Revised Definition of Optical Flow: Integration of Radiometric and Geometric Cues for Dynamic Scene Analysis

Shahriar Negahdaripour

Abstract—Optical flow has been commonly defined as the apparent motion of image brightness patterns in an image sequence. In this paper, we propose a revised definition to overcome shortcomings in interpreting optical flow merely as a geometric transformation field. The new definition is a complete representation of geometric and radiometric variations in dynamic imagery. We argue that this is more consistent with the common interpretation of optical flow induced by various scene events. This leads to a general framework for the investigation of problems in dynamic scene analysis, based on the integration and unified treatment of both geometric and radiometric cues in time-varying imagery. We discuss selected models, including the generalized dynamic image model in [21], for the estimation of optical flow. We show how various 3D scene information are encoded in, and thus may be extracted from, the geometric and radiometric components of optical flow. We provide selected examples based on experiments with real images.

Index Terms—Motion vision, optical flow, time-varying image analysis, dynamic visual cues, integration of radiometric and geometric cues.

1 Introduction

A static image encodes two types of more or less distinct information about a scene. *Geometric cues*, based on the projection geometry, constrain the position of points in the scene in terms of the coordinates of their projections onto the image plane. *Radiometric cues* are tied to a large number of scene properties, including illumination condition, medium properties, sensor spectral response characteristics, as well as shape, position, and reflectance characteristics of the scene surfaces.

In selected *Shape-from-X* techniques, certain radiometric or geometric cues in a *static image* are exploited to recover the 3D shape of the scene (see Fig. 1). In contrast, dynamic scene analysis involves the recovery of 3D information about a scene from time-varying imagery. By time-varying imagery, we refer to spatiotemporal data acquired by varying *passively* or *actively*, from one frame (sampling time) to the next, some aspect(s) of the imaging condition: Illumination, sensor, scene, or in some special applications the medium characteristics. (Binocular and photometric stereo vision have been classified as problems in dynamic scene analysis since they require two or more images from different viewpoints or lighting conditions. Furthermore, we are not concerned about the distinction between passive and active variations in imaging conditions in this paper.)

Referring to Fig. 2, $e(r)$ denotes the image brightness of some scene point, where $\mathbf{r} = [x, y, t]^T$ is a point in the spatiotemporal volume. In the next frame acquired at a later time $t + \delta t$, the scene point may move relative to the camera and thus projects onto a new point $\mathbf{r} + \delta\mathbf{r} = [x + \delta x, y + \delta y, t + \delta t]^T$. In addition, the image brightness of this point can vary by $\delta e = e(\mathbf{r} + \delta\mathbf{r}) - e(\mathbf{r})$, due to relative motion and/or other scene events. A large body of work in motion vision literature is concerned with estimating the geometric transformation field $[\delta x, \delta y]^T$ from an image sequence, the problem of *optical flow computation*. To do this, it is typically assumed that the optical flow is primarily due to the *motion of brightness patterns* from one image to the next. Knowledge of image displacement field permits the calculation of object and/or camera motion, 3D scene shape in the form of a depth map, as well as segmentation of the scene based on image motion and depth discontinuity boundaries (e.g., [5], [15], [16], [26], [27]). In contrast, other methods exploit the radiometric transformation field δe, the *optical flow induced by light source motion*, to extract 3D shape information[1] (e.g., [4], [14], [31], [35], [36]). Instead of optical flow, other terminologies including *shading, diffuse, or specular flow* are

1. There is a relationship to the photometric stereo problem, involving shape recovery based on employing images obtained under illumination from (three) sources in different positions (e.g., [6], [33]). However, photometric stereo requires reasonably large separations between source positions (equivalent to large baselines in binocular stereo), while δe information corresponds to a differential source motion. For three nearby source directions, resulting image shadings are quite similar, and reflectance maps in the gradient space are nearly tangent in the neighborhood of the solution. As a result, photometric stereo formulation becomes highly ill-conditioned. However, Wolff [31] has shown that the image from one view and the photometric flows from two differential rotations of the source (i.e., two δe fields) can provide orthogonal constraints in the gradient space, and thus may be used to recover the surface shape information robustly. The relationship between the two problems strongly resembles that of motion and stereo.

• *The author is with the Electrical and Computer Engineering Department, University of Miami, Coral Gables, FL 33124.*
E-mail: shahriar@ece.miami.edu.

Manuscript received 27 Feb. 1997; revised 18 May 1998. Recommended for acceptance by K. Ikeuchi.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107075.

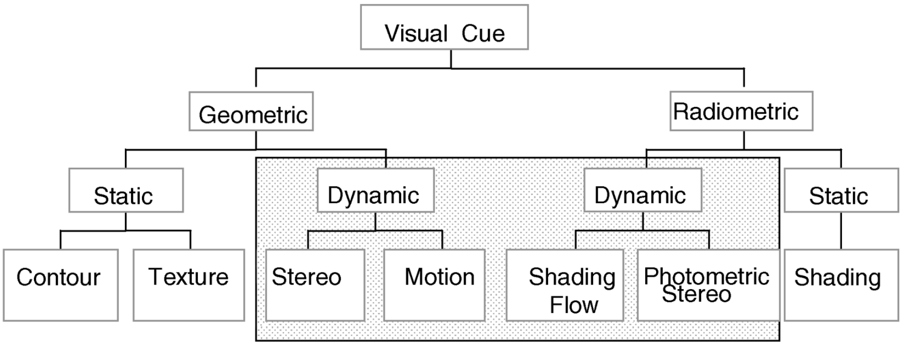


Fig. 1. Shape-from-X methods.

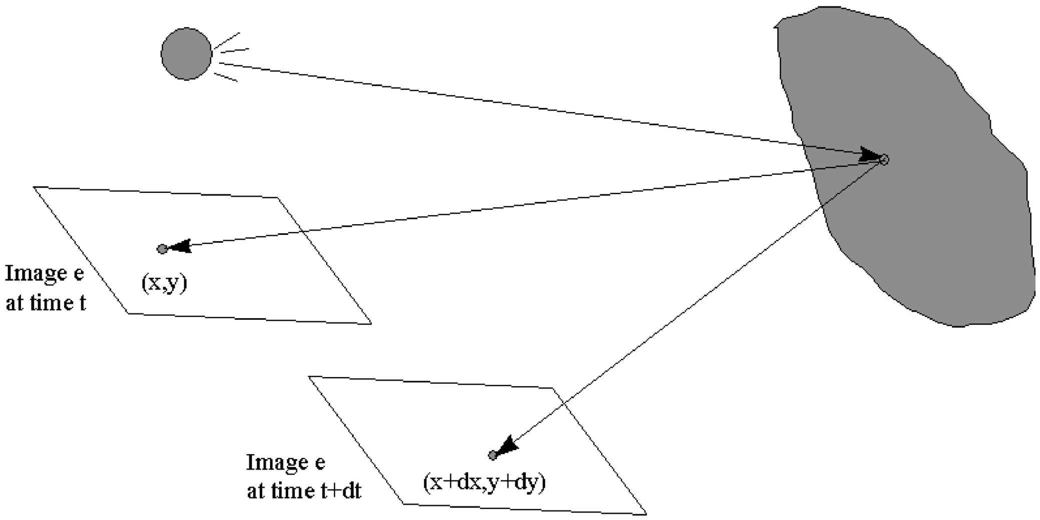


Fig. 2. Time-varying imagery encodes information about both geometric and radiometric variations of image brightness patterns. These, respectively, are the displacement $[\delta x, \delta y]^T$ and intensity variation $\delta e = e(x + \delta x, y + \delta y, t + \delta t) - e(x, y, t)$.

often used to distinguish the radiometric transformation induced by light source motion from the well-established interpretation as a geometric transformation induced by object or camera motion.

The primary objective in this paper is to establish a revised definition of optical flow, first proposed in [22]. Optical flow is viewed not simply as a geometric entity, but one that also encodes radiometric cues. It is argued that the new representation, describing both the radiometric and geometric variations in an image sequence, is more consistent with the common interpretation of optical flow. Furthermore, it is a richer entity, related to 3D scene properties, than the image motion alone. Knowledge of optical flow can permit the recovery of various types of information about, and certain dynamical events in, the scene (e.g., albedo, shape, and reflectance properties of surfaces, camera, and (or) light source position and motion). The revised definition calls for a more general interpretation of the aperture problem, as well as a number of problems in dynamic scene analysis. We also investigate the so-called *generalized dynamic image model* for the study of problems in dynamic scene analysis and show its application for the computation of (the radiometric and geometric components of) optical flow [21]. Comparisons with other image motion models provide insight on the physical interpretation of the radiometric transformation field.

The generalized dynamic image model and the new definition of optical flow provide a natural framework for the integration and unified treatment of radiometric and geometric visual cues in time-varying imagery. The significance of the new interpretation and accurate computation of optical flow in 3D scene analysis and reconstruction from time-varying imagery is demonstrated through selected examples. It needs to be emphasized that the issues addressed in this paper are tied to a large body of the previous work on shape-from-X problems. Citations made here are merely selected examples and are by no means an exhaustive listing.[2]

The remainder of the paper is organized as follows. In Section 2, we first review the classical definition of optical flow with selected examples that clarify the distinction from image motion. This leads to the revised definition and the reinterpretation of the aperture problem. In Section 3, we present the generalized dynamic image model (GDIM) and

2. In particular, a number of citations have been made to articles that have been collected in *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, which was edited by L.W. Wolff, S.A. Shafer, and G.E. Healey (Boston, Mass.: Jones & Bartlett Publishers, 1992), but were originally published in various other journals.

its application for the computation of optical flow. Comparison with selected other image motion models is made to establish the physical interpretation of the optical flow transformation fields. In Section 4, a generalized definition is given for the problem of 3D scene (shape) recovery from optical flow, and existing methods are classified according to the type(s) of geometric and/or radiometric cues that they employ. Selected experiments are given in Section 5 to provide examples of radiometric and geometric cues in time-varying imagery and the decoupling of this information using the GDIM-based optical flow algorithm. We summarize the contributions of this paper in Section 6.

2 Optical Flow

Optical flow has been commonly described as the *apparent motion of brightness patterns* in an image sequence [12]. Computation of optical flow is known as the problem of estimating, based on an assumed image motion constraint equation, *a displacement field* $[\delta x, \delta y]$ for transforming the brightness patterns from one image to the next. As such, the choice of an appropriate constraint equation, which represents accurately (enough) the significant events in a dynamic scene, is rather critical; this becomes clear in subsequent discussions.

Optical flow can be induced by the relative motion between the camera and a scene. However, it may also exist, in the absence of camera and/or scene motion, due to other events (e.g., a moving light source). "Extreme" examples, as the following two from [12], are often used to distinguish between optical flow and the *true geometric transformation field* $[\delta x, \delta y]$, the so-called *image motion*.[3] For instance, a uniform sphere rotating about an axis through its center is perceived to be stationary since there is no apparent image variation from one time instant to the next; that is, $e_t = \partial e/\partial t = 0$. Thus, it is commonly stated that there is no optical flow though there is an image motion, as each image point p that is the projection of a sphere point P moves with the surface. Conversely, we say that a fixed viewer perceives an optical flow when observing a stationary scene that is illuminated by a moving light source. Though there is no image motion (relative motion between the camera and the scene is zero), there is a nonzero optical flow because of the *apparent motion* of image brightness patterns induced by the moving source.

These examples have been used to clarify that image motion is a unique geometric transformation that is induced by the relative motion between the viewer and the scene. That optical flow field may be subdued or induced in the presence of various scene events (in the presence and absence of image motion). We argue that these examples signify the shortcomings in optical flow interpretation as merely an image displacement field. In this framework, optical flow can be informative if and only if it is

1) due to the relative motion between the scene and camera and
2) calculated with reasonable accuracy.

