Robust motion estimation under varying illumination

Yeon-Ho Kim[a],*, Aleix M. Martínez[b], Avi C. Kak[a]

[a]*School of Electrical and Computer Engineering, Purdue University, 2015 Neil Avenue, Columbus, OH 43210, USA*
[b]*Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA*

Abstract

The optical-flow approach has emerged as a major technique for estimating scene and object motion in image sequences. However, the results obtained by most optical flow techniques are strongly affected by motion discontinuities and by large illumination changes. While there do exist many separate techniques for robust estimation of optical flow in the presence of motion discontinuities and for dealing with the problems caused by illumination variations, only a few integrated approaches have been proposed. However, most of these previously proposed integrated approaches use simple models of illumination variation; a common assumption being that illumination changes by either just a multiplicative factor or just an additive factor from frame to frame, but not both. Some other previously proposed integrated approaches are limited to specialized tasks such as image registration or change recovery.

To remedy this shortcoming, this paper presents a new robust approach to general motion estimation in an integrated framework. Our approach deals simultaneously with motion discontinuities and large illumination variations. Our model of illumination variation is general, in the sense that it admits both multiplicative and additive effects.
© 2004 Published by Elsevier B.V.

Keywords: Motion estimation; Optical-flow; Illumination; Robust statistics

1. Introduction

The estimation of motion in images is a basic task in computer vision with many interesting applications. A primary goal in the field is to estimate the scene or object motion as precisely as possible. A classical way to compute the scene or object motion from a sequence of images is the optical-flow technique introduced by Horn and Schunck in [17]. The optical-flow can be defined as 'the apparent motion of the brightness pattern in an image sequence' [18]. The optical flow is a computational way of estimating the *motion filed*, which is the 'the two-dimensional (2D) projection of 3D motion on the image plane'.

The problem of motion estimation, in general, is made difficult by motion discontinuities, caused by relative motions between different objects in a scene, and by large illumination variations. When an object moves in front of another at a different velocity, that creates a motion discontinuity in the image plane which violates the assumptions embedded in the formulation of the classical method of Horn and Schunck. Black and Anandan [6] solve this problem by using a robust statistical framework which allows them to recover the structure of the motion as indicated by the majority of the pixels over a region while eliminating those pixels that are inconsistent with the majority. The latter pixels are referred to as 'outliers'.

As mentioned above, another basic assumption that often gets violated in the formulation of Horn and Schunck is that of constant illumination—the formulation assumes that an image pixel, representing a point on an object, does not change its brightness value from an instant of time t to the next instant of time $t + \delta t$. However, in a realistic scene this is almost never the case. A pixel can change its brightness value because an object moves (translates or rotates) to another part of the scene with different lighting or because the illumination of the scene (globally or locally) changes in time. The algorithm of Black and Anandan, if used on

* Corresponding author

E-mail addresses: yeonho@ecn.purdue.edu (Y.-H. Kim), aleix@ee.eng.ohio-state.edu (A.M. Martínez), kak@ecn.purdue.edu (A.C. Kak).

motion sequences with illumination variations, should be able to cope with small and localized illumination changes, because these will be seen as outliers. However, this method would not be adequate for image sequences with large or global illumination variations.

One way to deal with large illumination variation is to replace in the Horn and Schunck's formulation the assumption of constancy of pixel brightness values with a more realistic model. For example, Gennert and Negahdaripour [15] have assumed that the brightness at time $t + \delta t$ is related to the brightness at time t through a set of parameters that can be estimated from the image sequence. Their formulation, however, breaks down at motion discontinuities.

For obvious reasons, it would be desirable to solve these two problems—the problem caused by large motion discontinuities and the problem caused by large illumination variations—simultaneously. To be sure, there have been some studies in the past to fulfill this need [7,19,26,30]. To briefly review these contributions, the work of [26,30] uses the simplifying assumption that illumination changes by either just a multiplicative factor or just an additive factor from frame to frame, but not both. The work of Black et al. [7] uses a complicated framework that is not designed for general motion estimation but for recovering appearance changes in an image region due to a complex combination of events and processes, including motions, illumination variations, occlusions, dis-occlusions, etc. Finally, the work of Lai and Fang [19] uses an illumination variation model that comes closest to the one we use in our work reported here—this is the model that was proposed originally by Gennert and Negahdaripour [15]. However, the focus of the work reported by Lai and Fang is not motion estimation, but image alignment.

To remedy this shortcoming, this paper proposes a new integrated approach that incorporates Gennert and Negahdaripour's [15] formulation within the robust statistical framework of Black and Anandan [6]. This provides us with a clearly articulated illumination variation model embedded in a computationally well-organized and efficient framework for general motion estimation that can be expected to be simultaneously robust to both motion discontinuities and illumination changes. We show experimental results on synthetic and real data that demonstrate the superiority of our approach. Our results compare the performances of our approach with other methods that solve separately the problems caused by motion discontinuities and illumination variations.

2. Optical flow analysis

An optical flow field is a 2D field relating brightness patterns in an image recorded at one instant of time (say, t)

to the brightness patterns in the image recorded at the next instant of time $(t + \delta t)$.

The Horn and Schunck approach for recovering optical flow exploits the *data conservation* and the *spatial coherence* constraints. The data conservation constraint is derived from the observation that the brightness coming off object surfaces generally persists in time, although the position of the surface might change. If $I(x, y, t)$ is the pixel brightness at image coordinates (x, y) at time t, this constraint can be formally expressed as

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

where (u, v) are the horizontal and the vertical image velocities at (x, y) and δt is small. With a first-order Taylor series expansion of the right-hand side about (x, y, t), this constraint can be expressed as the minimization of the following integral form

$$E_D \int \int_R \rho(I_x u + I_y v + I_t) \, dx dy \tag{2}$$

where I_x, I_y and I_t are the partial spatial and time derivatives of the image brightness values, and ρ a prescribed function of the differential error term, a quadratic in the Horn and Schunck formulation. The quantity R refers to the image region over which this integration is carried out.

We can also assume that the motion of a surface changes gradually and thus enforce a smoothness constraint on the motion of the neighboring points in the image plane. Mathematically, this constraint can be expressed as the minimization of the form

$$E_S = \int \int_R \rho(\nabla(u, v)) \, dx dy \tag{3}$$

where ∇ denotes the gradient. This is known as the spatial coherence constraint.

