NISTIR 4919

Optimal Estimation of Optical Flow, Time-to-contact and Depth

Herbert Lau
Research Associate

Department of Electrical Engineering
University of Maryland
College Park, MD
and
Sensory Intelligent Group
Robot Systems Division

Tsai-Hong Hong
Martin Herman

Sensory Intelligent Group
Robot Systems Division

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

NIST

NISTIR 4919

Optimal Estimation of Optical Flow, Time-to-contact and Depth

Herbert Lau
Research Associate

Department of Electrical Engineering
University of Maryland
College Park, MD
and
Sensory Intelligent Group
Robot Systems Division

Tsai-Hong Hong
Martin Herman

Sensory Intelligent Group
Robot Systems Division

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

September 1992

U.S. DEPARTMENT OF COMMERCE
Barbara Hackman Franklin, Secretary

TECHNOLOGY ADMINISTRATION
Robert M. White, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
John W. Lyons, Director

Optimal Estimation of Optical Flow, Time-to-contact and Depth

Herbert Lau, Tsai-Hong Hong, and Martin Herman
Robot Systems Division
Bldg. 220, Rm B124, National Institute of Standards and Technology,
Gaithersburg, MD 20899

Keywords: closed-formed solutions, least squares estimation, optical flow, optimal estimates, spatiotemporal disparities, time-to-contact.

Abstract

Extraction of the optical flow field, time-to-contact[1] and depth from an image sequence is a vitally important problem in many contexts. Many robotic applications, e.g., obstacle avoidance, autonomous vehicles and the like, can benefit from this information. In this report, the problem is formulated under a least squares minimization framework. Based only upon a set of *spatiotemporal disparity*[2] measurements from a correlation algorithm, the algorithm in general recovers optimal estimates, in the sense of least squares, of both the *instantaneous* optical flow field and time-to-contact simultaneously. It does not utilize knowledge of camera translation or any internal camera parameters. Nevertheless, additional information can be recovered if prior information is incorporated into the algorithm. It is shown that if complete translational camera motion as well as some internal camera parameters are known, optimal estimates of *absolute depth* can directly be inferred from the disparity measurements. Closed-form solutions are obtained in all cases. The algorithm is simple, effective, and easy to implement, which enhances the possibility of hardware realization for real-time applications. Experimental results, both synthetic and real, are included to show the performance of the algorithm and to verify its robustness against random noise perturbation.

1 Introduction

1.1 Visual Motion

It is generally believed that the process of motion perception consists of two stages: *extraction* and *interpretation* of the *optical flow field.* Optical flow, the instantaneous 2D velocity flow field of the brightness patterns in an image [10] or on the retina, is a fundamental and important visual cue useful in many contexts. It can be induced by motion of

[1] The quantity, time-to-contact, can also be interpreted as scaled depth because the two quantities only differ by a scale factor. The two terms will be used interchangeably in this report.

[2] The relative displacement, $(\Delta x, \Delta y, \Delta t)$, of a feature point. A detailed description will be given later.

objects in the world, by motion of the observer, or by both. It is therefore of great interest to extract optical flow from a sequence of images. There are two main methodologies for extracting optical flow: gradient-based methods and correlation-based methods.

In general, gradient-based methods extract the optical flow field by computing spatial and temporal gradients of the image sequence [10, 13, 14, 22, 33, 41]. However, only one component of the flow (termed the normal flow) can be recovered from these spatiotemporal gradients. In some literature [4, 10], this shortcoming is referred to as the *aperture problem*, which states that the local information is not sufficient to recover the full flow field. Additional assumptions about the nature of the flow field [10], such as smoothness, are imposed to provide additional constraints to solve the problem, usually in an iterative fashion. However, these constraints are sometimes violated [22] and introduce problems at the regions with flow discontinuities. In addition, spatiotemporal derivatives are inherently noisy because of the noise-amplification nature caused by numerical differentiation. Hence, results are usually unstable and highly susceptible to noise.

The other methodology for extraction of optical flow is correlation-based using either features or block-matching techniques. In one approach, prominent features (e.g., corners, edges, or contours) are first extracted from one image and then matched with successive images to track the movement of the features [20]. In another approach, small patches of the image (e.g., a 5×5 template) are directly compared with nearby patches in the next frame by maximizing the correlation response [19, 21] or minimizing the sum-of-squared difference (SSD) [9, 11, 12, 16] in a window to estimate the relative displacement. The optical flow is calculated based upon the relative shift $(\Delta x, \Delta y)$ of the feature over the frame interval Δt. The relative shift $(\Delta x, \Delta y)$ is termed the spatial disparity for the feature point.

It is by no means trivial to establish and maintain such correspondence. In fact, the major complication in the correlation approach is sometimes referred to as the "correspondence problem" and results from failing to identify the feature in two images that are projections of the same entity in the 3D world. However, despite the difficulties associated with correlation-based methods, these methods appear to give stable and useful results. They have been widely adopted in many practical vision systems to track the movements of feature points on the image plane [8, 9, 16, 35].