When induced by other scene events, e.g., a moving source, we have problems with both calculation and interpretation of a meaningful optical flow: We are forced to arrive at a geometric interpretation, when the actual underlying image transformation is a radiometric one. Consequently, there is no unique interpretation since the problem does not have a (physically meaningful) solution. This is a serious inadequacy since the induced optical flow in these cases still encodes valuable information about the three-dimensional scene being viewed; e.g., shape cues from moving shadows (e.g., [14]) and from diffuse or specular flows (e.g., [31], [35]) induced by a moving source. We claim that we need a definition (or representation) of optical flow that permits accurate interpretation under most, if not all circumstances: an interpretation that is consistent with the underlying scene properties, events, and dynamics.

2.1 Revised Definition

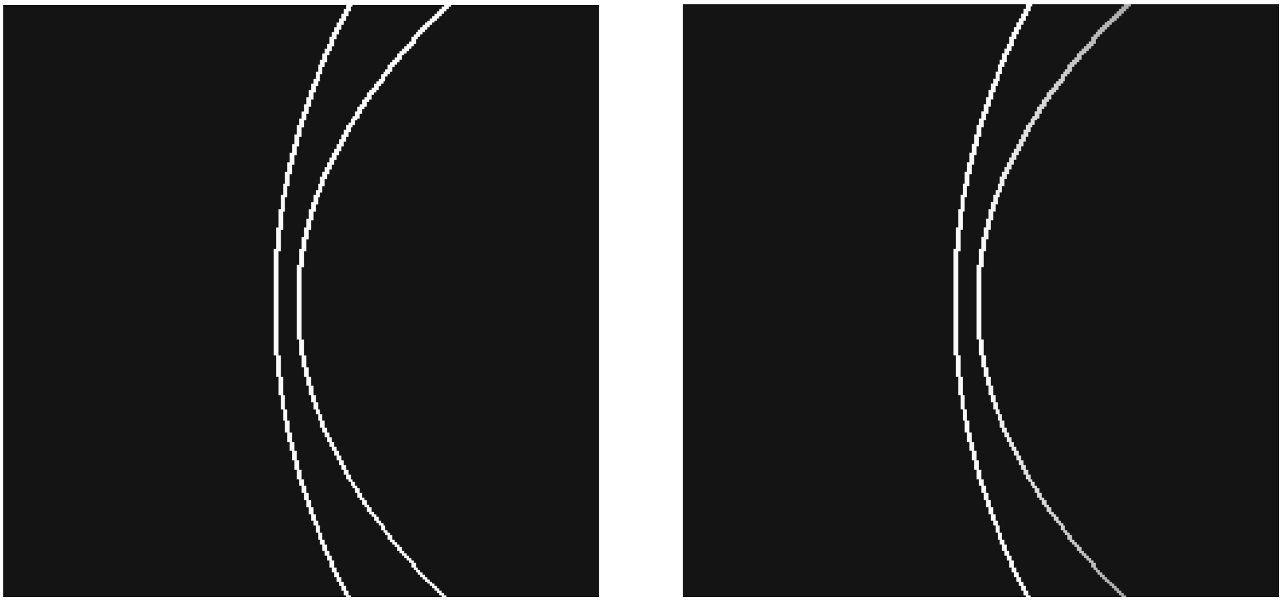
In [22], a revised definition of optical flow is proposed as a three-dimensional transformation field $v = [\delta x, \delta y, \delta e]^T$. In this representation, optical flow describes the *perceived transformation*, instead of the perceived motion, *of brightness patterns in an image sequence*. Alternatively, there exists an optical flow in the presence of an apparent or perceived transformation of image brightness patterns. The image motion is only the geometric component of optical flow. The last component, which we refer to as the *scene brightness variation field*, describes the radiometric transformation in an image sequence. We can appreciate the importance of the new definition if we take note of certain facts:

- Relative motion between the camera and the scene is only one special scene event that induces optical flow, namely, in the form of a geometric transformation. It therefore makes perfect sense that image motion is only the geometric component of optical flow.
- Optical flow from other scene events (e.g., a light source motion) induces a radiometric image transformation, uniquely and correctly described by the radiometric component δe. As stated, classical representation of optical flow forces an incorrect interpretation as a geometric displacement field.
- Optical flow as a three-dimensional field permits a *complete representation of all the information in an image sequence*. This permits perfect transformation of image $e(\boldsymbol{r})$ to image $e(\boldsymbol{r} + \delta\boldsymbol{r})$.

The new definition permits correct interpretations of optical flow in the two earlier examples. The true optical flow in the "degenerate" case[4] of a rotating uniform sphere includes an image motion component $[\delta x, \delta y]^T$ (due to rotation) and a corresponding scene brightness variation field δe (due to the change in the surface radiance as the object rotates relative to the source), in such a

3. The use of same notation to describe optical flow and image motion is somewhat misleading—probably as misleading as the classical interpretation of optical flow, as we explain.

4. By degenerate case, we mean one with multiple/infinite interpretations.



(a) (b)

Fig. 3. One interpretation of the generalized aperture problem in terms of corresponding contours. See text for details.

way that the temporal brightness variation e_t vanishes at every image point.[5] While the classical representation of optical flow permits only the zero image flow interpretation, the revised one allows both the zero and true flows as possible interpretations. Arriving at the correct interpretation, however, requires imposing appropriate constraints. Without these, any optical flow algorithm is expected to arrive at the "natural" interpretation, that of a stationary sphere, for an image sequence with no temporal variations. In the case of a stationary scene viewed by a fixed camera and illuminated by a moving source, the true interpretation of optical flow involves a zero motion field with a nonzero scene brightness variation field.

With the new definition, a natural question involves that of modeling the optical flow and its accurate computation from time-varying imagery. In many applications, the scene dynamical events induce both geometric and radiometric variations in time-varying image. In realistic situations, it may be impractical to ensure that one component of optical flow, either radiometric or geometric, is dominant over the entire image; this typically requires high control over lighting, camera motion, target position, and orientation, etc. It is highly desirable to accurately decouple image transformations into the geometric component $[\delta x, \delta y]^T$ and the radiometric component δe. Based on this information, Shape-from-X methods may be developed that exploit both cues in recovering information about the scene. Before we address these issues, we need to reinterpret the aperture problem.

2.2 Generalized Aperture Problem

The primary difficulty in the estimation of image motion from time-varying imagery is due to the *aperture problem*. This can be described using a particular example from [12]. Let C be an isobrightness contour in an image at time t, and C' is the corresponding contour in the image at some later time $t + \delta t$ (Fig. 3). We first assume that the contour maintains the same brightness value (Fig. 3a). Based on the *brightness constancy assumption*, the problem of estimating the motion of a point P on C translates to finding the correct match on C'. Though we have restricted the search to a point on C', we cannot uniquely determine the correct match P', as the contour generally changes its shape from one image to the next. We need other information and/or constraints (e.g., at corner points of a contour). This is one interpretation of the ambiguity known as the aperture problem.

Now, we allow the brightness of points in the second image to differ from their corresponding matches in the first image (Fig. 3b); that is, we assume a nonzero δe induced by some dynamic scene event (e.g., due to light source motion, variations in surface reflectance, etc.). The ambiguity now is even more severe; we cannot even match C and C' since the two contours have different brightness patterns. In fact, C' is not restricted to remain an isobrightness contour since points on C can change their brightness by different amounts. Therefore, we cannot restrict the search for matches of points on contour C to isobrightness contours (of intensity equal to or different from that of C). This is the *generalized aperture problem*.[6]

5. It is important to note the distinction between the rate of brightness change at a fixed point in the image $e_t = \partial e/\partial t$, and $\delta e \approx (de/dt)\delta t$ that is associated with tracking the image of a surface point.

6. It should be noted that the generalized aperture problem exists, independent of optical flow representation, once we permit a nonzero δe field.

3 Modeling of Dynamic Imagery

To extract information from time-varying imagery, we need a model that describes the variations (or transformations) in image brightness patterns. Such a model expresses the image at time $t + \delta t$ in terms of the previous image (or possibly images at times $t - k\delta t$, $k = 0, 1, \ldots$). Alternatively, the model can be written in terms of the scene brightness variation field $\delta e(\boldsymbol{r})$. Deriving the appropriate constraint equation involves taking into account various scene properties (e.g., shape, range, and reflectance of the scene surfaces, source position and orientation, etc.), as well as significant scene events that induce the optical flow, including the motion of camera, objects in the scene and/or the illumination source. We next review selected models for the computation of optical flow and investigate the interpretation of the scene brightness variation field δe.

3.1 Brightness Constancy Model

The simplest and most popular dynamic image model used in the motion vision literature is based on the so-called *brightness constancy assumption* (BCM):

$$e(\boldsymbol{r} + \delta\boldsymbol{r}) = e(\boldsymbol{r}). \tag{1a}$$

This model, and the corresponding image motion constraint derived from the first-order Taylor series expansion of the left-hand side about $\boldsymbol{r}$,

$$e_t + e_x\delta x + e_y\delta y = 0, \tag{1b}$$

has been the basis of a large number of methods[7] (see [3] for a survey). Here, $\boldsymbol{g} = (e_x, e_y, 0)^T$ and e_t are the spatial and time derivatives of the image. (Methods based on higher-order Taylor series representation of the brightness constancy model have also been proposed (e.g., [18]).)

Since the third component of optical flow, the scene brightness variation field, is explicitly assumed to be zero, the model is bound to produce erroneous results in image regions corresponding to surface patches with significant (ir)radiance variations from one image to the next (see examples and selected experiments in [21]).

Other methods for the estimation of image motion, subject to similar restrictions as the brightness constancy model, have incorporated the assumption of constancy of the image gradient [28], [29]:

$$\frac{d}{dt}(\nabla e) = 0. \tag{2}$$

In uncontrolled environments, various events can contribute to (ir)radiance variations of scene surfaces, including motion of the light source, object, or viewer; illumination nonuniformities; cast shadows; etc. In certain applications, image brightness variations are induced by active manipulation of camera settings. For instance, a sensor with computer-controlled aperture or exposure time may be employed to optimize image quality, to maintain a steady average brightness level, to prevent camera saturation, or to increase signal-to-noise ratio when illumination drops. Image models that take into account variations in scene (ir)radiance are necessary for applications of machine vision systems in uncontrolled environments.

3.2 Generalized Dynamic Image Model

In earlier work [20], [21], we have proposed the GDIM for representing transformations in dynamic imagery:

$$e(\boldsymbol{r} + \delta\boldsymbol{r}) = M(\boldsymbol{r})e(\boldsymbol{r}) + C(\boldsymbol{r}). \tag{3a}$$

The radiometric transformation from $e(\boldsymbol{r})$ to $(\boldsymbol{r} + \delta\boldsymbol{r})$ is explicitly defined in terms of the multiplier and offset fields $M(\boldsymbol{r})$ and $C(\boldsymbol{r})$, respectively. The geometric transformation is implicit in terms of the correspondence between points $\boldsymbol{r}$ and $\boldsymbol{r} + \delta\boldsymbol{r}$. If we write M and C in terms of variations from one and zero, respectively (the latter is simply for consistency in notation for both components),

$$M(\boldsymbol{r}) = 1 + \delta m(\boldsymbol{r}) \text{ and } C(\boldsymbol{r}) = 0 + \delta c(\boldsymbol{r}),$$

we can express the GDIM explicitly in terms of the scene brightness variation field:

$$e(\boldsymbol{r} + \delta\boldsymbol{r}) - e(\boldsymbol{r}) = \delta e(\boldsymbol{r}) = \delta m(\boldsymbol{r})e(\boldsymbol{r}) + \delta c(\boldsymbol{r}). \tag{3b}$$

This representation is also useful where comparisons may need to be made with the brightness constancy model. That is, the brightness constancy assumption holds where $\delta m = \delta c = 0$.