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

$$E_D + \lambda E_S \tag{4}$$

where λ controls the relative importance of the data conservation and spatial coherence terms.

2.1. A robust estimation of the optical flow

The data conservation and the spatial coherency constraints of Horn and Schunck are often violated in real circumstances. For example, the assumption of image brightness constancy is violated when motion hound shadows, occlusions or specular reflections are present. Motion boundaries and occlusions also violate the smoothness (spatial coherence) constraint.

The problems caused by these violations are compounded by the fact that there is no reliable way to choose R. The larger the region of integration, the more likely that

the two assumptions would be violated. At the same time, making this region too small would insufficiently constrain the solution. This is known as the aperture problem.

The aperture problem is particularly susceptible to the choice of a quadratic for ρ because it averages all of the motion present inside the region R. With a quadratic choice for ρ, the integral expressions in (2) and (3) will always be affected by all the motions that are included in the region R, and could even become dominated by these other motions that are unrelated to the motion at the center of R. The contributions to the integral expressions by the pixels that are not experiencing the same motion as the one at the center of R will be referred to as *outliers*.

Black and Anandan [6] have shown how one can deal with the problem of outliers by using a *robust estimator* for ρ. The basic notion behind the use of a robust estimator is to obtain a statistical characterization of the data that is immune to the outliers. Several approaches to robust estimation have been proposed for this task. These approaches span least median squares estimation, voting-based methods similar to the Hough method, the M-estimators, etc. For reasons of computational complexity, the last approach—the one based on M-estimators—is favored by most researchers.

Focusing on the M-estimators, Darrel and Pentland [9] have used the truncated quadratic estimator; Odobez and Bouthemy [26] have used Tukey's biweight estimator; Sawhney et al. [29] have used the Geman and McClure estimator; and the Lorentzian M-estimator has been used by Ayer and Sawhney [2] and Sawhney et al. [29]. In all of these contributions, robust estimation is used for only the weights assigned to the observed data within an iteratively re-weighted least squares (IRLS) framework.

Rather than the IRLS method in which robust estimation is carried out on a certain set of weights only, the work of Black and Anandan [6] applies robust estimation to the pixels directly through a straightforward minimization scheme. Black used the Lorentzian M-estimator and presented the following reasons for this choice among the different M-estimators available [5]:

- The estimator must be twice differentiable.
- The line-process based estimation-theoretic methods can be extended to the case when an objective function based on the Lorentzian is used. The concept of a line-process has been widely used to solve regularization problems in the presence of discontinuities in many computer vision applications such as segmentation, edge detection, optical flow, image reconstruction [8, 12–14,27,31].
- The estimator must allow us to determine the statistical distribution of the measurement errors with which the estimator works optimally.
- When an objective function that uses a certain M-estimator is non-convex, a simple gradient descent based optimization algorithm cannot search for a global

optimal solution. In that case, specifically, we can use the *Graduated Non-Convexity* algorithm of Blake and Zisserman [8]. This algorithm constructs a sequence of convex approximations to the non-convex objective function. The estimator is provided with a parameter that controls the shapes used for the approximations.

Our work reported here also uses the Lorentzian as the M-estimator for the same reasons as mentioned above.

This estimator is defined as:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right). \tag{5}$$

where σ controls the weighting to be given to x. At this time there does not exist an analytical approach for specifying a value for this parameter. Black and Anandan have proposed values for a in [5,6] that we will use in our work reported here.

While it protects us against the outliers, a robust estimator, such as the one shown here, does extract a computational price: it can cause the objective function in Eq. (4) to become non-convex in the (u, v) space. This minimization problem can be solved by using the *Graduated Non-Convexity* algorithm originally advanced by Blake and Zisserman [8]. Starting from a larger than the desired value for sigma, this method constructs a series of approximations for the objective function, each yielding through minimization a more accurate estimate for the velocity (u, v). Large values of sigma give us convex approximations of the objective function in which the minimum can easily be located using a minimization technique such as simultaneous over-relaxation (SOR). Subsequently, one iterates the procedure using smaller and smaller values of σ so that at the end of each iteration a more accurate approximation to the global minimum of the original non-convex objective function is obtained. We will discuss this approach in Section 2.3, especially as it applies to our work.

2.2. Including the radiometric information

The formulation discussed so far does not pay attention to other image factors that might also be time dependent. For example, the brightness of a pixel might change on account of spatially and/or temporally variant illumination in a scene [16,18,23,34]. When only a small number of the pixels in a region R are affected by, say, the illumination change, the statistical framework introduced above can be robust to them by treating these pixels as outliers. However, when the illumination change affects a large number of the pixels of R, the framework presented above will be inappropriate.

Some researchers have proposed to first filter the image as an alternative to handle the illumination changes in a sequence of images [1,10,11,25]. This can be achieve by either: (i) spatial smoothing or phase filtering, or (ii) edge

detection. The former has the disadvantage of inaccuracies at the boundaries, while the latter has the disadvantage of only being able to compute the motion fields at the boundaries of the object.

To properly address the problems caused by large illumination variations, we need a formulation that explicitly takes into account the radiometric cues, such as the one provided by Gennert and Negahdaripour [15] and Negahdaripour [24]. They have extended the usual definition of optical flow to include radiometric changes into the computation of the flow. This definition extends the 2D motion field vector (u, v) to a 3D transformation field given by $(u, v, \delta e)$. The last component, δe, describes the radiometric transformation of the image sequence. Gennert and Negahdaripour replaced the brightness constancy constraint of Eq. (1) by the following more general form

$$I(x + u\delta t, y + v\delta t, t + \delta t) = M(x,y,t)I(x,y,t) + C(x,y,t) \quad (6)$$

in which the brightness at a pixel in two consecutive images is related via the motion parameters u and v and the radiometric parameters M and C, as shown.