In addition to spatial matching, one can also exploit the characteristics of temporal coherence and perform matching in the temporal domain. Temporal profiles are extracted from pairs of pixels which are then matched with one another to find the time shift of a feature point [7, 34]. By measuring this time shift, Δt, using cross correlation between the two signals, the optical flow can be deduced since the two pixels are separated by a known distance $(\Delta x, \Delta y)$. We term the measured quantity Δt the *temporal disparity* to explicitly distinguish this notion from spatial disparity $(\Delta x, \Delta y)$.

The second stage of the motion perception process is optical flow interpretation. There are two approaches used in this process. The first attempts to arrive at a description of the 3D scene (structure from motion problem) and the observer motion relative to the scene (passive navigation problem) from optical flow. The second approach, sometimes referred to as "purposive vision" [39], attempts to use optical flow directly, without 3D reconstruction, to derive information relevant to achieving a desired robot task or behavior. Scene information such as depth, surface orientation, and object motion can be inferred from optical flow. Many robotic applications such as obstacle avoidance, path planning, and object manipulation by a robot arm can benefit from this process. Heeger and Jepson [18] used least squares minimization to find 3D motion and depth. They solved for translation, then rotation, and finally depth in a step by step manner. Bolles *et al.* [6] determined structure from extended image sequences. Some authors have proposed incremental approaches to recover structure and motion from extended image sequences. Matthies *et al.* [9] recovered depth sequentially using the Kalman filter framework. Their approach, however, is limited to the lateral camera motion scenario[3]. Broida and Chellappa [15], Chandrashekhar and Chellappa [30], and Heel [23] used *extended* Kalman filtering techniques to sequentially recover motion and 3D structure. Young and Chellappa [17] generalized the use of Kalman filtering to a stereo system which recovered 3D motion and scene structure.

Time-to-contact is another quantity of interest in many applications particularly for autonomous navigation and obstacle avoidance. Time-to-contact is defined as the time duration for the robot to collide with a particular object which appears in the field of view. It is viewed as an important invariant in visual motion [37]. Subbarao [36] derived the upper and lower bounds for time-to-contact based on the first-order spatial derivatives of the optical flow field. Nelson and Aloimonos [38] used the divergence of flow field to estimate this quantity. Raviv and Herman [40] have described the relationship between time-to-contact and visual looming, and how these quantities can be used, without 3D reconstruction, to achieve obstacle avoidance or to approach an object without collision.

1.2 Framework Overview

The least squares algorithm is primarily designed for cameras undergoing rectilinear translation (e.g., a forward-looking, side-looking, or oblique-looking camera). It can be extended to handle camera rotation if *derotation* is first performed with the knowledge of interframe camera rotation furnished by some independent sensors. From the mathematical point of view, rotation complicates matters by introducing nonlinearity to the flow equation (see equation (3)). That makes the extraction and interpretation of optical flow more difficult [27]. Since the rotational term does not involve any 3D parameters, one can "undo" the effect due to rotation without any knowledge about the scene if the rotation is given by some independent motion sensors such as an Inertial Navigation System (INS). One way to do this is to subtract the flow due to rotation from the total flow. Bhanu *et al.* [8] performed matching to obtain a displacement field and then *derotated* the field based on rotational information resulting in an optical flow field that cor-

[3] In lateral camera motion, the camera optical axis is always perpendicular to the velocity vector.

responds to pure translation. This flow field is then interpreted. We devise a different scheme to compensate for the rotation by derotating the entire iconic image based on the given rotational information prior to doing any optical flow or feature extraction [32]. After derotation has been accomplished, an image sequence corresponding to pure translation remains. Hence, without loss of generality, the image sequence is assumed to be compensated for rotation. In the following, we show that flow and depth calculation can be simplified by exploiting some special properties of a purely translating camera.

Disparities are assumed to be obtained from correlation-based algorithms [7, 8, 9, 11, 19, 21, 34]. It can be a spatial correlation algorithm in which the relative displacements of feature points, or spatial disparities, are estimated over successive frames. It can also be a temporal cross correlation algorithm in which temporal, as opposed to spatial, disparities are estimated by performing 1D correlation between pairs of pixels. In this report, disparities are obtained by taking the latter approach [34].

This report deals with the problem of flow extraction and depth recovery for a camera undergoing pure rectilinear translation in a static environment under a least squares framework. It is based on the mathematical model of feature point trajectories (shown in section 2). Multiple disparity[4] measurements are taken and matched with a mathematical model to minimize the error, or discrepancy between the model and the observations. It is done by searching for an optimal set of model parameters, optical flow and scaled depth, to best explain the observed data.

A schematic diagram explaining the relationship between different modules and their functions is shown in figure 1. This paper concentrates on the least squares module. References [32, 34] provide methods for implementing the other modules.

1.3 Significance

Many researchers estimate optical flow field based on extracted spatiotemporal derivatives [10, 13, 14, 22, 23, 41]. For instance, Horn and Schunck [10] devised an iterative minimization procedure over the entire flow field based on the spatiotemporal derivatives. Lucas *et al.* [41] and Simoncelli *et al.* [42] described a least squares minimization procedure on the flow field in a small neighborhood based also on spatiotemporal derivatives. In this paper, the estimation of the optical flow field is based on spatiotemporal disparities which are believed to be more reliable.