One should not overlook the generality of (3), as the simple form of this transformation is superficially deceiving. Mathematically, one of the two fields, say M, is sufficient to describe the radiometric transformations in an image sequence if this is allowed to vary arbitrarily from point to point and one time instant to the next. In this case, the multiplier field M is typically a complicated function of several scene events that contribute to the radiometric transformation, each of which may vary sharply (or become discontinuous) in different isolated regions. This is not desirable since it then becomes very difficult, if not impossible, to compute optical flow due to the generalized aperture problem.[8] The representation of radiometric transformations in terms of multiplier and offset fields facilitates the computation of optical flow, as we describe next.

3.2.1 Smoothness Constraint

Application of some form of smoothness constraint has been employed to overcome the aperture problem in devising techniques for computing a unique image motion[9] (e.g., [11], [13]). This is motivated by the observation that most visual motion is the result of objects of finite size undergoing continuous and/or smooth rigid motions or deformations. Neighboring object points have similar displacements/velocities, and to the extent that they project onto neighboring image points, neighboring image points will also have similar motions. Therefore, the motion field should be smooth almost everywhere. Discontinuity in the

7. Pioneering work is by Barnard and Thompson [2] for the computation of object motion and Horn and Schunck [11] and Hildreth [10] for the computation of optical flow.

8. As an analogy, imagine the computation of image motion in the presence of a large number of depth or motion discontinuity boundaries.

9. Alternative strategies have been proposed, including use of multiple light sources [32].

image motion is due to depth and motion discontinuity boundaries, sparse regions in the image in most cases. To employ the smoothness constraint for the computation of optical flow fields δx, δy, and δe, we need to ensure that these are smooth over most of the image.

Based on the physics of imaging, many scene events induce a radiometric transformation that can be modeled by a "smooth" multiplier field M, with sparse (isolated) discontinuity boundaries. Examples include variations in source illumination or surface reflectance. However, a number of other events, including moving shadows and highlights, may be modeled by a "smooth" offset term (with possible isolated discontinuities). A general model, with multiplier and offset terms, permits the computation of these fields by imposing the smoothness constraint. In addition, this representation is suitable in dealing with sharp variations in $\delta e = \delta m\, e + \delta c$ due to large gradients in the brightness function. By imposing the smoothness constraint on δm and δc (not on δe), we expect accurate estimates of optical flow, to the extent that the underlying scene events induce smooth multiplier and offset fields. This is the main motivation (or justification) for the presentation of the scene brightness variation field in terms of multiplier and offset fields.

Does this suggest an ambiguity is decomposing the radiometric transformation δe into the multiplier and offset terms, δm and δc, respectively? Alternatively, if each is sufficient to represent the radiometric component of optical flow, what scene events are embedded in one term or the other? An ambiguity arises when, in addition to δm and δc, e varies smoothly as well (not enough texture as we show in the next section). Consequently, there can be several decompositions of δe into smooth multiplier and offset fields (depending on how smoothness is imposed). Though the decomposition may not be unique, any interpretation of the two transformation fields is sufficient to reconstruct δe. However, it may be meaningless to assign a physical interpretation to these fields in such cases. We discuss related issues more formally in Section 3.3 using selected examples of image motion models.

3.2.2 GDIM-Based Optical Flow Computation

We may represent the GDIM explicitly in terms of the various components of the flow. A linear approximation can be derived from the Taylor series expansion of the left side of (3) about $\mathbf{r} = [x, y, t]^T$ up to first-order terms:

$$e_t + e_x \delta x + e_y \delta y - (e\,\delta m + \delta c) = 0. \tag{4}$$

For each scene point, (4) constrains the instantaneous image displacement $[\delta x, \delta y]^T$ and radiometric transformation fields δm and δc in terms of the spatiotemporal gradient of the corresponding image point. If we rewrite (4),

$$e_t = \delta g + \delta e \quad \text{where} \quad \delta g = -(e_x \delta x + e_y \delta y), \tag{5}$$

we can quantify explicitly the geometric and radiometric components (δg and δe, respectively) of variations in an image sequence. The ratio $\delta g / \delta e$ can be used as a measure for relative strength of geometric and radiometric cues. For example, shape cues are stronger from motion than shading flow, where δg is large relative to δe (and vice-versa). This is quite useful in evaluating the reliability of 3D scene information extracted from each or a combination of these components. This measure can also serve as a signal-to-noise measure for traditional methods based on the brightness constancy model; that is, the image motion estimates may be deemed accurate mainly in regions with large $\delta g / \delta e$.

Application of (4) for the computation of optical flow has been discussed in earlier work. One implementation based on the regularization approach, as proposed by Horn and Schunck [11], involves imposing smoothness through minimization of cost (energy) functionals [20]. Alternatively, smoothness can be imposed with the assumption that sought after fields are constant within small regions around each point. This approach, commonly employed in solving various ill-posed low-level vision problems, exploits a least-square formulation with a simple closed-form solution (e.g., see [13] for application to optical flow computation based on BCM). In our case, the four fields (δx, δy, δm, and δc) are calculated from

$$\sum_W \begin{bmatrix} e_x^2 & e_x e_y & -e_x e & -e_x \\ e_x e_y & e_y^2 & -e_y e & -e_y \\ -e_x e & -e_y e & e^2 & e \\ -e_x & -e_y & e & 1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta m \\ \delta c \end{bmatrix} = \sum_W \begin{bmatrix} -e_x e_t \\ -e_y e_t \\ e e_t \\ e_t \end{bmatrix} \tag{6}$$

where W is a neighborhood region [21]. (Results in this paper are computed for a 9×9 window centered at each pixel.) The multiplier and offset fields can become discontinuous at isolated boundaries, just as image motion is discontinuous at occluding or motion boundaries. As a result, the estimated radiometric and geometric components of optical flow may be inaccurate in these regions. Erroneous results may be detected by evaluating the residual sum squared-error. Based on this, the estimation process can be repeated by adjusting the image region. In practice, other visual cues (e.g., intensity edges) can be employed to guide this process.

Finally, there can exist situations where the solution of (6) is not unique (ill-conditioned). These can occur in regions with weak/regular texture where the motion and shading flows interact to yield multiple interpretations. Mathematically, these happen for points or regions where the above 4×4 matrix is (nearly) singular. In Appendix 2, we show that the matrix has full rank (we obtain a unique solution) if brightness within window W varies at least quadratically and if both the determinant and trace of the Hessian of $e(x, y, t)$ are nonzero. However, investigation of other types of ambiguities (e.g., nonunique decomposition of δm and δc) may be of interest to psychophysical studies and certain machine vision applications; these studies are yet to be carried out.

We next review a number of constraint equations that, in agreement with GDIM, permit radiometric variations in an image sequence. We show through comparison with these models the physical interpretations of the multiplier and offset fields for certain cases.

3.3 Selected Dynamic Image Models

Cornelius and Kanade [7] have proposed employing the equation

$$e_t + e_x\delta x + e_y\delta y - \delta e = 0 \tag{7}$$

for the estimation of image motion. The unknown δe, in the form of an offset field, was constrained to vary smoothly over the image. As discussed earlier, this formulation can yield erroneous estimates in regions with sharp gradients $\boldsymbol{g}$.

Using Phong's shading model, Mukawa [17] has derived constraints on δe in (7) in terms of light source direction and surface albedo. Using this physical interpretation, the algorithm of Cornelius and Kanade was revised to incorporate additional constraints in the computation of optical flow. It is also shown how source direction, surface shape, and albedo are estimated from the information in various components of optical flow.

Nagel [19] derives analytical expressions for image irradiance variations due to perspective projection effects for a moving Lambertian surface patch under isotropic illumination. In his result, the scene brightness variation field is in the form of a multiplier field,

$$\delta m = 4\left(\frac{t_z}{Z} - \frac{\boldsymbol{R}' \cdot \boldsymbol{R}}{\boldsymbol{R} \cdot \boldsymbol{R}}\right), \tag{8}$$

where $\boldsymbol{R}'$ is the velocity of scene point $\boldsymbol{R} = (X, Y, Z)$; $t = (t_x, t_y, t_z)$ is translational component. Thus, radiometric transformation field encodes information about 3D shape and motion of the scene (objects).

Schunck [25] proposed the image motion constraint

$$e_t + e_x\delta x + e_y\delta y + e(\nabla \cdot v) = 0, \tag{9}$$

to overcome the limitations of the BCM ($\nabla \cdot v$ is the divergence of image motion). Based on this model, the scene brightness variation term takes on the multiplier form with $\delta m = -(\nabla \cdot v)$. Nagel [19] questions the underlying assumptions in the derivation of (9), and no experimental results have yet been given to support its validity (to our knowledge). We have examined this model through comparison with Nagel's result. We show in Appendix 1 that the results of Nagel and Schunck are mathematically equivalent under the following conditions:

- perspective projection,
- pure translational motion,
- coefficient four set equal to one in (8),
- scene surface expressed in image coordinates is given by

$$Z(x, y) = K(1 + x^2 + y^2)^{1/2} \quad \text{for} \quad K > 0.$$

In practice, such a surface can be presented by a frontal plane over a reasonably large field of view without significant error (e.g., maximum error in depth, at image boundary, is 8 percent for a half-angle field of view as large as $\tan^{-1}(0.4) = 22$ degrees).

Verri and Poggio [30] have derived related results, expressed in terms of the normal component of image motion (component along local gradient $\boldsymbol{g}$), for a surface with Lambertian and specular reflectance undergoing motion with translation and rotation components $\boldsymbol{t}$ and ω, respectively. The magnitude of scene brightness variation field is calculated from

$$|\delta e| = \left|\rho\, \hat{n} \cdot \left(\hat{\imath} \times \omega\right) + \frac{sk}{\|\boldsymbol{R}\|^3}\left(\frac{\boldsymbol{R} \cdot \hat{s}}{\|\boldsymbol{R}\|}\right)^{k-1}\{.\}\right|, \tag{10}$$

$$\{.\} = (t \times \boldsymbol{R}) \cdot (\hat{s} \times \boldsymbol{R}) - 2\|\boldsymbol{R}\|^2$$
$$\Big((\boldsymbol{R} \cdot \hat{n})\big(\hat{\imath} \cdot (\omega \times \hat{n})\big) + \big(\hat{\imath} \cdot \hat{n}\big)\big(\boldsymbol{R} \cdot (\omega \times \hat{n})\big)\Big)$$

Here, ρ is surface albedo, $\hat{\imath}$ denotes the source direction, $\boldsymbol{R}$ and $\hat{n}$ are position vector and unit normal of a surface point, s gives the fraction of specularly reflected light according to

$$\text{Specular Component} = s\left(\frac{\boldsymbol{R} \cdot \hat{s}}{\|\boldsymbol{R}\|}\right)^k,$$

and $\hat{s} = \hat{\imath} - 2\left(\hat{\imath} \cdot \hat{n}\right)\hat{n}$ is unit vector for the direction of perfect specular reflection. We note that the scene brightness variation field δe encodes information about the 3D surface shape and motion, light source direction, as well as reflectance properties of the surface.