To be sure, there do exist other formulations also for modeling the illumination variations in a scene [7,16,21,22]. These formulations decompose image changes on the basis of one or more physical phenomena into a nonlinear set of physical events that cause the image changes. By contrast, the formulation of Negahdaripour describes the illumination-caused image changes as a combination of a multiplicative effect and an additive effect regardless of the underlying physical events. This makes the formulation of Negahdaripour much more suitable for incorporation in an estimation-theoretic framework for motion estimation especially when only two consecutive images are used as data.

The data conservation constraint corresponding to Eq. (6) is

$$E_D = \int\int_R \rho(I_x u + I_y v + I_t - (Im_t + c_t)) \quad (7)$$

where m_t and c_t are the time derivatives of M and C, respectively. These derivatives can be derived from the observation that for small δt, M is expected to be close to 1 and C close to 0. Since we are dealing with incremental changes, we can write $M = 1 + \delta m$ and $C = \delta c$, where $\delta m \to 0$ and $\delta c \to 0$ as $\delta t \to 0$. This allows us to define the time derivatives used above as:

$$m_t = \lim_{\delta t \to 0} \frac{\delta m}{\delta t} \quad \text{and} \quad c_t = \lim_{\delta t \to 0} \frac{\delta c}{\delta t}. \quad (8)$$

Since, we now have two more variables to estimate, the smoothness constraint, needs also to include the following minimizations:

$$E_M = \int_R \rho(\nabla m_t) \quad \text{and} \quad E_C = \int_R \rho(\nabla c_t). \quad (9)$$

By imposing a smoothness constraint on the illumination variables, one also prevents M and C to model other physical events not related to the lighting of the scene; e.g. noise (which will be tackle by the robust statistics). The motion of a moving square (with constant texture) cannot be represented by M and C either, because its brightness changes sharply at several local areas.

The problem of estimating optical flow can finally be defined as that of minimizing the functional

$$E = \lambda_D E_D + \lambda_S E_S + \lambda_M E_M + \lambda_C E_C \quad (10)$$

where λ_D, λ_S, λ_M and λ_C are the weights of the respective contributions.

When using the least-square formulation (i.e. $\rho(x) = x^2$), the above definition corresponds to the approach presented in [15]. Furthermore, in the more restricted case where $M = 1$ and $C = 0$ (i.e. $m_t = c_t = 0$) this formulation reduces to the approach proposed in [17].

2.3. A framework for the robust estimation of optical flow in illumination variant scenes

The method of [15] successfully couples the motion transformation of the object with the radiometric cues. However, the assumptions embedded in the constraint equations (i.e. the data conservation and spatial coherence) would still be violated at motion discontinuities in a sequence of images.

We note, however, that when formulating the approach of Section 2.2, we intentionally kept the estimator (ρ) undefined. This allows us to use the robust statistical framework (defined in Section 2.1) in the re-defined optical flow method of Section 2.2. In the rest of this section, we will adapt the SOR technique to the case of the revised optical flow equations of the previous section. For simplicity we show the derivation that minimizes the objective function of Eq. (3) with respect to u only. The same treatment also applies to the minimizations with respect to the other parameters, v, m_t and c_t.

We can write as follows the iterative equation for minimizing E at the image pixel i at step $r+1$ using the nonlinear gradient-descent method presented in [8]:

$$u_i^{r+1} = u_i^r - w \frac{1}{T(u_i)} \frac{\partial E}{\partial u_i} \quad (11)$$

where w is an over-relaxation parameter that is used to over-correct the estimate of u_i^{r+1}. The parameters v_i^{r+1}, $m_{t_i}^{r+1}$ and $c_{t_i}^{r+1}$ can be updated in the same way as u_i^{r+1}, that is by replacing u in the above equation with v, m_t and c_t, respectively. It has been shown that with $0 < w < 2$ the method converges [33].

The partial derivative $\frac{\partial E}{\partial u_i}$ is:

$$\frac{\partial E}{\partial u_i} \approx \sum_{s \in R} \left[\lambda_D I_x \psi(I_x u_i + I_y v_i + I_t - Im_{t_i} - c_{t_i}, \sigma_D) \right.$$
$$\left. + \lambda_S \sum_{s \in \mathscr{R}_s} \psi(u_i - u_s, \sigma_S) \right] \tag{12}$$

where $\mathscr{R}_s$ represents the set of 4-neighboring pixels (left, right, up and down) and $\psi(x, \sigma) = 2x/2\sigma^2 + x^2$ is the derivative of the Lorentzian. The term $T(u_i)$ used in Eq. (11) is an upper bound on the second partial derivative of E and is given by

$$T(u_i) = \frac{\lambda_D I_x^2}{\sigma_D^2} + \frac{4\lambda_S}{\sigma_S^2} \geq \frac{\partial^2 E}{\partial u_i^2} \tag{13}$$

The values of I_x, I_y and I_t can be calculated efficiently using the simple discrete approximation at a center of a cube formed by eight measurements as presented in [17]:

$$I_x \approx \frac{1}{4}(I_{i+1,j,k} + I_{i+1,j,k+1} + I_{i+1,j+1,k} + I_{i+1,j+1,k+1})$$
$$- \frac{1}{4}(I_{i,j,k} + I_{i,j,k+1} + I_{i,j+1,k} + I_{i,j+1,k+1})$$

$$I_y \approx \frac{1}{4}(I_{i,j+1,k} + I_{i,j+1,k+1} + I_{i+1,j+1,k} + I_{i+1,j+1,k} + 1)$$
$$- \frac{1}{4}(I_{i,j,k} + I_{i,j,k+1} + I_{i+1,j,k} + I_{i+1,j,k+1})$$

$$I_t \approx \frac{1}{4}(I_{i,j,k+1} + I_{i+1,j,k+1} + I_{i,j+1,k+1} + I_{i+1,j} + 1, k + 1)$$
$$- \frac{1}{4}(I_{i,j,k} + I_{i+1,j,k} + I_{i,j+1,k} + I_{i+1,j+1,k}) \tag{14}$$

where $I_{i,j,k}$ is the image pixel at the coordinate represented by x, y and t as indicated by the values of i, j and k.

Using the same approach as in [6,8], we can now find the global minimum of the objective function, E, by first choosing a large enough σ that creates a convex approximation for E. This will yield a minimum for the objective function. In general, this solution will not be the desired one, but it will serve as a good starting approximation. We can then decrease the value of σ and find a new minimum for E. When doing this, however, we will use the minimum obtained previously as an initial guess. By repeating this process several times, we can converge to the correct solution [8].