We explicitly distinguish the difference between disparities and optical flow. Optical flow is the instantaneous velocity of feature points on the image plane at a particular instance of time. It can also be interpreted as the tangent or slope evaluated at a particular instance of time along feature trajectories. On the other hand, disparities are only discrete measurements of the differences in feature point positions. A widely accepted method is to approximate the instantaneous flow (u, v) with disparities using

[4] Disparities are assumed to be known. A precise definition of disparity will be given in later sections.

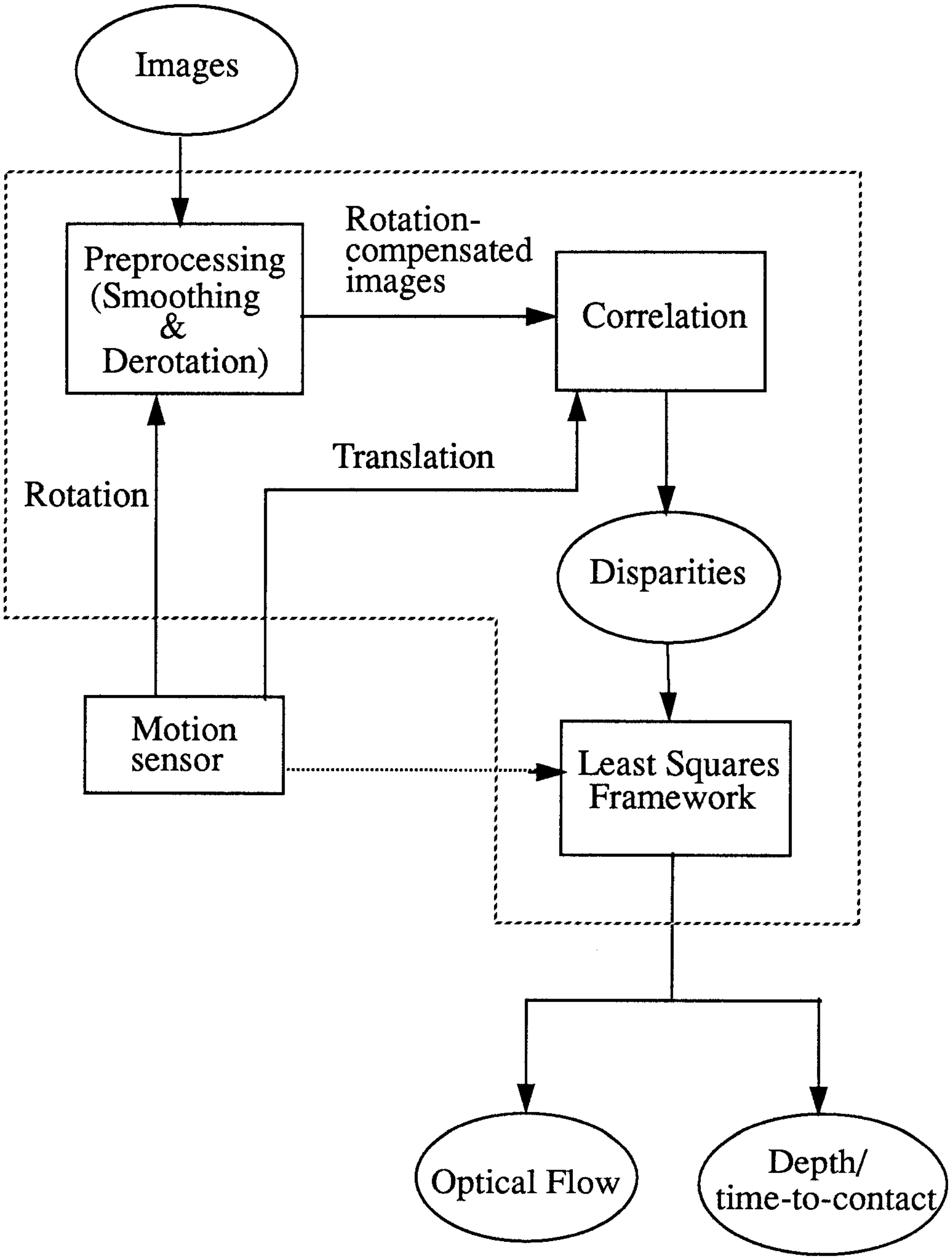

FIGURE 1. A schematic diagram showing the relationship between different modules in the flow and depth/time-to-contact estimation process. Camera motion is assumed to be known in the derotation and correlation modules and is optional in the least squares module.

$u \approx \Delta x/\Delta t$ and $v \approx \Delta y/\Delta t$. In this paper, no such approximations are made and they are treated as two different quantities. The exact instantaneous velocity of a feature point at a pre-selected instance of time is estimated by fitting a set of disparity measurements with a model. In other words, a better and more accurate optical flow estimate is obtained without any numerical approximations.