Using this example, we examine how δe is decomposed into smooth multiplier and offset components, in agreement with assumptions of the GDIM-based algorithm. Alternatively, given the estimated δm and δc fields from (6), what is the physical interpretation in terms of scene properties? First, we consider the highlight regions where δe is dominated by specular flow; that is,

$$|\delta e| \approx \left|\frac{sk}{\|\boldsymbol{R}\|^3}\left(\frac{\boldsymbol{R} \cdot \hat{s}}{\|\boldsymbol{R}\|}\right)^{k-1}\{.\}\right|, \tag{11}$$

where {.} (which depend on $\boldsymbol{R}$ and n) is given above. The only variables, $\boldsymbol{R}$ and n, vary slowly within small highlight regions. We conclude that δe is a smooth offset term (with $\delta m = 0$) within highlight regions. Next, consider regions where diffuse shading flow is dominant. In this case, we obtain

$$|\delta e| \approx \rho\, \hat{n} \cdot \left(\hat{\imath} \times \omega\right). \tag{12a}$$

Noting that the diffuse shading is given by $e = \rho\left(\hat{n} \cdot \hat{\imath}\right)$, we can write

$$|\delta e| = \left|\frac{\hat{n} \cdot \left(\hat{\imath} \times \omega\right)}{\left(\hat{n} \cdot \hat{\imath}\right)}\right| e. \tag{12b}$$

In small regions on an object, surface normal $\hat{n}$ typically varies gradually, and thus the coefficient on the right-hand side in (12b) varies smoothly. Consequently, a smooth multiplier field (with $\delta c = 0$) can represent diffuse shading flow. Both δm and δc are necessary in regions where diffuse and specular shading flows are comparable.

In summary, we have presented the GDIM where the radiometric transformation is decomposed into multiplier and offset fields. A GDIM-based optical flow algorithm has been devised by imposing the smoothness constraint

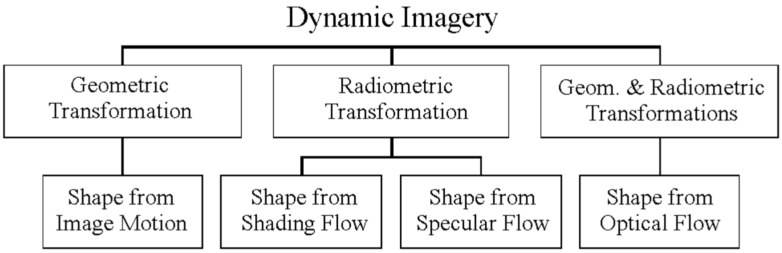


Fig. 4. Classification of methods for 3D scene (shape) recovery from time-varying imagery.

(constancy of geometric and radiometric transformation fields within small image regions). We have reviewed a number of optical flow constraints, where the radiometric transformation field is expressed as a multiplier and/or an offset field(s), in agreement with the GDIM. Finally, physical interpretation of the radiometric component δe is given through explicit representation in terms of various scene parameters.

4 3D Scene Recovery From Optical Flow

Many vision problems for 3D scene recovery and interpretation employ image data that are acquired under passive or active variation of one or more conditions or parameters of the imaging environment. Geometric cues in time-varying imagery, namely, image motion, have been employed to extract 3D motion and shape information (e.g., [1], [5], [15], [16]). Similarly, methods have been devised by exploiting radiometric cues in dynamic imagery, e.g., shape from diffuse flow [31], specular flow [4], [35], [36], and moving shadows [14]. It is hardly the case, in dealing with images of real scenes, that only one type of visual cue, geometric or radiometric, is dominant over the entire image. Consequently, it becomes inevitable to integrate information from all components of optical flow for robust 3D scene (shape) recovery from dynamic imagery. These can provide richer information about scene properties and also may resolve possible ambiguities in scene interpretation. For instance, when the first two components of the flow are zero (no motion field), any brightness variation is solely due to radiometric effects, including motion of the light source and/or a variation in its intensity, change in camera gain or exposure time (if these are allowed), etc. In the presence of motion, both geometric and radiometric effects contribute to the scene brightness variation field, and the geometric effects may be discounted using motion field information (the first two components of optical flow) and other constraints on surface shape, if these are available (e.g., shape information encoded in the image motion).

Interpretation of radiometric information generally requires knowledge of the imaging condition (e.g., point or collimated source, moving or stationary source, etc.). In the previous section, selected examples were presented where the dependency of the radiometric component on various 3D scene properties, conditions, or parameters has been derived explicitly.

Methods have been proposed to integrate the geometric and radiometric information for extracting 3D scene properties. For example, Pentland [24] argues for the significance of shading and motion cues in 3D shape recovery and proposes a method that exploits both types of information. Mukawa [17] employs (7) to compute optical flow components and employs these to recover illumination direction and the surface shape and reflectance coefficients (albedo). Grimson [9] proposes use of Lambertian and specular shading cues to guide surface interpolation in a feature-based stereo system, in contrast to imposing the heuristic smoothness assumption based on membrane and thin-plate energy models (e.g., [8]). The problem formulation involves the integration of depth estimates at zero-crossings and shading constraints on surface normals in other regions. This method, based on the information in a stereo pair, can be generalized to the case of monocular image sequences.

Based on our definition of optical flow, we can classify various approaches to 3D scene recovery from dynamic imagery according to Fig. 4. Here, explicit distinction is made of

1) *shape from image motion*, encompassing methods that employ solely the geometric cues,
2) *shape from shading flow* by merely exploiting the radiometric cues, and
3) *shape from optical flow* approach, comprising of techniques that integrate both visual cues.

5 Selected Examples

We next investigate four experiments, involving the calculation of the three components of optical flow [δx, δy, δe] from (6). The primary goal is to show through selected examples:

- Information contents of, and significance of determining, the three components of optical flow in various situations.
- Applicability of the generalized dynamic image model for accurate calculation of optical flow, based on the ability to decouple variations due to geometric and radiometric components in image sequences.

Some examples include the estimated image motion based on the BCM to demonstrate the effect of ignoring the radiometric variations in an image sequence (see also [21] for other comparisons). In experiments with real data, the original images are 512×480 and are reduced to 64×60 by block averaging for optical flow computations.

5.1 Synthetic Textured Sphere

We use synthetic images in the first experiment to demonstrate a number of issues using perfect knowledge of ground truth. The first image is that of a textured sphere (see Fig. 5a). Using a known motion, the image motion is calculated and used to transform points from this image onto the second image. (Note that there is some error in the data due to quantization.) A third image is then constructed from the second image using constant transformation fields δm and δc. (Fig. 5b shows the histograms of the three images used in these experiments.) Three sequences using frames 1-2, 1-3, and 2-3 are considered. Note that the image motions for sequences 1-2 and 1-3 are the same, while the radiometric transformation fields are equal for sequences 1-3 and 2-3. In this experiment, the motion vector is arbitrarily set to $\boldsymbol{t} = [0.97, -0.07, 0.25]^T$, $\delta m = .1$, and $\delta c = 10$.

Figs. 5c and 5d show the estimated image motions as needle diagrams using the GDIM and BCM, respectively, for sequence 1-3 (with both motion and radiometric variations). When employing BCM, we obtain an erroneous interpretation of geometric transformation due to the presence of radiometric variations. Figs. 5e and 5f are histograms of the estimated δx and δy components of image motions from the GDIM-based algorithm for the three sequences. These depict the accuracy and consistency of the results in the presence of geometric and (or) radiometric variations; an almost zero motion field for sequence 2-3; and similar ones in the absence (sequence 1-2) and presence (sequence 1-3) of radiometric variations. We have shown histograms instead of flows since the differences in the latter are visually indistinguishable in needle diagram plots (average flow for 2-3 sequence is .001 and average error between sequences 1-2 and 1-3 is .002 pixel).

Various radiometric transformation results in Fig. 6 include histograms of δm (Fig. 6a) and δc (Fig. 6b) fields for the three sequences, difference between the true and estimated (from sequence 1-3) δe fields (Fig. 6c), and difference between the estimated δe fields for sequences 1-3 and 2-3 (Fig. 6d). To quantify the error in these estimates, the average intensities in the second and third images are $\overline{e_2} = 35.7$ and $\overline{e_3} = 49.3$, respectively. Thus, the average true δe field is about 13.6 gray levels. In comparison, the average and standard deviation of the error field (difference between the true and estimated δe fields) are –0.04 and 0.56. The same numbers calculated for the sequence 1-3 are 0.15 and 0.44, respectively.

To summarize, results based on the solution of (6) demonstrate accuracy in estimating image motion in absence of scene brightness variation (sequence 1-2), capability to estimate radiometric variations in the absence of geometric transformation (sequence 2-3), and capability to decouple geometric and radiometric transformations (sequence 1-3).

5.2 Variable Camera Aperture

This example is a representative of applications where the camera aperture can be adjusted, either manually or under computer control (e.g., to compensate for ambient lighting variations) during the data acquisition process. Though we have not shown here, similar results are obtained when employing a camera with self-adjusting gain.

Original image of a popcorn-type textured scene (classroom ceiling) was compared to that acquired after increasing the camera aperture, without camera motion. The images have mean intensities of about 85 and 123 (a ratio of $\overline{\delta e}/\overline{e} = 38/85 = 0.45$), which is a rather large variation in brightness level. (Fig. 7a is the image after increased aperture.) It is assumed that the induced image motion is negligible in order to estimate the true δe field by subtracting the two images. Fig. 7b depicts the histograms of true and estimated δe fields (mean value of about 38 and 37.5, respectively). A roughly random field, in Fig. 7c, is the erroneous interpretation of the actual radiometric variation as a geometric transformation based on BCM (average flow is about 1.5 pixels). The estimated flow vectors from GDIM also have relatively random directions, but are very small (seen as dots at the same scale as in Fig. 7c). Instead, Fig. 7d gives histograms of δx and δy to quantify the means and distributions (mean magnitude of flow vectors is about 0.04 pixel). Histograms of estimated multiplier and offset fields, depicted in Figs. 7e and 7f, have averages of about 0.35 and 5.0. Based on these values, the average brightness changes from the two components are $\delta e_m = 0.35 * 85 \approx 30$ and $\delta e_c = 5$, respectively. This is in good agreement with the exception that scene brightness variation field induced by varying the camera aperture is due to a multiplier component.

From these results, we infer the importance of taking into account the radiometric variations in image sequences and ability of the GDIM to differentiate between radiometric and geometric variations with large brightness changes from one image to the next.

5.3 Specular Reflectance

Three image sequences, taking combinations of three real images of a Folgers coffee can, have been used as in the experiment with synthetic images of a textured sphere. However, scene brightness variation is induced by moving a light source parallel to the can, before the second image is captured (no camera motion). The camera is then moved parallel to the scene, and the third image is acquired (Fig. 8a is the image before motion).

Figs. 8b, 8c, and 8d are the estimated flows for the three sequences, drawn at the same scale (image motion for sequence 1-2 should ideally be zero). The average image motion for 1-3 and 2-3 sequences is 0.8 pixel of a 64×60 image. Fig. 8e is the difference of the two flows in Fig. 8c and Fig. 8d, which ideally should be zero. The x and y components of this "error" vector have means and standard deviations (.01, 0.09) and (.02, .1), respectively. (Fig. 8f shows the histograms.) The estimated flows based on BCM, not shown here, are fairly accurate, except where δe is large (within highlight regions), as expected.