When there is no illumination change in the image, we want $M = 1$ and $C = 0$, in which case the method described above would reduce to Black and Anandan algorithm.

As with all methods advanced to date for the computation of optical flow, the procedure outlined above will yield good results only when the object displacements between consecutive images are small; ideally no larger than a pixel. In order to detect large motions, a coarse-to-fine strategy [1,4,6] can be used. For the experiments reported below, a pyramid method was employed. In this approach we begin with a reduced-resolution representation of the images so that the small-displacement assumption is satisfied. The optical flow is computed for the low-resolution images and then projected to the next level of the pyramid where the images in the sequence have a higher resolution. At each level of the pyramid, the optical flow computed from the lower level is used to warp the images in the sequence so that the small displacement assumption holds for the new resolution. This process is repeated until the flow has been computed at the original resolution. The warping process can be formally described as:

$$I_{\text{warped}}(x, y) = I(x - u(x, y), y - v(x, y)). \tag{15}$$

3. Experimental results

Using the formulation presented here, we will now show optical flow results on synthetic and real data for which the ground-truth is known. We will compare our results with those obtained using the algorithm of Black and Anandan [6]; the algorithm of Gennert and Negahdaripour [15]; Negahdaripour's close-form solution to the revised definition of optical flow (which incorporates radiometric changes) [24]; Lucas and Kanade's approach [20]; and the classical method of Horn and Schunck [17].

We will then show the results on two consecutive images of the well-known *Flower Garden* sequence. The reason we do not show comparative results on these images is because the ground truth is not known for them. Yet, since the different objects in the scene are at very different depths, one has an intuitive sense of their relative motions. So it is instructive to examine, albeit in a qualitative sense only, the results obtained on these images. To introduce a large illumination variation in this data, we modified the second image by multiplying it with the Gaussian function shown in Fig. 1(b) and adding a gray scale offset.

As a final experiment, we will show the results obtained on a two frame sequence where a can move in front of a stationary background, and where the illumination source is also moving. This sequence incorporates thus a real

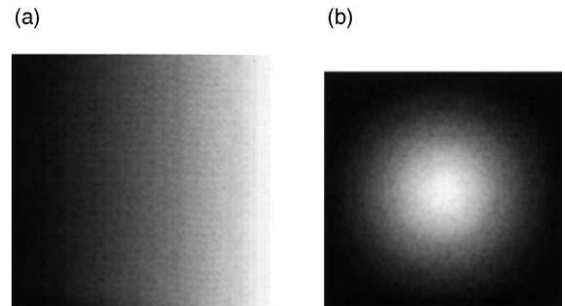(a)   (b)

Fig. 1. Shown here are the multipliers used in the *Square* sequences (a) and in the *Flower* sequence (b).

illumination change caused by a moving light source which located at the right hand-side of the camera.

The data sets used for the experiments consist of two synthetic sequences and one real:

- *Square 1*: This two 50×50 pixels frame sequence has a 40×40 pixel square moving at a velocity of (1, 1) pixels. The value of the pixels in the background is zero. The pixels of the square are set to 255. In the second image of this sequence we have incorporated an illumination change in the form of a multiplier. The multiplier is shown in Fig. 1(a).

- *Square 2*: A 40×40 pixel square moves at a velocity of (1, 1) pixels per image. The pixels in the background are set to zero. The pixels of the square take random values between zero and 255. Again, we have incorporated an illumination change in the second image of this sequence in the form of a multiplier; shown in Fig. 1(a).

- *Flower*: This two-frame sequence was created from a real scene of a flower. The sequence was constructed by cropping a 100×100 part from the original image at two different locations, separated by a motion of (1, 1) pixels. This simulates a (1, 1) velocity motion. The second image of this sequence incorporates an illumination
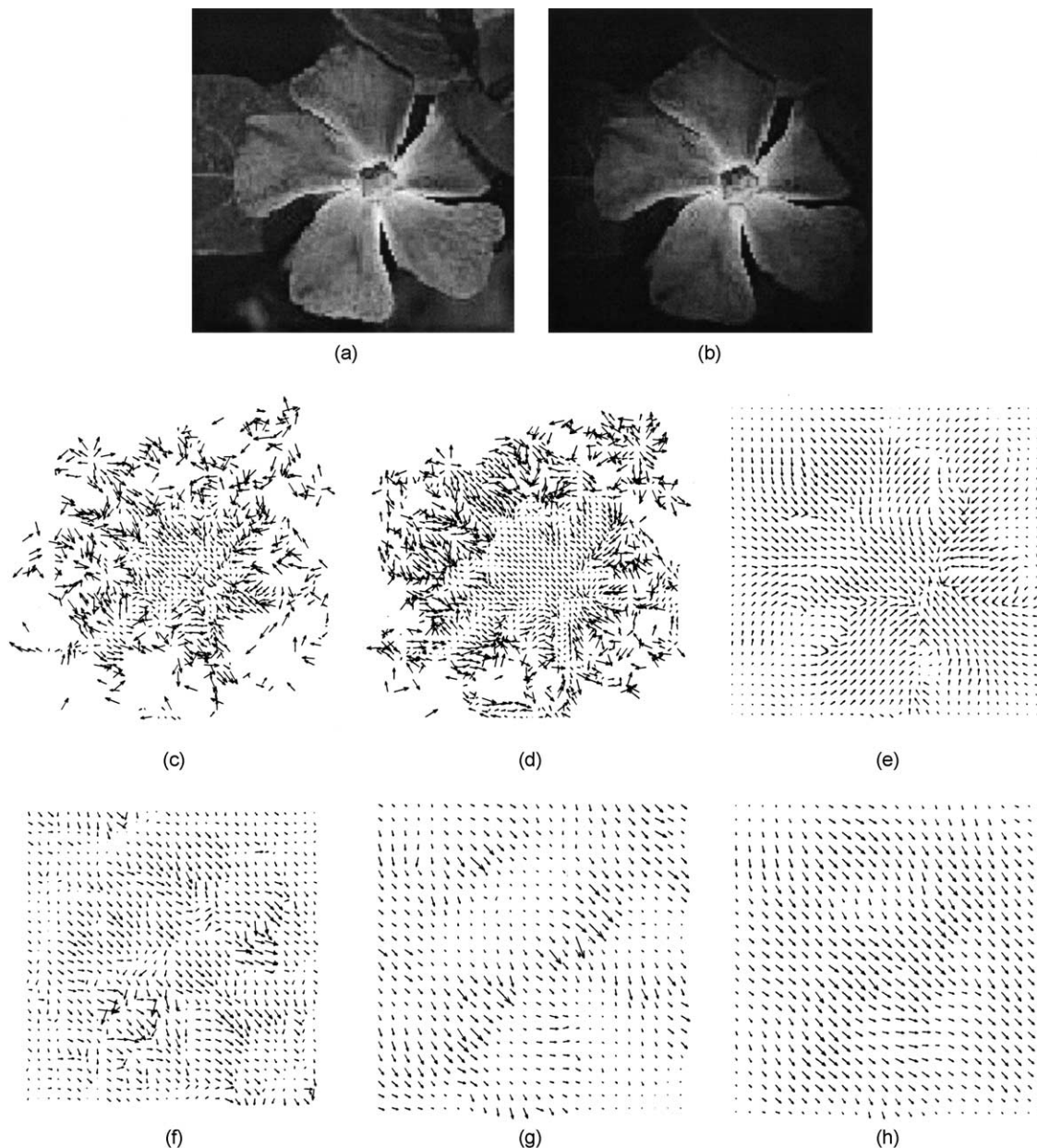