No camera translation or internal camera parameters, such as focal length, or camera center are required to estimate optical flow and time-to-contact (or scaled depth). A set of spatiotemporal disparities suffices for the estimation process. Hence, tedious camera calibration procedures can be avoided. However, more information can be recovered if additional prior information is incorporated into the algorithm. It is shown that the least squares estimation procedure can easily be modified to recover absolute depth directly if complete camera motion as well as some internal camera parameters are known a priori. The estimation of the time-to-contact is solely based on the disparity observations. It does not require any derived quantities of the optical flow field such as the first-order [36], second-order, higher-order derivatives or the divergence [38] of the flow field. This avoids the expensive and unreliable computations of those spatial derivatives.

1.4 Paper Overview

This report is organized as follows: In section 2, a general mathematical model relating optical flow to 3D motion parameters and depth is presented. The model is characterized by two coupled first order differential equations which depend nonlinearly on the translation, and rotation of the camera and the depth of the scene point. In section 3, we show that, if rotation is absent, the differential equation is decoupled and becomes a linear model. Moreover, the simplified differential equation is shown to be solvable which allows us to derive a model for the feature trajectory as a function of time. In section 4, a least squares formulation is presented to estimate the optimal (absolute/scaled) depth and optical flow on the basis of disparity measurements which are input to the algorithm. Some experimental results, both synthetic and real, are then presented in section 5. Final remarks are presented in section 6.

2 General Motion and Optical Flow

We first review the mathematical model for optical flow under perspective projection. This equation has been derived previously by numerous authors including Longuet-Higgins and Prazdny [1], Bruss and Horn [2], Waxman and Ullman [3], and Jepson and Heeger [18].

Consider a camera-centered coordinate frame (figure 2), where each scene point is associated with a position vector, $\vec{P} = (X, Y, Z)^T$, relative to the camera where T denotes the transpose. Due to the motion of the camera characterized by the 3D motion parameters $\vec{V} = (V_x, V_y, V_z)^T$ and $\vec{\Omega} = (\Omega_x, \Omega_y, \Omega_z)^T$, which denote the translation and

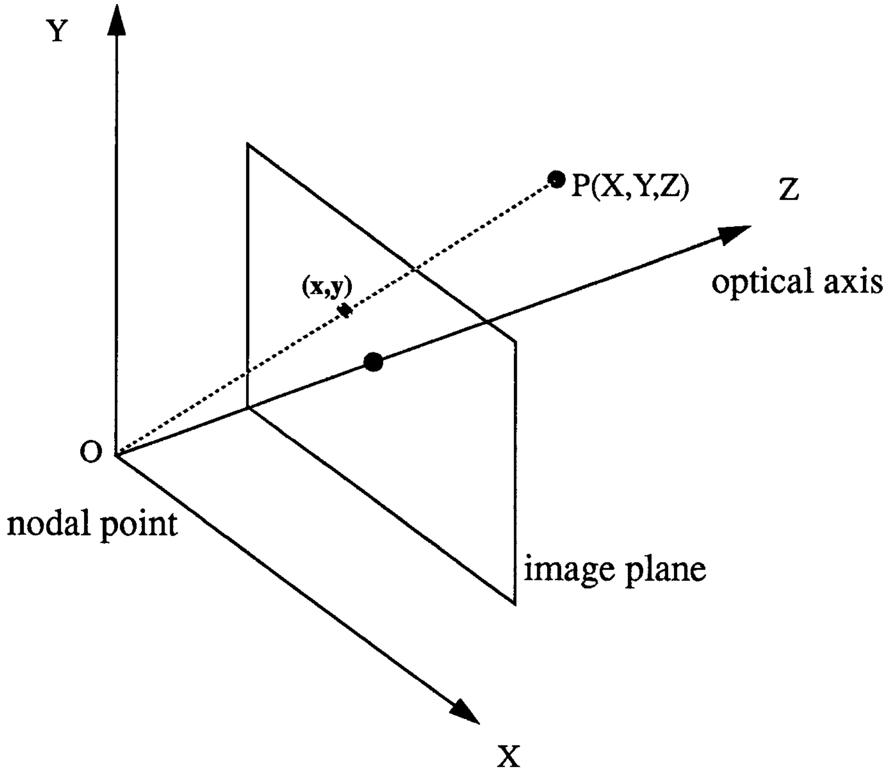

FIGURE 2. The camera-centered coordinate frame and the principle of perspective projection of a scene point P(X,Y,Z) onto the image plane at (x,y).