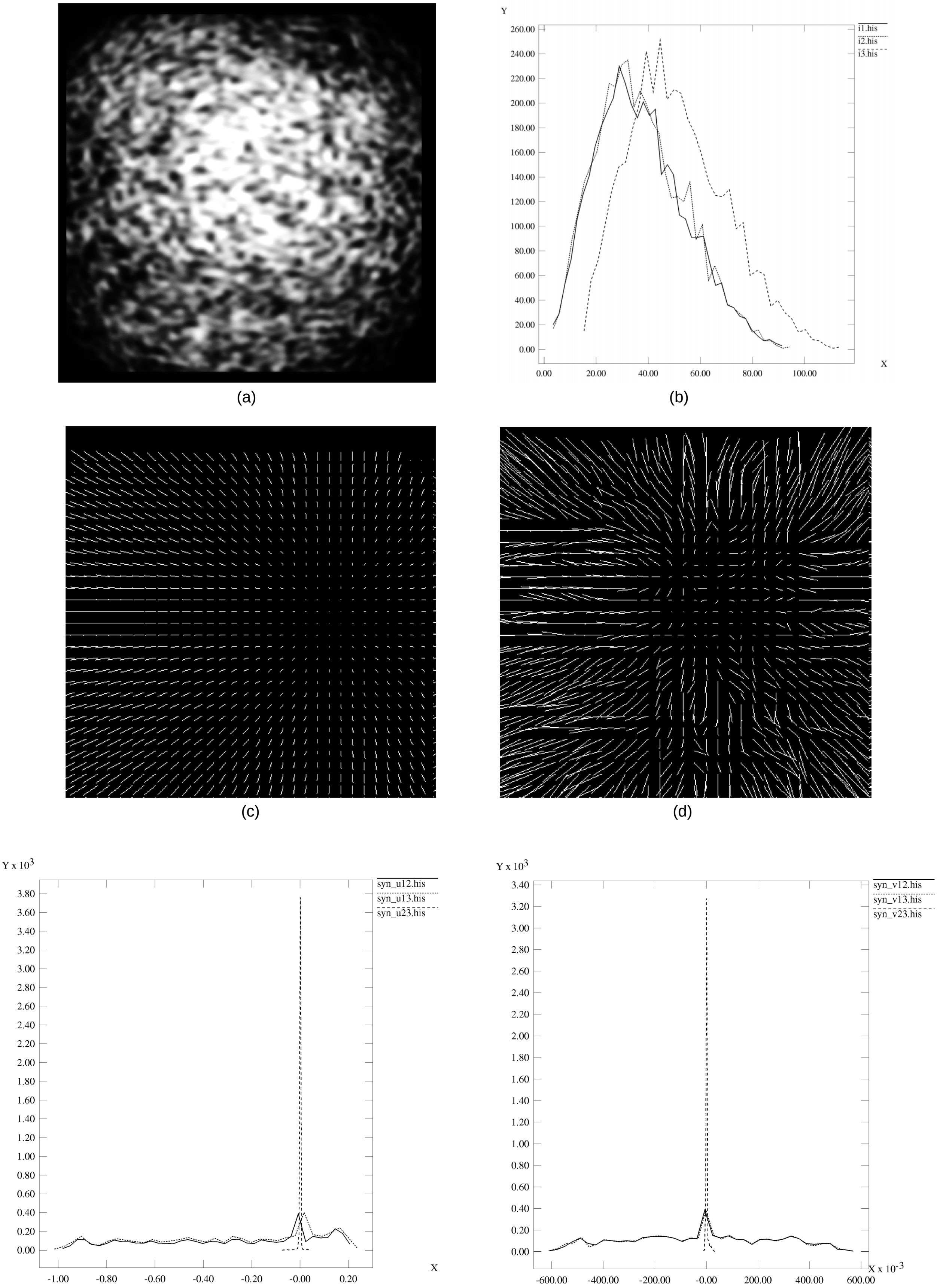


Fig. 5. Synthetic textured sphere. (a) Image 1. (b) Histograms of images in three sequences. (c), (d) Estimated flow for sequence 1-3 based on GDIM and BCM, respectively. (e), (f) Histograms of flow components using GDIM for three sequences.

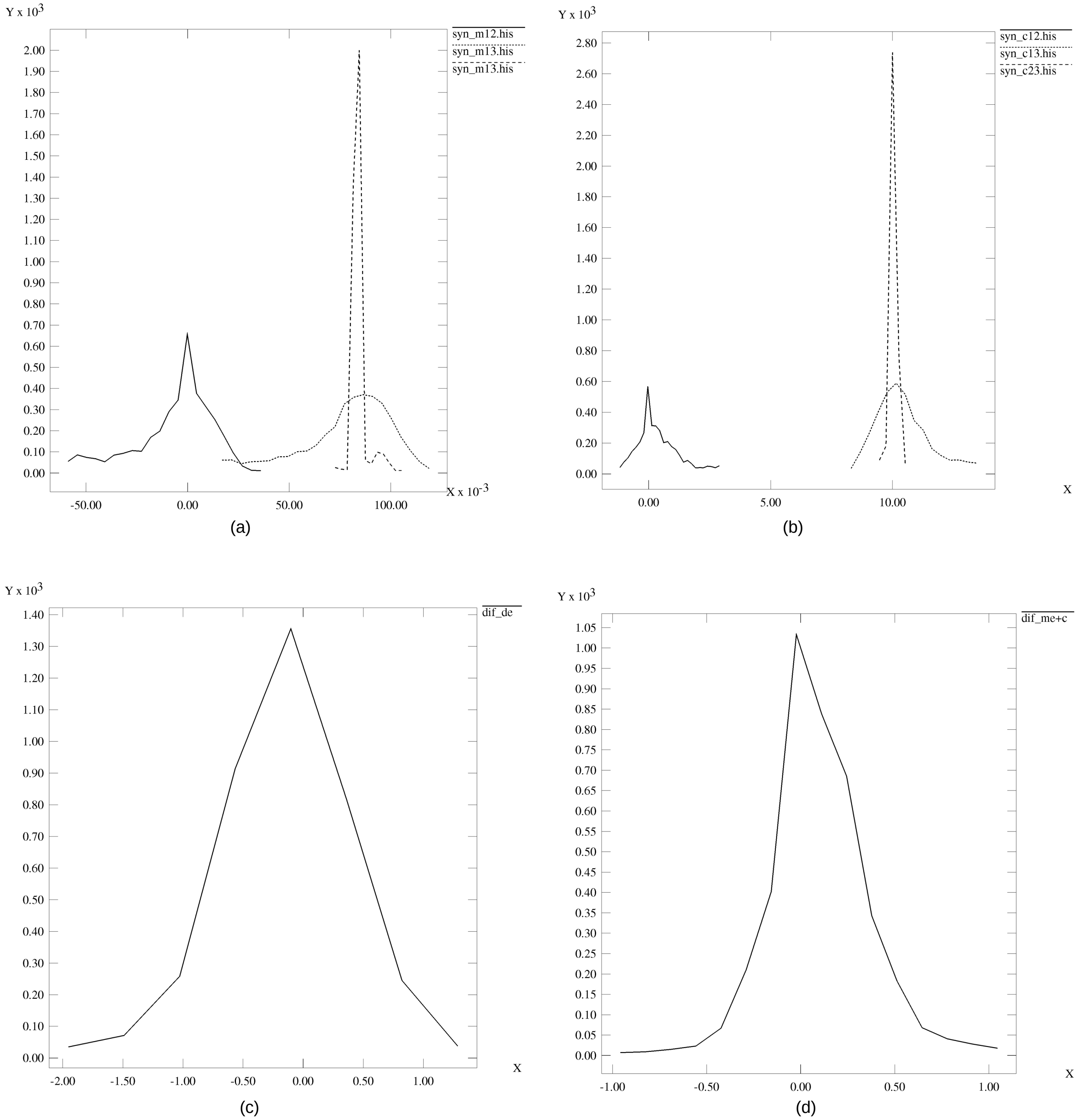


Fig. 6. Experiment with synthetic data; textured sphere (continued). (a), (b) Histograms of estimated δm and δc fields for the three sequences. (c), (d) Histograms of difference between true and estimated (sequences 2-3) δe fields, and difference between estimated δe fields for sequences 1-3 and 2-3 (both ideally should show zero fields).

Fig. 9a, the difference image after light source motion, is the true δe field. The largest brightness variations, about 72 gray levels, occur in highlight regions. Figs. 9b and 9c are the estimated δe images for 1-2 and 1-3 sequences, which agree well with the true field (note the smoothing effect in using (6) over a 9×9 window). Fig. 9d shows the histograms of these three fields. Most error magnitudes are no more than three gray levels, small relative to absolute magnitude of 72.

The results show once again that the GDIM allows accurate decoupling of the radiometric and geometric variations in image sequences. The radiometric component is useful in the detection of, and obtaining 3D shape information in, highlight regions. Shape information in other regions can be obtained by the integration of radiometric (diffuse flow δe) and geometric (image motion) cues.

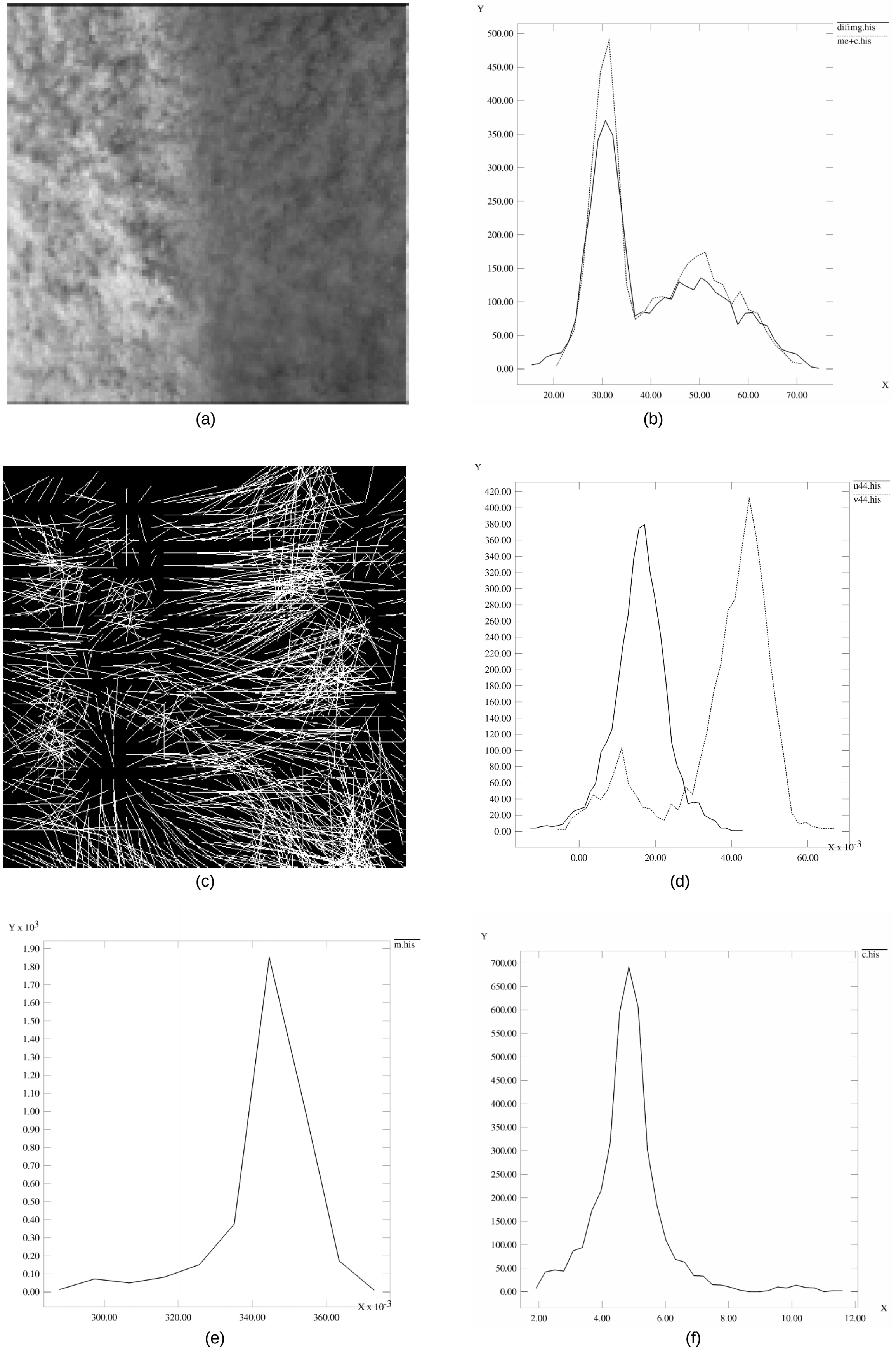


Fig. 7. Variable aperture. (a) Image with increased aperture. (b) Histograms of true and estimated δe fields. (c) Estimated image motion based on BCM. (d) Histogram of flow components based on GDIM. (e), (f) Histograms of δm and δc fields.