Fig. 2. The pair of images in the *Flower* sequence. Pictorial depiction of motion vectors obtained with the algorithms of (c) Horn and Schunck, (d) Lucas and Kanade, (e) Black and Anandan, (f) Negahdaripour, (g) Gennert and Negahdaripour, and (h) this paper.

Table 1
Comparative performance results for the *Square 1* image sequence

	Err_{ang}	$Stdv_{ang}$	$Density_{ang}$ (%)	Err_{vel}	$Stdv_{vel}$	$Density_{vel}$ (%)
Horn and Schunck	35.987	13.246	1.89	1.204	0.558	100.00
Lucas and Kanade	19.424	27.043	1.32	0.886	0.660	23.20
Black and Anandan	21.347	14.335	75.61	0.479	0.205	100.00
Negahdaripour	19.844	16.499	2.96	0.825	0.484	3.94
Gennert and Negahdaripour	35.039	12.658	53.07	1.045	0.263	100
This paper	**15.221**	**8.701**	75.61	**0.472**	**0.191**	100

change in the form of a multiplier. The sequence is shown in Fig. 2 and the multiplier is shown in Fig. 1(b). Note that since the motion is created by shifting the entire cropped region, we do not really have any relative motions between the flower and the background in the resulting image pair. In that sense, the results shown for this data are more for testing the robustness of computations for large illumination variations only[1].

To evaluate the different optical-flow results, we define two error measurement functions, one that calculates the velocity magnitude error and another that computes the orientation of motion error over all pixels. Other researchers have used error measurements which integrate both errors (velocity and orientation) in one. However, integrated measures of error tend to hide the contribution of each component. Since we are specifically interested in knowing as to what extent our method performs better than others, either in estimating the velocity or the orientation or both, we have found more appropriate to show both error values separately.

We record three different aspects of velocity error: (1) the mean velocity magnitude error associated with optical-flow estimation; (2) the standard-deviation of the error; and (3) the number of pixels used to calculate the mean divided by the total number of pixels in the image frame.

The mean velocity magnitude error is given by

$$Err_{vel} = \frac{\sum_{i=1}^{n_v} |\,||O_i|| - ||G_i||\,|}{n_v} \qquad (16)$$

where $||O_i||$ is the magnitude of the velocity vector at the pixel position i of the optical flow result and $||G_i||$ is the magnitude of the corresponding ground-truth velocity vector at the same pixel. The quantity n_v is the total number of pixels where the optical flow is calculated.

The standard deviation associated with this mean value will be denoted by Std_{vel}.

In general, it is also necessary to show the number of pixels that were used to compute the above error vales. This

is needed because not all the methods can compute a dense estimation of the optical flow, because they require matrix inversions which might not have solution in an area of the image that, for example, has no gradient values on account of absence of any texture. The number of pixels where optical flow is computed will be expressed as a percentage of the total number of pixels in the image frame. This percentage value will be referred to as 'density'.

The second group of error measurements records the errors in the orientation of the velocity vectors as obtained from the optical-flow calculations vis-a-vis the orientations of the corresponding velocities in the ground-truth data. As with velocity magnitudes, the orientation errors will be measured in the form of average angle error, as given by

$$Err_{ang} = \frac{\sum_{i=1}^{n_o} \arccos\left(\frac{O_i \cdot G_i}{||O_i|| ||G_i||}\right)}{n_o}, \qquad (17)$$

the standard deviation associated with this mean, and the density. In this formula $O_i \cdot G_i$ represents the dot product between these two vectors, and n_o is the total number of pixels that either have non-zero velocity magnitudes in both the ground-truth and the optical-flow estimation or zero velocity magnitudes in both values. So, if a pixel has zero velocity in the ground-truth but non-zero in the estimation, or vice versa, that pixel will not be included in the count n_o.

Table 1 shows the performance of all the methods for the *Square 1* image sequence. To compute the results shown in Table 1, we have used the values $\lambda_D = \lambda_S = \lambda_M = 1$ and $\lambda_C = 10^8$ for those methods that deal with illumination changes and $\lambda = 1$ for the methods that do not include the radiometric information in their formulation. This choice was made on the basis that the illumination change included had only a multiplicative component[2]. From the results in Table 1, it is clear that the method of Negahdaripour [24] and our method perform the best. However, the former

---

[1] It would, of course, be trivial to create a real two-image sequence that incorporates different relative motions for different objects in the scene. But it would be much too difficult to generate the ground-truth information for such a sequence in order to apply the error measures defined later in this section.