rotation vectors respectively, the motion of the 3D point is described by a first order vector differential equation:

$$\frac{d\vec{P}}{dt} = -\vec{\Omega} \times \vec{P} - \vec{V} \tag{1}$$

Under perspective projection, $\vec{P}$ projects onto the image plane at $\boldsymbol{x} = (x, y)^T$ according to:

$$\boldsymbol{x} = (x, y)^T = (fX/Z, fY/Z)^T \tag{2}$$

Optical flow at point $\boldsymbol{x}$ is defined as the time derivative of (2). By taking the derivative of (2) and substituting (1), we arrive at the following mathematical model for optical flow:

$$\dot{\boldsymbol{x}} = \boldsymbol{h}_t(\vec{V}, Z, \boldsymbol{x}) + \boldsymbol{h}_r(\vec{\Omega}, \boldsymbol{x}) \tag{3}$$

where $\dot{\boldsymbol{x}} = (\dot{x}, \dot{y})$ is the optical flow at point (x, y), Z is the depth of the point, $\boldsymbol{h}_t$ and $\boldsymbol{h}_r$ are, respectively, the translational and rotational components of the optical flow given by:

$$\boldsymbol{h}_t(\vec{V}, Z, \boldsymbol{x}) = \frac{1}{Z}\begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix}\vec{V}$$

$$\boldsymbol{h}_r(\Omega, \boldsymbol{x}) = \begin{bmatrix} \frac{xy}{f} & \frac{-(f+x^2)}{f} & y \\ \frac{f^2+y^2}{f} & \frac{-xy}{f} & x \end{bmatrix}\vec{\Omega}$$

Equation (3) is the vector equation for optical flow characterized by two coupled first order ordinary nonlinear differential equations which describe the image velocity for each feature point on the image plane as a function of translation, rotation of the camera and depth. The equation is purposely separated into two terms to emphasize the two different flow components, the translational and the rotational. One very important observation is that the rotational flow component is fully determined by the rotational velocity and is entirely independent of the structure of the scene while the translational component does depend on the depth of scene points. Moreover, from (3) we see that depth and velocity are multiplied together in the translational component. Therefore, they can only be estimated up to a scale factor. This mathematically explains why a monocular camera can at most recover scaled information from an image sequence, unless either depth or velocity is known a *priori.*

3 Feature Point Trajectory

Rotation is assumed to be compensated by derotation of the image sequence using the camera rotational information. Hence, without loss of generality, let us assume that images have already been compensated for rotational effects and that the camera undergoes only rectilinear motion. The flow model can be greatly simplified with the absence of rotation. The flow induced by a translating camera is governed by:

$$\dot{\boldsymbol{x}} = \boldsymbol{h}_t(\vec{V}, Z, \boldsymbol{x}), \tag{4}$$

or in terms of its components:

$$\begin{aligned} \dot{x} &= \frac{dx}{dt} = \frac{1}{Z}(xV_z - fV_x) \\ \dot{y} &= \frac{dy}{dt} = \frac{1}{Z}(yV_z - fV_y) \end{aligned} \tag{5}$$

The equation of the path for any feature point on the image plane as a function of time can then be derived from (5). As will be shown, the parameters for the path model are the *optical flow* and *scaled depth.* The derivation of the mathematical model is partly motivated by the work of Bolles and Baker [31]. Sethi and Jain [20] have also investigated the problem of finding trajectories of feature points.

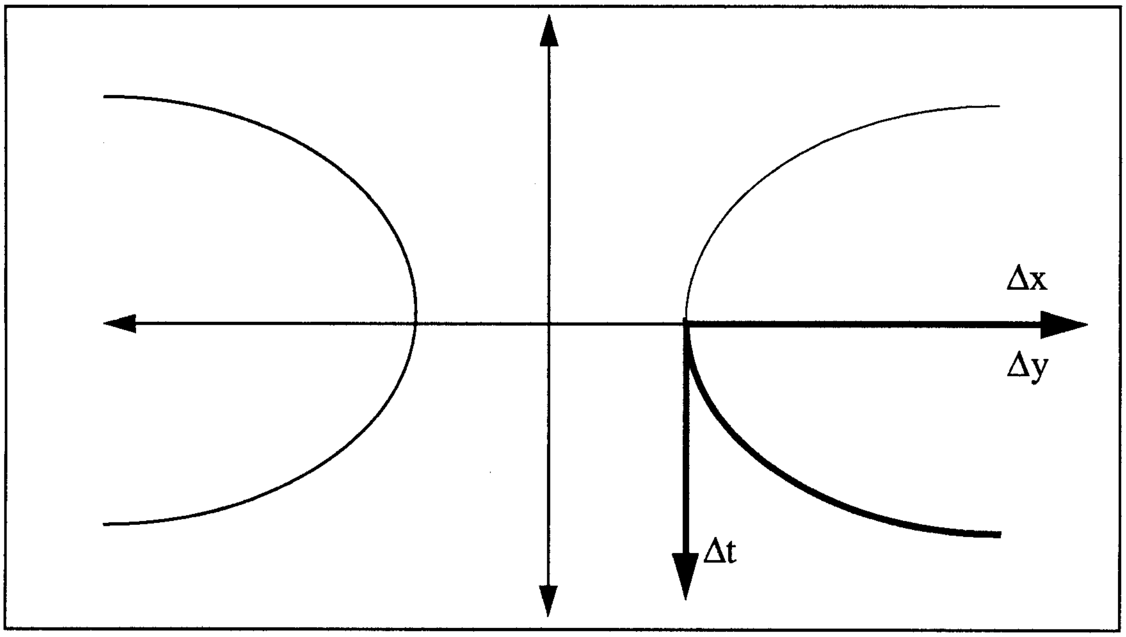

FIGURE 3. The trajectory of feature points (bold line) as a function of time: The *x* or *y* component of the trajectory has a hyperbolic shape in general and the curvature depends on the model parameters: optical flow and depth. In the case of a laterally moving camera, the curves degenerate into straight lines.