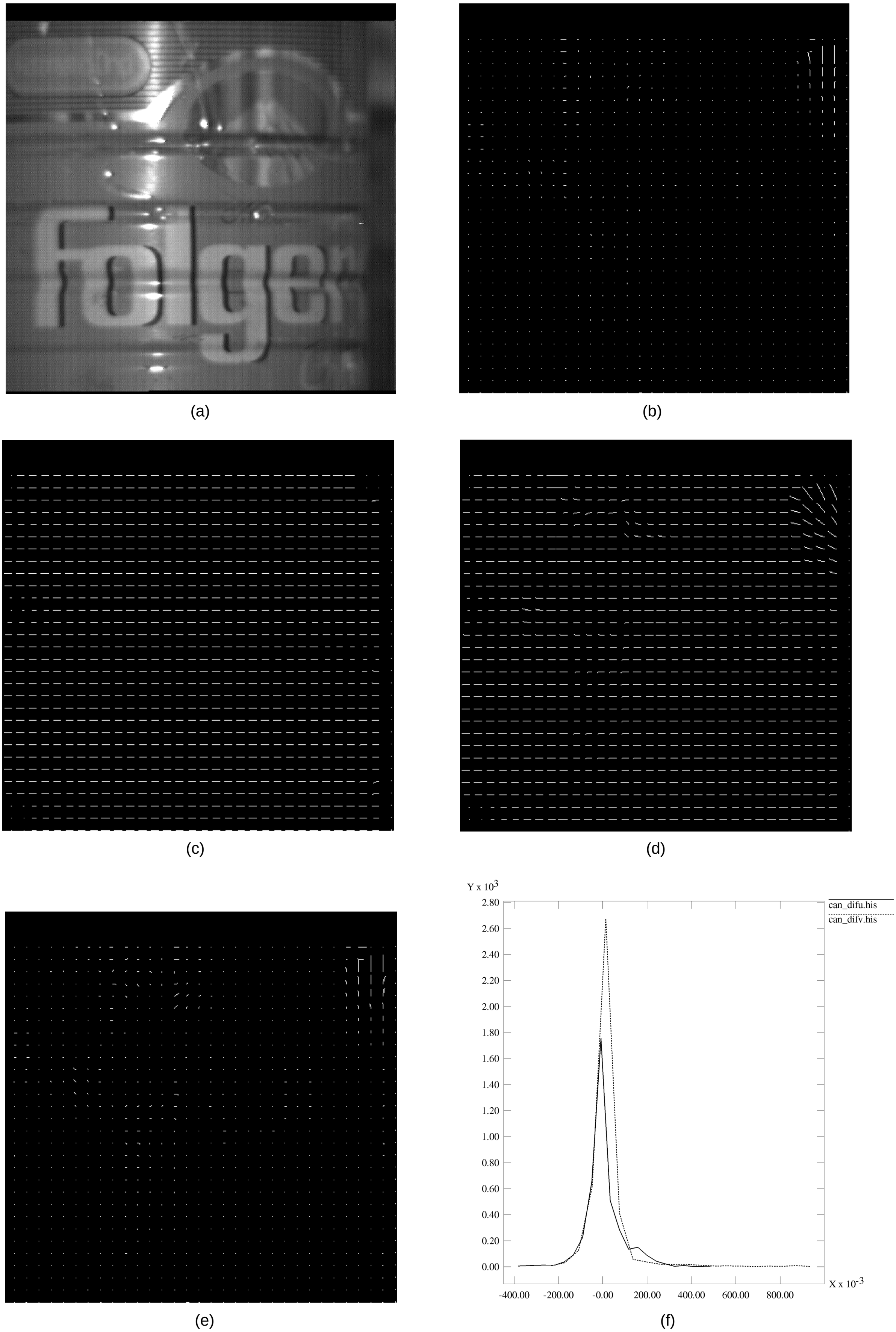


Fig. 8. Coffee can. (a) First image. (b), (c), and (d) Estimated flow from GDIM-based algorithm for sequences 1-2, 2-3, and 1-3. (e) Difference in estimated image motions for sequences 1-3 and 2-3. (f) Histograms of components of the error flow in (e).

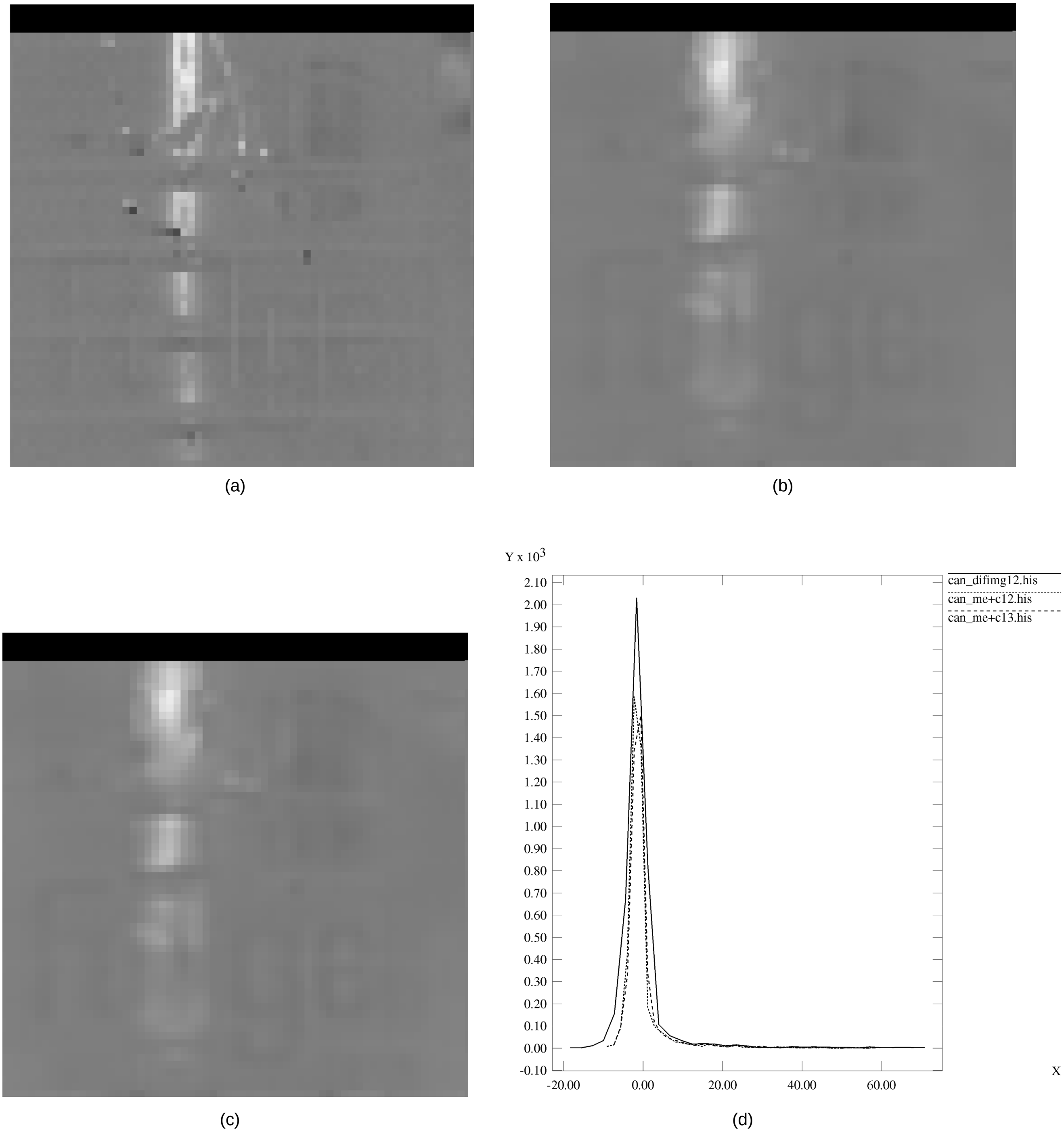
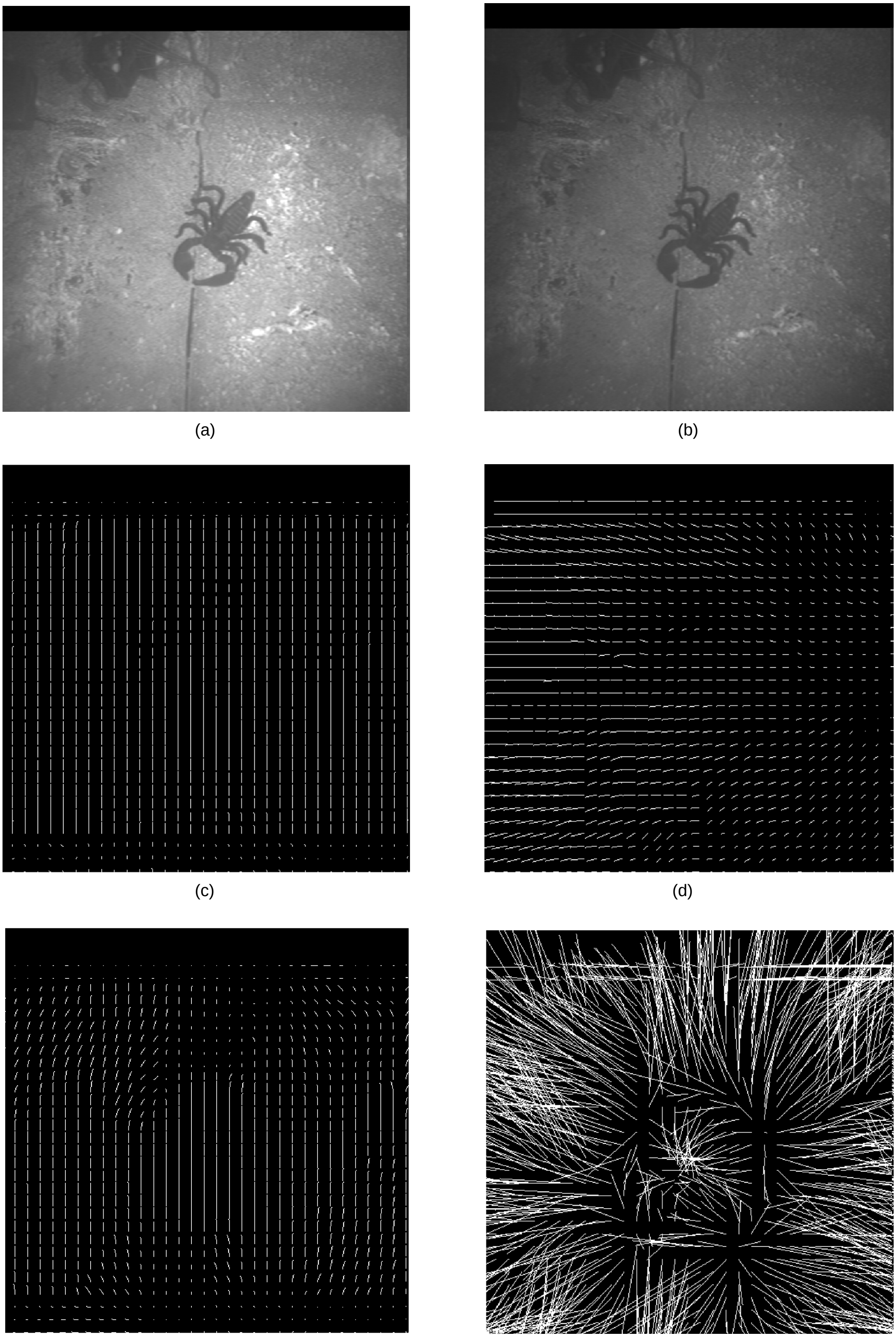


Fig. 9. Coffee can experiment (continued). (a) True δe. (b), (c) Estimated δe fields with brightness variations induced by light source motion (sequence 1-2), and light source and camera motions (sequence 1-3), respectively. (d) Histograms of three δe fields in (a), (b), and (c).