[2] In [24], Negahdaripour shows that the multiplicative radiometric parameters are related to diffuse shading effects and that the additive radiometric parameters are related to illumination offsets in highlighted regions. Illumination highlighting may be caused by spotlight effects or the presence of specular surfaces in a scene. For this synthetic sequence, we can assume that there are no specularly reflecting surfaces in the scene and that therefore the additive contributions to the varying illumination can be ignored. To neglect the offset term, we do over-smoothing on it by assigning a large value to λ_C.

Table 2
Comparative performance results for the *Square 2* image sequence

	Err$_{ang}$	Stdv$_{ang}$	Density$_{ang}$ (%)	Err$_{vel}$	Stdv$_{vel}$	Density$_{vel}$ (%)
Horn and Schunck	69.878	56.591	66.87	1.043	0.763	100.00
Lucas and Kanade	44.850	49.128	74.05	0.768	0.371	95.56
Black and Anandan	52.260	49.557	74.53	0.718	0.315	100.00
Negahdaripour	11.258	10.538	94.74	0.939	0.257	100.00
Gennert and Negahdaripour	14.430	16.184	75.52	0.596	0.401	100.00
This paper	**8.653**	**8.301**	75.61	**0.490**	**0.248**	100

Table 3
Comparative performance results for the *Flower* sequence

	Err$_{ang}$	Stdv$_{ang}$	Density$_{ang}$ (%)	Err$_{vel}$	Stdv$_{vel}$	Density$_{vel}$ (%)
Horn and Schunck	78.144	65.407	39.96	1.388	0.662	100
Lucas and Kanade	73.061	64.264	62.43	1.268	0.782	100
Black and Anandan	73.264	57.6	97.82	0.531	0.386	100
Negahdaripour	18.652	21.086	99.32	0.562	0.313	100
Gennert and Negahdaripour	14.712	16.236	98.97	0.822	0.283	100
This paper	9.81	9.26	98.87	0.595	0.31	100

yields optical flow values at less than 4% of the pixels, whereas our method computes the flow at all the pixels—100%, as evidenced by the density value for the velocity magnitude. The density for the orientation with our method is only 75.61% because, as stated earlier, the orientation errors are calculated at only those pixels where the velocity magnitude is non-zero or zero in both the ground-truth and the estimation value. Since the number of pixels outside of the central square that moves are roughly 25%, that explains the difference between the two densities. The reason that most of the method yield low densities is because they depend on the gradient of gray levels in the image. In this test sequence, the gradient is zero inside the moving square in the image.

Table 2 does the same thing as Table 1 but for the *Square 2* image sequence. As mentioned earlier, this image sequence differs from *Square 1* in that the central square that moves has random texture inside it. It is clear that our method outperforms the other methods, both for velocity magnitudes and for orientations. The λ's value are the same used above.

Finally, for real data, Table 3 shows the results for the *Flower* sequence. In this sequence, the values for the λ were the same as those used above in the *Square 1* image sequence except for λ_M which was now set to 10 that produces the best result for this experiment[3]. In this case three of the methods (Black and Anandan, Negahdaripour and the method described in this paper) do a reasonably good job in estimating the velocities. However, only our method is able to successfully estimate the direction of

motion with some precision. Pictorial depictions of the motion vectors obtained for this sequence are presented in Fig. 2 for all the methods.

As promised at the beginning of this section, Fig. 3 shows the two images of the famous Flower Garden sequence used here, and the results with the method described in this paper and the algorithms of Black and Anandan, Gennert and Negahdaripour, and Negahdaripour. As the observer can see, the different objects in the scene are at very different depths. The tree in the immediate foreground has the maximum motion. The pixels corresponding to the flower bed behind the tree experience the next largest motion, followed the houses, and, finally, the sky pixels undergo the smallest motions. Our results, shown in the form of arrows (Fig. 3(c)), are in keeping with these differing motions for the different objects. In this case, $\lambda = 0.5$ for Black and Anandan, and $\lambda_D = 1$ $\lambda_S = 0.5$, $\lambda_M = 10$, and $\lambda_C = 100$ for all other methods.

As a final example, we would like to show results for a real sequence of two images with a real illumination change. This sequence corresponds to a scene of a stationary background and a can in the foreground. Two images were taken, the second after the can had moved along the plane parallel to the scene. Because the objects of this scene are at different depth, their motion with respect to the camera (which does not move) is different. The can should have a constant motion towards the right. The background should have zero motion. We show in Fig. 4, the results obtained with our method. For this example, we also show comparisons with the methods of Black and Anandan, Gennert and Negahdaripour, and Negahdaripour. However, since no ground truth is available for a meaningful comparison, the comparison can only be visual. We can see that our algorithm was not affected by neither motion discontinuities or illumination changes. Also note that the motion within the can (foreground) and that of