Let us assume that the camera velocity is constant; hence (5) is solvable if one knows how the depth, *Z*, evolves over time. Since the velocity is constant, depth will change linearly over time as follows:

$$Z(t) = Z_0 - V_z t \tag{6}$$

where Z_0 denotes the initial depth of the feature point at $t = 0$.

Substituting (6) into (5):

$$\begin{aligned} \frac{dx}{dt} &= \frac{xV_z - fV_x}{Z_0 - V_z t} \\ \frac{dy}{dt} &= \frac{yV_z - fV_y}{Z_0 - V_z t} \end{aligned} \tag{7}$$

Equation (7) is a set of first order linear differential equations and is easily solvable in (x, y) as a function of t subject to the following initial condition:

$$\text{for } t = 0, Z = Z_o, \text{ and } x = (x_0, y_0).$$

The solution (x, y), is given by two implicit relationships as follows:

$$
\begin{aligned}
Z_o(x-x_o) + (fV_x - V_z x_o)\,t - V_z(x-x_o)\,t = 0 \\
Z_o(y-y_o) + (fV_y - V_z y_o)\,t - V_z(y-y_o)\,t = 0
\end{aligned}
\qquad (8)
$$

Equation (8) is the equation of the path of any feature point as a function of time on the image plane. It will play an important role in subsequent analysis. Let us rewrite (8) by introducing a change of coordinates given by:

$$
\begin{aligned}
x &= x_o + \Delta x \\
y &= y_o + \Delta y \\
t &= \Delta t
\end{aligned}
\qquad (9)
$$

The solution, in vector form $\boldsymbol{g} = (g_1, g_2)$, can then be expressed as:

$$
\boldsymbol{g}(\Delta x, \Delta y, \Delta t) = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} Z_0\Delta x + (fV_x - V_z x_0)\,\Delta t - V_z\Delta x\Delta t \\ Z_0\Delta y + (fV_y - V_z y_0)\,\Delta t - V_z\Delta y\Delta t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (10)
$$

From (10), the trajectories of a feature point in the $(\Delta x\text{-}\Delta t)$ and $(\Delta y\text{-}\Delta t)$ domains are hyperbolas[5] (part of the trajectory is shown in figure 3). The quantity $(\Delta x, \Delta y)$ can be treated as the displacement of feature points relative to the initial position (x_o, y_o) after time Δt has elapsed. Many researchers refer to the term $(\Delta x, \Delta y)$ as *spatial* or *stereo disparity* [26] and the entire disparity map as the displacement field [11]. In most cases, spatial disparity is obtained by correlating a pair of images in order to set up the correspondence of feature points. As described above, we can approach the problem differently by estimating the time elapsed, Δt, for a feature to travel $(\Delta x, \Delta y)$ on the image plane by doing temporal correlation between intensity profiles extracted from pairs of pixels [7, 34]. We call this time elapsed, Δt the *temporal disparity.*

The least squares formulation presented in the next section is general and is independent of whether the input is spatial or temporal disparity. For this reason, we will introduce a unified notation, $(\Delta x, \Delta y, \Delta t)$, termed the *spatiotemporal disparity* or *spatiotemporal displacement vector* with respect to the initial feature point position at time equal to 0. The spatiotemporal disparity of a feature point can be interpreted as the relative displacement of that feature in the spatiotemporal volume (x-y-t).

[5] If we rotate the axes by 45°, the cross product terms $\Delta x\Delta t$ in the x-component of (10) can be eliminated. The new equation is a general equation of hyperbola and is given by the following:

$$
\frac{c}{2}(\Delta x'^2 - \Delta t'^2) + \frac{a+b}{\sqrt{2}}\Delta x' + \frac{a-b}{\sqrt{2}}\Delta t' = 0
$$

where $\Delta x'$, $\Delta t'$ are the new coordinates after axes rotation and a, b, and c are the coefficients of the terms Δx, Δt, and $\Delta x\Delta t$ in the x-component of (10) respectively. Similar axis transformation can be performed to obtain the general hyperbola equation for the y-component of (10).

4 Model-Based Least Squares Formulation

4.1 Motivation

In principle, two disparity measurements suffice to solve the equations for the optical flow and scaled depth (see below). But in practice, due to noise in the observations, it is more reliable to combine multiple measurements to reduce the effect of *outliers* in disparity observations. The idea of using multiple measurements for estimation to increase accuracy is adopted by numerous researchers. Bolles *et al.* [6] used long, densely sampled image sequences to estimate depth by exploiting the epipolar constraint. Broida and Chellappa [15], and Matthies *et al.* [9] tracked feature locations in extended sequences of images to improve the accuracy in estimation of parameters such as depth, translation and rotation. Okutomi and Kanade [24] used multiple images with different baselines to acquire a set of sum-of-squared-difference (SSD) curves, added them up and obtained the depth derived from the minimum of the resultant SSD curve. The latter authors claim that their method effectively eliminates ambiguity in matching and can be implemented in a massively parallel manner. However, the method is restricted only to lateral camera motion while the least squares integration approach described below has no such restriction and can be applied to arbitrary rectilinear translation. Lucas *et al.* [41] and Simoncelli *et al.* [42] proposed a least squares minimization procedure of spatiotemporal derivatives to a constant optical flow model in a local neighborhood. However, it is difficult to calculate all the spatiotemporal derivatives from numerical differentiation reliably. Our approach uses a feature point trajectory model for the optical flow and formulates the problem as a least squares minimization problem using spatiotemporal disparities.