5.4 Underwater Scene

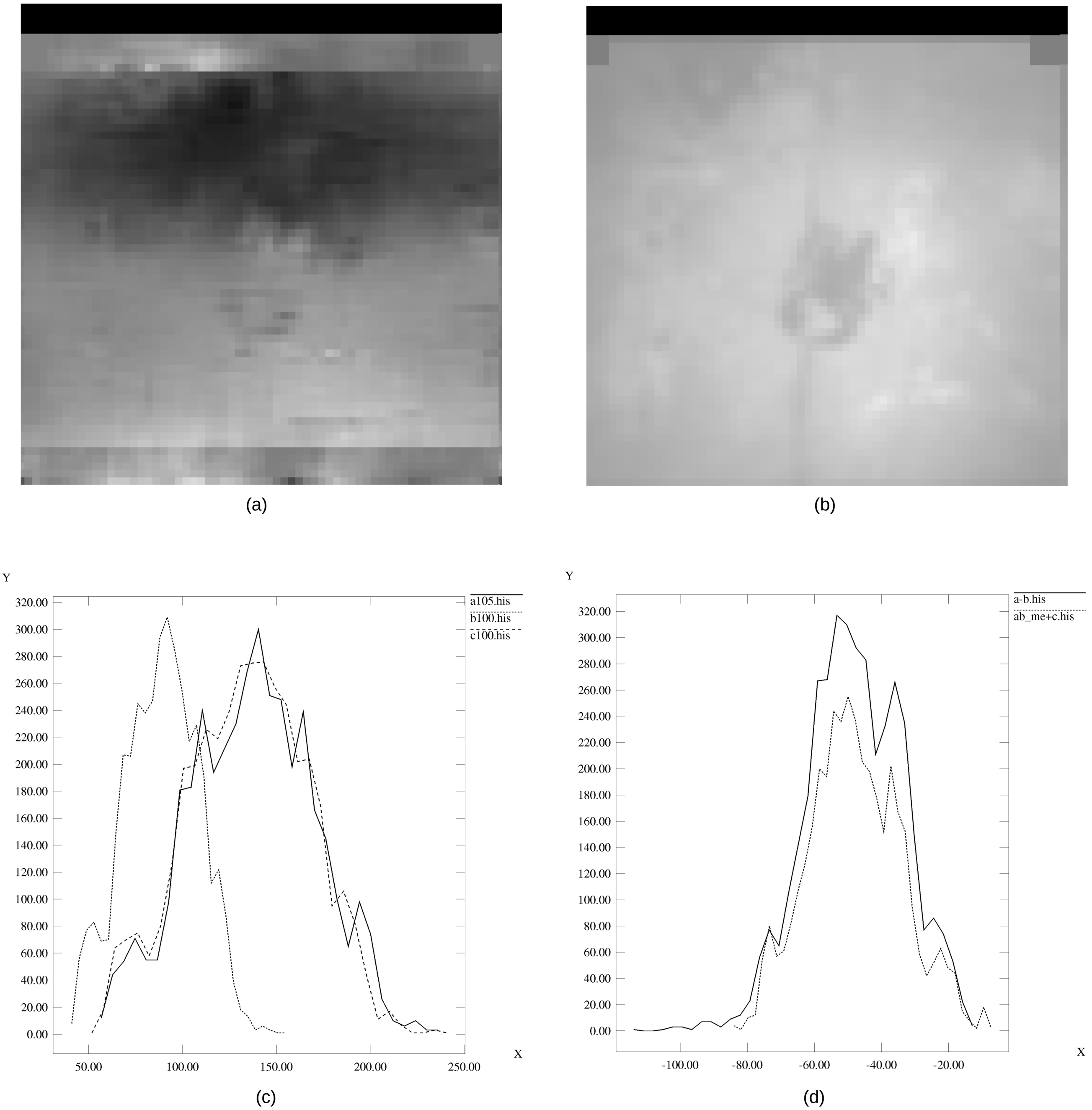
Fig. 10 shows the results of experiments involving two different motions of a "down-look" camera-light source platform. The imaging condition in these experiments is common in many applications of underwater robotic vehicles near the seabed in deep waters.

In the first experiment, the platform moves parallel to a scene, which is at the bottom of a water tank at a distance of four feet from the camera (north to south motion direction). The camera moves down and to the right in the next experiment. Corresponding images at the near and far positions relative to the scene are shown in Figs. 10a and 10b. The image motions in these two experiments, depicted in Figs. 10c and 10d and in Figs. 10e and 10f, have been estimated based on the GDIM and BCM, respectively.



(a) (b) (c) (d) (e) (f)

Fig. 10. Underwater scene. (a), (b) Near and far images in the sequence. (c), (d) Estimated flows for sequences 1-2 and 1-3 based on GDIM. (e), (f) Estimated flows based on BCM.



(a) (b) (c) (d)

Fig. 11. Underwater scene (continued). (a), (b) Estimated δe fields for sequences 1-2 and 1-3. (c) Histograms of three images in the two sequences 1-2 and 1-3. (d) Histograms of estimated δe field and the difference image in 1-3 sequence.

From Fig. 11c, showing the histograms of the three images in these experiments, we observe that there is not a significant change in brightness level in experiment one (parallel motion) though local regions undergo small changes. Correspondingly, the estimated δe image in Fig. 11a has a maximum value of about five gray levels in the lower part, minimum value of –6 gray levels in the upper region, and about no change in the middle (consistent with N-S motion of the light source). Interestingly, this is consistent with the estimated flow based on the BCM; accurate estimates are obtained in the central region with no intensity change, however, results are erroneous in the upper and lower regions, despite only a small amount of brightness variation. There is a significant intensity change in the second experiment where the camera moves toward the scene. Fig. 11d is the histograms of the estimated $-\delta e$ field (estimated irradiance change of each scene point from first to second image) and the actual difference of these two images; the difference image is not the true δe field since the two images are not geometrically registered, but is given for rough comparison. Both graphs have a similar distribution with a mean of about 45 gray levels. Note in Fig. 11b that the estimated δe image has the largest variation in the area to the right of the lobster, the region in the direction of the camera/source platform motion. We have shown (ongoing work) that the platform motion can be estimated directly from the δe field (in both experiments) without knowledge of image motion (see also [34]). In general, the dependency of δe on the source (and hence camera) motion can be integrated with the information from the image motion in order to impose more physical constraints on the motion direction.

6 SUMMARY

The significance of redundant information in image analysis for 3D scene interpretation has motivated exploiting visual cues in time-varying imagery. Most of the previous work is based on employing either geometric or radiometric cues (i.e., shape from shading flow versus shape from motion flow). More robust and accurate results can be obtained by the integration and simultaneous exploitation of information from both cues in time-varying imagery, as has been proposed in a limited number of techniques (e.g., in stereo vision [9] or motion vision [17], [24]).

We have proposed a new definition of optical flow based on the apparent transformation, rather than motion, of image brightness patterns in an image sequence. The transformation is defined in terms of two geometric fields (components of image motion) and a radiometric field (rate of change of image brightness of scene points). The new definition provides a complete representation of all possible transformations in dynamic imagery and is more consistent than the traditional one with the interpretation of optical flow. We have presented the generalized dynamic image model, proposed in earlier work for the computation of image motion [20], [21] and 3D motion [23], as a suitable tool for the simultaneous estimation of geometric and radiometric transformations in dynamic imagery. The radiometric transformation field is described in terms of multiplier and offset fields, in consistency with various dynamic scene events and to permit accurate estimation by imposing the smoothness constraint. We have shown how the GDIM relates to other image motion models that represent the radiometric component of optical flow in terms of various 3D scene properties or parameters (e.g., [17], [19], [30]), in order to establish physical interpretation in various situations. Finally, we have provided selected examples to demonstrate the capability to decouple and accurately estimate the radiometric and geometric variations in image sequences, using a GDIM-based optical flow algorithm.

We conclude with the claim that the revised definition of optical flow and the generalized dynamic image model provide a natural framework for the integration and unified treatment and exploitation of the radiometric and geometric transformations in time-varying imagery. These can be used to extract various 3D scene information from both visual cues.

APPENDIX 1

Assume the following conditions:

- Camera-centered coordinate system XYZ with origin O at focal point of camera; Z-axis chosen as optical axis.
- Image plane at $Z = f = 1$ (without loss of generality), and axes of image coordinate system are parallel to XY axes.
- Scene surface given by depth map $Z(X, Y)$.
- Perspective projection model with image coordinates given by $(x, y) = (X/Z, Y/Z)$.
- Camera translational velocity $\boldsymbol{t}$ and rotational velocity ω about an axis through O.

The velocity of a scene point relative to camera is

$$\boldsymbol{R}' = -(\boldsymbol{t} + \omega \times \boldsymbol{R}).$$

Substitution into (8) of Nagel (setting coefficient four to k for now), we obtain the multiplier

$$\delta m_{\text{nagel}} = \frac{k}{Z}\left(\frac{xt_x + yt_y + \left(2 + x^2 + y^2\right)t_z}{1 + x^2 + y^2}\right).$$

This expression is independent of the rotational motion (the first term in δm depends only on translation along Z-axis, and the second term reduces to $\boldsymbol{R} \cdot \boldsymbol{R}' = (\boldsymbol{R} \cdot (t + \omega \times \boldsymbol{R})) = \boldsymbol{R} \cdot \boldsymbol{t}$.

For Schuck's results, we write the image motion derived by Longuet-Higgins [15]:

$$v = \begin{bmatrix}\delta x\\ \delta y\end{bmatrix} = \begin{bmatrix} xy & -\left(x^2+1\right) & y \\ \left(y^2+1\right) & -xy & -x\end{bmatrix}\omega + \frac{1}{z}\begin{bmatrix}-1 & 0 & x\\ 0 & -1 & y\end{bmatrix}t.$$

The multiplier, divergence of image flow, simplifies to

$$\delta m_{\text{schunck}} = 3\left(y\omega_x - x\omega_y\right) + \frac{1}{Z^2}\left(\frac{\partial Z}{\partial x}t_x + \frac{\partial Z}{\partial y}t_y + \left(2Z - \left(x\frac{\partial Z}{\partial x} + y\frac{\partial Z}{\partial y}\right)\right)t_z\right).$$

For results of Nagel and Schuck to be equivalent, we need to restrict the motion to pure translation. In this case, we have

$$\delta m_{\text{schunck}} = \frac{1}{Z^2}\left(\frac{\partial Z}{\partial x}t_x + \frac{\partial Z}{\partial y}t_y + \left(2Z - \left(x\frac{\partial Z}{\partial x} + y\frac{\partial Z}{\partial y}\right)\right)t_z\right).$$

Setting the two multipliers equal to each other, we derive one expression for each of the three components of the translational motion:

$$\frac{\partial Z}{\partial x} = \frac{kx}{\left(1 + x^2 + y^2\right)}Z, \qquad \frac{\partial Z}{\partial y} = \frac{ky}{\left(1 + x^2 + y^2\right)}Z,$$

$$x\frac{\partial Z}{\partial x} + y\frac{\partial Z}{\partial y} - 2Z = -k\frac{2 + x^2 + y^2}{1 + x^2 + y^2}Z.$$

It can be readily shown that these are simultaneously satisfied for $k = 1$. That is, Schunck's model has to be revised to include a factor of four on the divergence term. Alternatively, the coefficient of four in Nagel's results should be set to one. Furthermore, the first two conditions can be solved for the surface function:

$$Z(x, y) = K\left(1 + x^2 + y^2\right)^{1/2} \text{ for any } K > 0.$$

The scaling K represents the scale factor ambiguity inherent of perspective projection model.