---

[3] There exists no clear analytical formulation to decide the value of smoothness parameters for the radiometric terms and the motion terms. These parameters usually need to be tuned experimentally depending on the spatial properties of image intensities. This is a general disadvantage of any global method that needs explicit smoothness constraints for regularization.
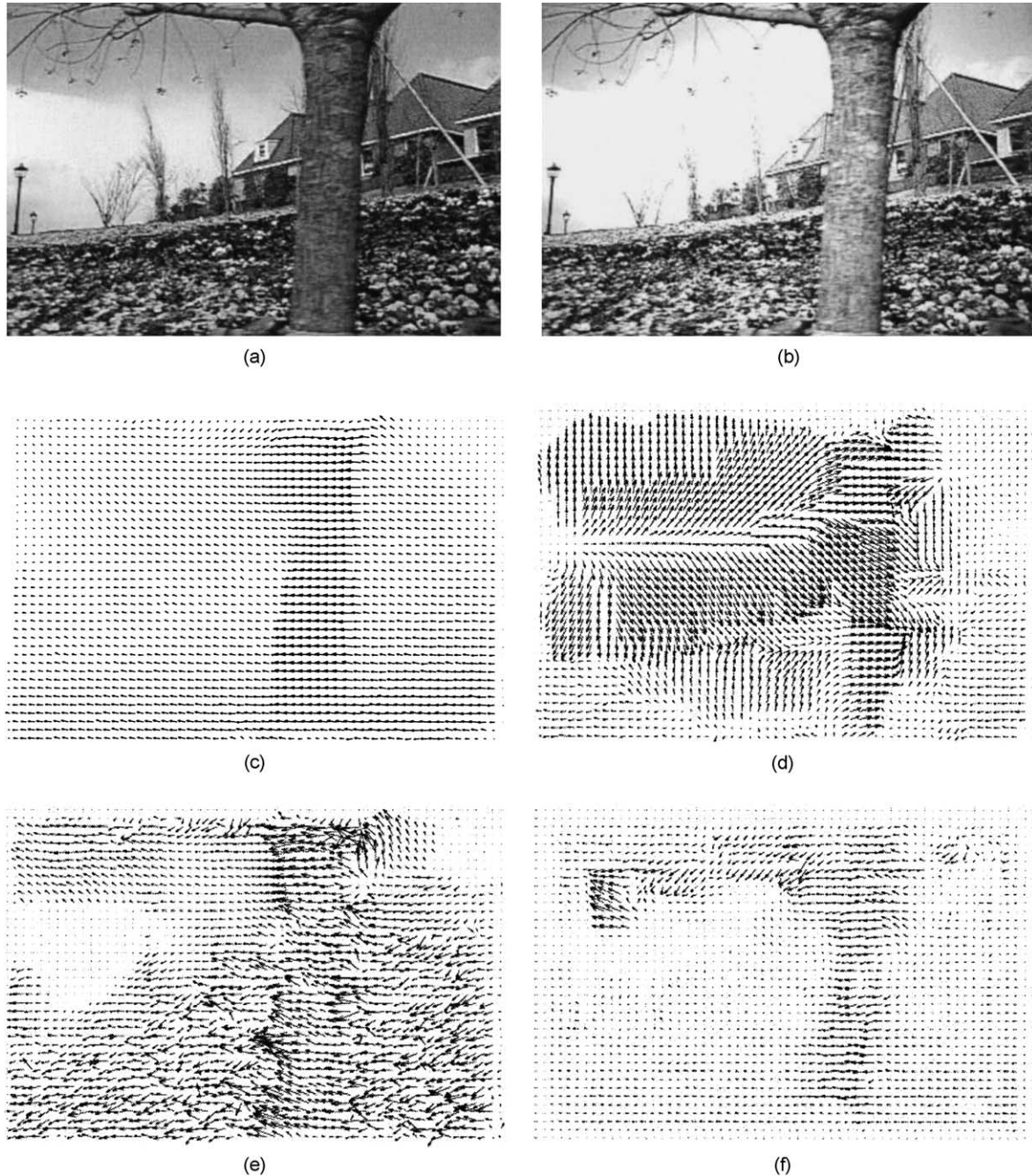
Fig. 3. (a–b) The two images of the *Flower Garden* sequence used. Results as obtained by the method of (c) this paper, (d) Black and Anandan, (e) Gennert and Negahdaripour, and (f) Negahdaripour.

the background are quite stable and constant. In this example, $\lambda = 1$, and $\lambda_D = \lambda_S = 1$, $\lambda_M = \lambda_C = 100$.

4. Conclusions

We have presented in this paper a new approach for estimating optical flow when image sequences are recorded under conditions of varying illumination and when there are large motion discontinuities present. To achieve this, we reformulated the approach of Gennert and Negahdaripour [15] within the robust statistical framework proposed by Black and Anandan [6]. As a result, our method is robust not only with respect to illumination changes, but also with respect to motion discontinuities.

The experimental results reported in this paper show the superiority of our method when illumination changes significantly from one image to the next. We have shown this via statistical comparisons on a set of synthetic image sequences for which the ground truth is known.
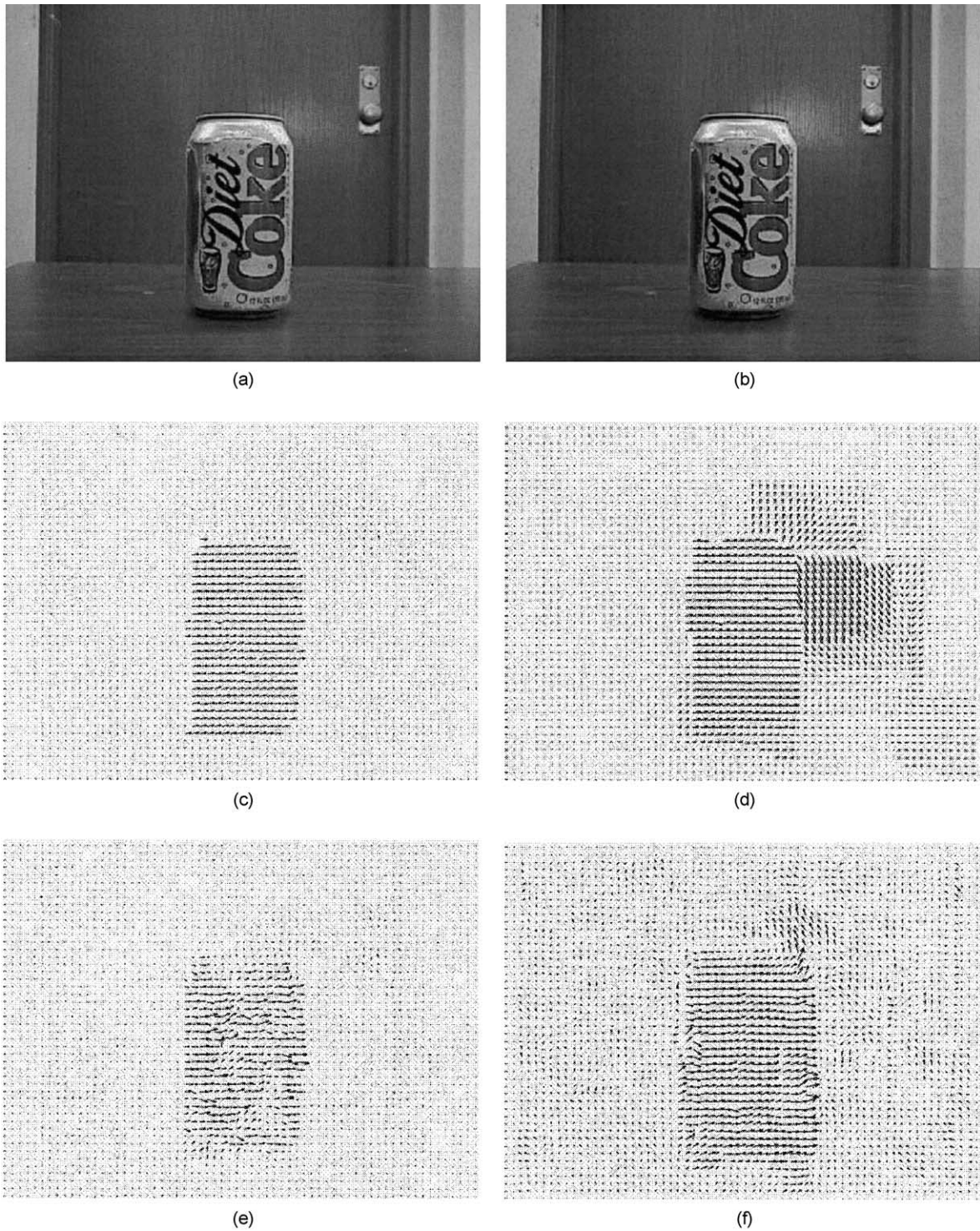
Fig. 4. (a–b) The two images of our sequence. Results obtained by the method of (c) this paper, (d) Black and Anandan, (e) Gennert and Negahdaripour, and (f) Negahdaripour.

