] present possible methods for achieving this grouping and do not appear to exhibit the kinds over-segmentation we see with our method. While it should be relatively straightforward to extend these methods to group our segmented patches the local deformations to some extent already provide this grouping, or merging, at a low level. It may be desirable to exploit the deformed motion at the region boundaries in deciding which regions to group into larger regions.

Additionally, having the motion of segmented image regions means that the occlusion relationships between the regions can be analyzed over time. The addition of temporal integration might also improve the accuracy, efficiency, and robustness of the method. Moreover, it may be possible to incorporate a layered representation which can represent occluded portions of regions viewed over many frames as in the work of Wang and Adelson [46]. This segmented and layered representation of a video stream might be useful for video coding; for example, MPEG-4.

Under-segmentation is also an issue. For example, in our experiments with moving people, their legs are often segmented into a single region based on brightness. We would like to be able to detect that a single motion does not give a good fit to this region and break it into parts in the appropriate places. One possibility is to use the local deformation as a measure of strain and introduce breaks when the strain is too great [24]. An alternative way to cope with undersegmentation is to allow multiple motions within a region and use either a robust estimation approach [12] or a mixture model approach [23] to recover the multiple motions.

The segmentation approach presented here is merely used to illustrate the idea of exploiting static segmentation in motion estimation. It is interesting to note that the idealized brightness model used for segmentation is one of piecewise constant brightness. Not only is this an unrealistic model of brightness in natural scenes but, if the segmented regions were actually of constant brightness, then the optical flow constraint equation would provide no motion information within the image regions. The segmentation method actually produces regions with small variations around the mean brightness within the region. This variation is controlled by the σ parameters and is necessary for reliable motion estimation. Future work should explore the use of texture segmentation techniques (e.g. [18]) which would yield regions with adequate texture for motion estimation.

The segmentation and motion approaches presented here rely on a number of parameters, particularly the scale parameters σ. The segmentation method used here is somewhat sensitive to the choice of parameters but, since the segmentation information is used primarily for illustration, we chose these parameters by hand. The motion estimation method, on the other hand, is not very sensitive to the choice of parameters as has been demonstrated elsewhere [12]. In nearly all the experiments, the parameters for the motion estimation are identical and standard statistical techniques can be used to estimated these parameters automatically [3, 28, 30, 39]. Additionally, statistical measures of the accuracy of the motion estimation within a region might be used to provide a *confidence measure*.

Finally, this paper has presented the fitting and deformation process as a one-shot algorithm. In fact, it may be useful for this process to iterate in the context of an incremental estimation scheme where estimates are refined over an image sequence (cf. [4, 10, 35]).

8 Conclusion

This paper has presented a new model for estimating optical flow based on the motion of planar regions plus local deformations. The approach exploits brightness information to organize and constrain the interpretation of the motion by using segmented regions of piecewise smooth brightness to hypothesize planar regions in the scene. Parametric flow models are estimated in these regions in a two step process which first computes a coarse fit and then refines it using a generalization of the standard area-based regression approaches. Since the planar-patch assumption is likely to be violated, we allow local deformations from the parametric flow using a physically-based model in which a regularized optical flow estimate is partially constrained by the parametric motion estimate.

The approach produces good results on a wide variety of image sequences for two primary reasons. The first is that the segmented regions provide large areas for integrating multiple constraints and, as opposed to methods which choose a particular fixed region size/shape (eg. [31]), the segmented regions are less likely in general to contain multiple motions; or, said another way, are more likely to correspond to actual planar surfaces in the scene. The second reason is that the brightness segmentation provides good localization of motion boundaries since motion discontinuities often coincide with brightness discontinuities. The approach illustrates the importance of both of these properties (large areas of integration and use of brightness in localizing motion boundaries).

Acknowledgements

We would like to thank David Fleet and David Heeger for discussions that helped to clarify this work.

References

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(4):384–401, July 1985.

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

[3] S. Ayer and H. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. In *Fifth International Conference on Computer Vision*, pages 777–784, Boston, MA, 1995.

[4] S. Ayer, P. Schroeter, and J. Bigün. Segmentation of moving objects by robust motion parameter estimation over multiple frames. In J. Eklundh, editor, *European Conf. on Computer Vision, ECCV-94*, volume 801 of *LNCS-Series*, pages 317–327, Stockholm, Sweden, 1994. Springer-Verlag.

[5] J.L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1), 1994.

[6] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS-Series*, pages 237–252. Springer-Verlag, May 1992.

[7] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):886–896, September 1992.

[8] P. J. Besl and J. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2), March 1988.

[9] M. J. Black. Combining intensity and motion for incremental segmentation and tracking over long image sequences. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS-Series*, pages 485–493. Springer-Verlag, May 1992.

[10] M. J. Black. Recursive non-linear estimation of discontinuous flow fields. In J. Eklundh, editor, *European Conf. on Computer Vision, ECCV-94*, volume 800 of *LNCS-Series*, pages 138–145, Stockholm, Sweden, 1994. Springer-Verlag.

[11] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. on Computer Vision, ICCV-93*, pages 231–236, Berlin, Germany, May 1993.