APPENDIX 2

The GDIM provides one constraint on the four unknown fields, δx, δy, δm, and δc. To estimate this information using (6), the window W should include a minimum of four points. A sufficiency condition can be derived in terms of the local structure of the intensity pattern. To do this, we rewrite the solution given by (6) in terms of the data from the minimum necessary number of points. To maintain symmetry, we assume that the window W consists of the four (N, S, W, and E) neighbors of the center pixel (x_c, y_c) at which the optical flow information is computed. In this case, the solution is obtained from:

$$\begin{bmatrix} e_x^1 & e_y^1 & -e^1 & -1 \\ e_x^2 & e_y^2 & -e^2 & -1 \\ e_x^3 & e_y^3 & -e^3 & -1 \\ e_x^4 & e_y^4 & -e^4 & -1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta m \\ \delta c \end{bmatrix} = -\begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^3 \\ e_t^4 \end{bmatrix},$$

where the superscript $i = 1, \ldots, 4$, is data point index. To obtain a unique solution, the 4×4 matrix $\boldsymbol{Q}$ on the left-hand side should be nonsingular; that is, $Det\{\boldsymbol{Q}\} \neq 0$.

Assume that the brightness function varies at least quadratically in the neighborhood of the center pixel. That is,

$$e(x, y) = \overline{e_0} + \overline{e_x}x + \overline{e_y} + \frac{1}{2}\overline{e_{xx}}x^2 + \overline{e_{xy}}xy + \frac{1}{2}\overline{e_{yy}}y^2 + O(\epsilon),$$

where $O(\epsilon)$ denotes higher-order variations, and (x, y) are the coordinates of neighboring image points relative to the center pixel (x_c, y_c).

Carrying out the algebra, it can be shown that

$$Det\{\boldsymbol{Q}\} = -2Det\{\boldsymbol{E}\}\ Tr\{\boldsymbol{E}\} + O(\epsilon)$$

where $\boldsymbol{E}$ is the Hessian of $e(x, y)$:

$$E = \begin{bmatrix} e_{xx} & e_{xy} \\ e_{xy} & e_{yy} \end{bmatrix}.$$

Ignoring terms of order $O(\epsilon)$, $Det\{\boldsymbol{Q}\}$ is nonzero if the determinant and trace of $\boldsymbol{E}$ are nonzero. We conclude that second-order brightness variations at the minimum, as well as conditions on the Hessian, are necessary for a nonsingular $\boldsymbol{Q}$. We may employ a least-square formulation using more than four points (for redundancy), to reduce sensitivity to various sources of noise (e.g., camera noise, image digitization, intensity quantization, and estimation of derivatives by finite differences). In this case, we have

$$\begin{bmatrix} e_x^1 & e_y^1 & -e^1 & -1 \\ e_x^2 & e_y^2 & -e^2 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ e_x^n & e_y^n & -e^n & -1 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta m \\ \delta c \end{bmatrix} = -\begin{bmatrix} e_t^1 \\ e_t^2 \\ \vdots \\ e_t^n \end{bmatrix},$$

where $n = N \times N$ is the number of points in the region. A least-square solution is given by $\left(\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{Q}}\right)^{-1}\left(\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{q}}\right)$, where $\hat{\boldsymbol{Q}}$ is the $n \times 4$ matrix on the left-hand side, and $\hat{\boldsymbol{q}}$ is the n-dimensional vector on the right-hand side of the above equation. This solution is equivalent to that given by (6). The necessary conditions given earlier must hold to ensure a nonsingular $\left(\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{Q}}\right)$.

ACKNOWLEDGMENTS

This paper presents research work with funding support from the U.S. National Science Foundation under grants IRI-8910967 (Revised to IRI-9296040) and BCS-9017179. The views expressed herein are those of the author and do not necessarily reflect the views of the funding agency. I am grateful to the reviewers for their constructive criticism and comments.

REFERENCES

[1] D.H. Ballard and O.A. Kimball "Rigid Body Motion From Depth and Optical Flow," *CVGIP*, vol. 22, no. 1, Apr. 1983.
[2] S.T. Barnard, and W.B. Thompson, "Disparity Analysis of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 333-340, July 1980.
[3] J.L. Baron, D.J. Fleet, and S.S. Beauchemin, "Performance of Optical Flow Techniques," *Int'l J. Computer Vision*, vol. 12, no. 1, 1994 (also TR-299, Dept. of Computer Science, Univ. Western Ontario, Mar. 1992).
[4] A. Blake and G. Brelstaff, "Geometry From Specularities," *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, L.W. Wolff, S.A. Shafer, and G.E. Healey eds. Boston, Mass.: Jones & Bartlett Publishers, 1992.
[5] A.R. Bruss and B.K.P. Horn, "Passive Navigation" *CVGIP*, vol. 21, 1983.
[6] E.N. Coleman and R.A. Jain, "Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, L.W. Wolff, S.A. Shafer, and G.E. Healey, eds. Boston, Mass.: Jones & Bartlett Publishers, 1992.
[7] N. Cornelius and T. Kanade, "Adapting Optical Flow to Measure Object Motion in Reflectance and X-Ray Image Sequences," *ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception*, Toronto, Canada, Apr. 1983.
[8] W.E.L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Vision System*, Cambridge, Mass.: MIT Press.
[9] W.E.L. Grimson, "Binocular Shading and Visual Surface Reconstruction," *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, L.W. Wolff, S.A. Shafer, and G.E. Healey eds. Boston, Mass.: Jones & Bartlett Publishers, 1992.
[10] E.C. Hildreth, *The Measurement of Visual Motion*. Cambridge, Mass.: MIT Press, 1983.
[11] B.K.P. Horn and B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, 1981.
[12] B.K.P. Horn, *Robot Vision*. Cambridge, Mass.: MIT Press, 1986.
[13] J.K. Kearney, W.B. Thompson, and D.L. Boley, "Optical Flow Estimation: An Error Analysis of Gradient-Based Methods With Local Optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, Mar. 87.
[14] J.R. Kender and E.M. Smith, "Shape From Darkness: Deriving Surface Information From Dynamic Shadows," *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, L.W. Wolff, S.A. Shafer, and G.E. Healey, eds. Boston, Mass.: Jones & Bartlett Publishers, 1992.
[15] H.C. Longuet-Higgins and K. Prazdny, "The Interpretation of a Moving Retinal Image," *Proc. Royal Soc. London B*, vol. 208, 1980.
[16] L. Matthies, R. Szeliski, and T. Kanade, "Incremental Estimation of Dense Maps From Image Sequences," *Proc. Computer Vision and Pattern Recognition*, Ann Arbor, Mich., June, 1988.
[17] N. Mukawa, "Estimation of Shape, Reflection Coefficients and Illumination Direction From Image Sequences," *Proc. Int'l Conf. Computer Vision '90*, Osaka, Japan, Dec. 1990.

[18] H.H. Nagel and Enkelmann "An Investigation of the Smoothness Constraints for the Estimation of Vector Fields From Image Sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 565-593, Sept. 1986.

[19] H.H. Nagel, "On a Constraint Equation for the Estimation of Displacement Rates in Image Sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 13-30, Jan. 1989.

[20] S. Negahdaripour, A. Shokrollahi, and M. Gennert, "Relaxing the Brightness Constancy Assumption in Computing Optical Flow," *Proc. ICIP-89*, Singapore, Sept. 1989.

[21] S. Negahdaripour, and C. H. Yu, "A Generalized Brightness Change Model for Computing Optical Flow," *Proc. Int'l Conf. Computer Vision*, Germany, 1993.

[22] S. Negahdaripour "Revised Representation of Optical Flow for Dynamic Scene Analysis," *Proc. ISCV*, Coral Gables, Fla., Nov. 1995.

[23] S. Negahdaripour, and L. Jin, "Direct Recovery of Motion and Range From Images of Scenes With Time-Varying Illumination," *Proc. ISCV*, Coral Gables, Fla., Nov. 1995.

[24] A. Pentland, "Photometric Motion," *Proc. ICCV-90* Osaka, Japan, Sept. 1990.

[25] B.G. Schunck, "Image Flow: Fundamentals and Future Research," *Proc. Computer Vision and Pattern Recognition '85*, San Francisco, Calif., June 1985.

[26] A. Spoerri and S. Ullman, "The Early Detection of Motion Boundary," *Proc. ICCV*, London, England, June 1987.

[27] W.B. Thompson and T.C. Pong, "Detecting Moving Objects," *Proc. Int'l Conf. Computer Vision*, June 1987.

[28] M. Tistarelli "Computation of Coherent Optical Flow by Using Multiple Constraints," *Proc Int'l Conf. Computer Vision*, Boston, Mass., 1995.

[29] S. Uras, F. Girosi, A. Verri, and V. Torre, "A Computational Approach to Motion Perception," *Bio. Cybernetics*, vol. 60, 1988.

[30] A. Verri, and T. Poggio, "Motion Field and Optical Flow: Qualitative Properties," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 490-498, May 1989.

[31] L.B. Wolff, "Shape From Lambertian Photometric Flow Fields," *Proc Computer Vision and Pattern Recognition*, San Diego, Calif., 1989.

[32] R.J. Woodham "Multiple Light Source Optical Flow," *Int'l Conf. Computer Vision* Osaka, Japan, Dec. 1990.

[33] R.J. Woodham, "Photometric Method for Determining Surface Orientation From Multiple Images," *Physics-Based Vision: Principles and Practice (Volume on Shape Recovery)*, L.W. Wolff, S.A. Shafer, and G.E. Healey, eds. Boston, Mass.: Jones & Bartlett Publishers, 1992 (also in *Optical Engineering*, vol. 19, no. 1, Jan. 1980).

[34] C.H. Yu, S. Negahdaripour, "Motion Recovery in Light-Attenuating Media From Image Shading Variations," *J. Optical Soc. of Am.*, vol. 9, no. 7, July 1992.

[35] J.Y. Zheng, Y. Fukagawa, and N. Abe, "Shape and Model From Specular Motion," *Proc. Int'l Conf. Computer Vision*, Boston, Mass., 1995.

[36] A. Zisserman, P. Giblin, and A. Blake, "The Information Available to a Moving Observer From Specularities," *Image and Vision Computing*, vol. 7, no. 1, 1989.

Shahriar Negahdaripour received SB, SM, and PhD degrees in 1979, 1980, and 1987 from MIT, Cambridge, MA. He joined the University of Hawaii as an Assistant Professor of Electrical Engineering in January 1987, and University of Miami as an Associate Professor of Electrical and Computer Engineering, in August 1991. He has been a member of the technical committee and session chair of major conferences and workshops in computer vision, pattern recognition and image processing, including the ICCV (Osaka, Japan, 1991), CVPR (Seattle, Washington, 1994), ICPR (Jerusalem, Israel, 1994), and IASTED International Conference on Signal and Image Processing (Las Vegas, Nevada, 1995, 1998). He was the General Co-Chair of CVPR (Maui, HI, 1991) and ISCV (Miami, FL, 1995). He has co-organized two NSF sponsored workshops, *Workshop on Challenges in Computer Vision: Future Directions of Research* (Maui, HI, June 1991), and *Workshop on Future Research Directions in Underwater Robotics,* (Maui, HI, August 1994). His current primary areas of research interest is the application of machine vision technology for autonomous and remotely operated submersible systems.