Furthermore, we have presented results on two real image sequences. The first of these incorporates a known synthetic illumination change. The second real image sequence has a real illumination change that is caused by physically moving the illumination source. We have shown that our method can still produce good estimates of the optical flow despite the illumination changes in these two sequences.

Acknowledgements

References

[1] P. Anandan, A computational framework and an algorithm for the measure of visual motion, International Journal of Computer Vision 2 (3) (1989) 283–310.

[2] S. Ayer, H.S. Sawhney, Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding, IEEE International Conference on Computer Vision 1995; 777–784.

[3] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision 12 (1) (1994) 43–77.

[4] J. Bergen, P. Anandan, K. Hanna, R. Hingorani, Hierarchical model-based motion estimation in: G. Sandini (Ed.), Proceedings of Second European Conference on Computer Vision, vol. 588 of Lecture Notes in Computer Science, Springer, Berlin, 1992, pp. 237–252.

[5] M.J. Black, Robust incremental optical flow, PhD thesis, Yale University, 1992.

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

[7] M.J. Black, D.J. Fleet, Y. Jacoobs, Robustly estimating changes in image appearance, Computer Vision and Image Understanding 78 (1) (2000) 8–31.

[8] A. Blake, A. Zisserman, Visual reconstruction, The MIT Press, Cambridge, MA, 1987.

[9] T. Darrel, A. Pentland, Robust estimation of a multi-layered motion representation In, IEEE Workshop on Visual Motion 1991, pp. 173–178.

[10] D.J. Fleet, A.D. Jepson, M. Jenkin, Phase-based disparity measurement, Computer Vision and Image Understanding 53 (2) (1991) 198–210.

[11] D.J. Fleet, A.D. Jepson, Stability of phase information, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (12) (1993) 1253–1268.

[12] D. Geiger, F. Girosi, Parallel and deterministic algorithms form MRFs: surface reconstruction, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (5) (1991) 401–412.

[13] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1984) 721–741.

[14] D. Geman, G. Reynolds, Constrained restoration and the recovery of discontinuities, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (3) (1992) 367–383.

[15] M.A. Gennert, S. Negahdaripour, Relaxing the brightness constancy assumption in computing optical flow, A.I. Memo No. 975, M.I.T., June 1987.

[16] H.W. Haussecker, D.J. Fleet, Computing optical flow with physical models of brightness variation, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 661–673.

[17] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artificial Intelligence 17 (1981) 185–203.

[18] B.K.P. Horn, Robot vision, MIT Press, Cambridge, MA, 1986.

[19] S.-H. Lai, M. Fang, Robust and efficient image alignment with spatially varying illumination models In Proceedings of In, IEEE Conference on Computer Vision and Pattern Recognition 2 1999, pp. 167–172.

[20] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision In, Proceedings of Seventh IJCAI, Vancouver, BC, Canada 1981, pp. 674–679.

[21] N. Mukawa, Estimation of shape, reflection coefficients and illuminant direction from image sequences In, Proceedings of International Conference on Computer Vision 1990, pp. 507–512.

[22] N. Mukawa, Optical-model-based analysis of consecutive images, Computer Vision and Image Understanding 66 (1) (1997) 25–32.

[23] H. Nagel, On a constraint equation for the estimation of displacement rates in image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1) (1989) 13–30.

[24] S. Negahdaripour, Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (9) (1996) 961–979.

[25] O. Nestares, R. Navarro, Probabilistic estimation of optical flow in multiple band-pass directional channels, Image and Vision Computing 19 (2001) 339–351.

[26] J.-M. Odobez, P. Bouthemy, Robust multiresolution estimation of parametric motion models, Visual Communication and Image Representation 1995; 348–365.

[27] A. Rangarajan, R. Chellappa, A continuation method for image estimation using the adiabatic approximation, in: R. Chellappa, A.K. Jain (Eds.), Markov Random Fields: Theory and Applications, Academic Press, New York, 1993.

[28] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.

[29] H.S. Sawhney, S. Ayer, M. Gorkani, Model-based 2D and 3D dominant motion estimation for mosaicing and video representation, IEEE International Conference on Computer Vision 1995; 583–590.

[30] D. Shulman, J.-Y. Herve, Regularization of discontinuous flow fields, In Proceedings of IEEE Workshop on Visual Motion, Irvine, CA, 1989.

[31] T.Y. Tian, M. Shah, Motion estimation and segmentation, Machine Vision and Applications 9 (1) (1996) 32–42.

[32] P. Treves, J. Konrad, Motion estimation and compensation under varying illumination In, Proceedings of IEEE International Conference on Image Processing 1 1994, pp. 373–377.

[33] R.S. Varga, Matrix iterative analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[34] A. Verri, T. Poggio, Motion field and optical flow: qualitative properties, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (5) (1989) 490–498.