In the following sections, a general framework based on the least squares (LS) formulation to combine several disparity observations and to obtain the best estimate of both optimal flow and scaled depth in the least squares sense will be presented. Method for obtaining the disparities is discussed in [34]. If we impose the additional constraint of known camera motion and internal camera parameters to the LS problem, the solution will be simplified and *absolute* depth can be directly inferred from disparity measurements without going through the optical flow calculation.

4.2 Least Squares Minimization

Combining (5) and (10), we obtain a format which is more suitable for our derivation:

$$\begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} \Delta x - u_0 \Delta t - \zeta_0 \Delta x \Delta t \\ \Delta y - v_0 \Delta t - \zeta_0 \Delta x \Delta t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{11}$$

where $\zeta_0 \equiv \frac{V_z}{Z_0}$ is the ratio between the forward velocity and depth, and (u_0, v_0) is the optical flow $(\dot{x}, \dot{y})$ for the point (x_0, y_0) at time equal to 0. The parameter, ζ_0, is the *inverse time-to-contact* or the *inverse scaled depth.* Equation (11) emphasizes that the trajectory model has optical flow (u_0, v_0) and inverse scaled depth ζ_0 as the model parameters.

In principle, two disparity vectors suffice to solve (11) in terms of (ζ_0, u_0, v_0). In practice, however, it is desirable to consider more than the minimum number of conditions. We therefore formulate the problem as a LS minimization problem. Assume that there are N (> 2) observations of the disparity vector, with each one denoted by

$$\boldsymbol{d}_i = (\Delta x_i, \Delta y_i, \Delta t_i) \text{ for } i = 1 \ldots N. \tag{12}$$

The set of vectors $\boldsymbol{d}_i$ are disparities relative to the initial position (x_0, y_0) at time equal to 0. These vectors can be seen as a set of discrete positions at a set of discrete time instants. Note that we are not restricted to any particular method for measuring the disparity. We can use spatial matching or temporal profile matching for disparity measurement and, using (12), we maintain a consistent description of the disparities in a vector format. The disparities can then be fed into the LS framework described below.

The following objective function or minimization criterion, J, defined as:

$$J = \sum_i \|g(\boldsymbol{d}_i)\|^2 = \sum_i (g_1^2(\boldsymbol{d}_i) + g_2^2(\boldsymbol{d}_i)) \tag{13}$$

where $\|\cdot\|$ is the Euclidean norm will be minimized.

Intuitively, J is a measure of the deviation of observations from the model in the square sense. The goal is to search for a set of model parameters, (ζ_0, u_0, v_0), which are most consistent with the recorded disparity observations in the least squares sense. In other words, we intend to solve the following residual minimization problem, namely,

$$\min_{\zeta_0, u_0, v_0} J \tag{14}$$

The solution, denoted by $(\hat{\zeta}_0, \hat{u}_0, \hat{v}_0)$ which minimizes (14), satisfies the following necessary condition:

$$\nabla J = \left(\frac{\partial J}{\partial \zeta_0}, \frac{\partial J}{\partial u_0}, \frac{\partial J}{\partial v_0}\right)\Bigg|_{(\hat{\zeta}_0, \hat{u}_0, \hat{v}_0)} = (0, 0, 0) \tag{15}$$

Expanding (15) term by term yields the following:

$$\left.\frac{\partial J}{\partial \zeta_0}\right|_{(\hat{\zeta}_0,\,\hat{u}_0,\,\hat{v}_0)} = \hat{\zeta}_0\sum_i \Delta t_i{}^2\,(\Delta x_i^2+\Delta y_i^2) - \sum_i (\Delta x_i^2+\Delta y_i^2)\,\Delta t_i + \hat{u}_0\sum_i (\Delta t_i)^2\Delta x_i + \hat{v}_0\sum_i (\Delta t_i)^2\Delta y_i = 0$$

$$\left.\frac{\partial J}{\partial u_0}\right|_{(\hat{\zeta}_0,\,\hat{u}_0,\,\hat{v}_0)} = -\sum_i \Delta x_i\Delta t_i + \hat{\zeta}_0\sum_i \Delta x_i\,(\Delta t_i)^2 + \hat{u}_0\sum_i (\Delta t_i)^2 = 0 \qquad (16)$$

$$\left.\frac{\partial J}{\partial v_0}\right|_{(\hat{\zeta}_0,\,\hat{u}_0,\,\hat{v}_0)} = -\sum_i \Delta y_i\Delta t_i + \hat{\zeta}_0\sum_i \Delta y_i\,(\Delta t_i)^2 + \hat{v}_0\sum_i (\Delta t_i)^2 = 0$$