[12] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[13] M. J. Black and A. Rangarajan. The outlier process: Unifying line processes and robust statistics. In *Proc. Computer Vision and Pattern Recognition, CVPR-94*, pages 15–22, Seattle, WA, June 1994.

[14] T. Darrell and A. Pentland. Robust estimation of a multi-layer motion representation. In *Proc. IEEE Workshop on Visual Motion*, pages 173–178, Princeton, NJ, October 1991.

[15] M. Dubuisson and A. K. Jain. Object contour extraction using color and motion. In *Proc. Computer Vision and Pattern Recognition, CVPR-93*, pages 471–476, New York, June 1993.

[16] C. L. Fennema and W. B. Thompson. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9:301–315, 1979.

[17] D.J. Fleet and A.D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[18] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

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

[20] F. Heitz and P. Bouthemy. Multimodal motion estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1217–1232, December 1993.

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

[22] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS-Series*, pages 282–287. Springer-Verlag, May 1992.

[23] A. Jepson and M. J. Black. Mixture models for optical flow computation. In Ingmer Cox, Pierre Hansen, and Bela Julesz, editors, Partitioning Data Sets: With Applications to Psychology, Vision and Target Tracking, pages 271–286, DIMACS Workshop, April 1993. AMS Pub., Providence, RI.

[24] I. A. Kakadiaris, D. Metaxas, and R. Bajcsy. Active part-decomposition, shape and motion estimations of articulated objects: A physics-based approach. In *Computer Vision and Pattern Recognition, CVPR-94*, pages 980–984, Seattle, WA, 1994.

[25] J. K. Kearney, W. B. Thompson, and D. L. Boley. Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:229–244, 1987.

[26] R. Koch. Automatic reconstruction of buildings from stereoscopic image sequences. *EURO-GRPHICS'93*, 12(3):339–350, 1993.

[27] R. Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views: A parallax based approach. In *ARPA Image Understanding Workshop*, 1994.

[28] R. Kumar and A. R. Hanson. Analysis of different robust methods for pose refinement. In *Proc. Int. Workshop on Robust Computer Vision*, pages 167–182, Seattle, WA, October 1990.

[29] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation as the search for the best description of the image in terms of primitives. Technical Report MS–CIS–90–30, GRASP LAB 215, May 1990.

[30] G. Li. Robust regression. In F. Mosteller and J. W. Tukey, editors, *Exploring Data, Tables, Trends and Shapes*. John Wiley & Sons, New York, 1985.

[31] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. 7th IJCAI*, pages 674–679, Vancouver, B. C., Canada, 1981.

[32] W. Luo and H. Maître. Using surface model to correct and fit disparity data in stereo vision. In *Proc. IEEE Int. Conf. on Pattern Recognition*, pages 60–64, June 1990.

[33] W. J. MacLean, A. D. Jepson, and R. C. Frecker. Recovery of egomotion and segmentation of independent object motion using the em algorithm. In *Proc. British Machine Vision Conf.*, York, U.K., 1994.

[34] G.J. McLachlan and K.E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker Inc., N.Y., 1988.

[35] F. G. Meyer and P. Bouthemy. Region based tracking using affine motion models in long image sequences. *CVGIP: Image Understanding*, 60(2):119–140, September 1994.

[36] H. H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33(3):299–324, November 1987.

[37] M. Otte and H. H. Nagel. Optical flow estimation: Advances and comparisons. In J. Eklundh, editor, *European Conf. on Computer Vision, ECCV-94*, volume 800 of *LNCS-Series*, pages 51–60, Stockholm, Sweden, 1994. Springer-Verlag.

[38] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):715–729, July 1991.

[39] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

[40] H. S. Sawhney. 3D geometry from planar parallax. In *Computer Vision and Pattern Recognition, CVPR-94*, pages 929–934, Seattle, WA, 1994.

[41] A. Singh. *Optic Flow Computation: A Unified Perspective*. IEEE Computer Society Press, 1992.

[42] S. Sull and N. Ahuja. Segmentation, matching and estimation of structure and motion of textured piecewise planar surfaces. In *Proc. IEEE Workshop on Visual Motion*, pages 274–279, Princeton, NJ, October 1991.

[43] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):703–714, July 1991.

[44] W. B. Thompson. Combining motion and contrast for segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2:543–549, 1980.

[45] S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60:79–97, 1989.

[46] J. Y. A. Wang and E. H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3(5):625–638, September 1994.

[47] A. Waxman. An image flow paradigm. In *Proc. IEEE Workshop on Computer Vision: Representation and Control*, pages 49–55, Annapolis, MD, 1984.

[48] A. M. Waxman, B. Kamgar-Parsi, and M. Subbarao. Close-form solutions to image flow equations forr 3D-structure and motion. *International Journal of Computer Vision*, (3):239–258, 1987.

[49] A. M. Waxman and K. Wohn. Contour evolution, neighbourhood deformation and global image flow: Planar surfaces in motion. *Int. J. of Robotics Research*, 4:95–108, 1985.

[50] J. Weber and J. Malik. Robust computation of optical flow in a multi-scale differential framework. In *Proc. Int. Conf. on Computer Vision, ICCV-93*, pages 12–20, Berlin, Germany, May 1993.