Rewriting (16) in matrix form:

$$\begin{bmatrix} \sum_i \Delta t_i{}^2\,(\Delta x_i^2+\Delta y_i^2) & \sum_i (\Delta t_i)^2\Delta x_i & \sum_i (\Delta t_i)^2\Delta y_i \\ \sum_i (\Delta t_i)^2\Delta x_i & \sum_i (\Delta t_i)^2 & 0 \\ \sum_i (\Delta t_i)^2\Delta y_i & 0 & \sum_i (\Delta t_i)^2 \end{bmatrix} \begin{bmatrix} \hat{\xi}_0 \\ \hat{u}_0 \\ \hat{v}_0 \end{bmatrix} = \begin{bmatrix} \sum_i \Delta t_i\,(\Delta x_i^2+\Delta y_i^2) \\ \sum_i \Delta x_i\Delta t_i \\ \sum_i \Delta y_i\Delta t_i \end{bmatrix} \qquad (17)$$

The solution $(\hat{\zeta}_0, \hat{u}_0, \hat{v}_0)$ can be obtained by solving the system of linear equations defined in (17) which gives the least squares solution to the optical flow for point (x_0, y_0) at time equal to 0. In fact, a closed-form solution can be written down explicitly. If we re-express (17) in a more compact notation as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & 0 \\ a_{13} & 0 & a_{22} \end{bmatrix} \begin{bmatrix} \hat{\xi}_0 \\ \hat{u}_0 \\ \hat{v}_0 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \qquad (18)$$

then the solution to (17) can be written down explicitly as follows:

$$\hat{\xi}_0 = \frac{1}{\Delta}\,(a_{22}^2 b_1 - a_{12}a_{22}b_2 - a_{13}a_{22}b_3)$$

$$\hat{u}_0 = \frac{1}{\Delta}\,(-a_{12}a_{22}b_1 - (a_{13}^2 - a_{11}a_{22})\,b_2 + a_{12}a_{13}b_3) \qquad (19)$$

$$\hat{v}_0 = \frac{1}{\Delta}\,(-a_{13}a_{22}b_1 + a_{12}a_{13}b_2 - (a_{12}^2 - a_{11}a_{22})\,b_3)$$

where $\Delta = -a_{12}^2a_{22} - a_{13}^2a_{22} + a_{22}^2a_{11}$. Note that ζ_0, the *scale factor* between V_z and Z_0, is recovered. Absolute depth cannot be obtained from a monocular camera without further camera motion information.

4.3 Using Complete Motion Information in a Calibrated Camera

If full motion information (heading and magnitude) is supplied by a reliable sensor (e.g., an inertial sensor) [8], and the camera is calibrated so that focal length, aspect ratio, and image center are known), the LS problem takes a different form and we can solve for the optical flow together with *absolute depth* by the following modified LS minimization approach.

Observe that the only unknown parameter in the model described by (10) is Z_0. We observe N disparity vectors and fit them to the model in order to identify the only unknown parameter Z_0. In this case, the dimensionality of the parameter space reduces to one instead of three.

The objective function, or the minimization criterion, is still the same as in (14), i.e., the quadratic criterion. Minimizing the only unknown of the equation, Z_0, the LS solution is the following:

$$\left.\frac{\partial J}{\partial Z_0}\right|_{\hat{Z}_0} = \hat{Z}_0 \sum_i (\Delta x_i^2 + \Delta y_i^2) - \sum_i \Delta t_i \Delta x_i (V_z \Delta x_i - fV_x + V_z x_0) - \sum_i \Delta t_i \Delta y_i (V_z \Delta y_i - fV_y + V_z y_0) = 0 \qquad (20)$$

Hence, optimal $\hat{Z}_0$ is given by:

$$\hat{Z}_0 = \frac{\sum_i \Delta t_i \Delta x_i (V_z \Delta x_i - fV_x + V_z x_0) + \sum_i \Delta t_i \Delta y_i (V_z \Delta y_i - fV_y + V_z y_0)}{\sum_i (\Delta x_i^2 + \Delta y_i^2)} \qquad (21)$$

With the optimal depth, the optical flow, $(\hat{u}_0, \hat{v}_0)$ can be calculated from (5), i.e.,

$$\hat{u}_0 = \frac{(x_0 V_z - fV_x) \cdot \sum_i (\Delta x_i^2 + \Delta y_i^2)}{\sum_i \Delta t_i \Delta x_i (V_z \Delta x_i - fV_x + V_z x_0) + \sum_i \Delta t_i \Delta y_i (V_z \Delta y_i - fV_y + V_z y_0)}$$

$$\hat{v}_0 = \frac{(y_0 V_z - fV_y) \cdot \sum_i (\Delta x_i^2 + \Delta y_i^2)}{\sum_i \Delta t_i \Delta x_i (V_z \Delta x_i - fV_x + V_z x_0) + \sum_i \Delta t_i \Delta y_i (V_z \Delta y_i - fV_y + V_z y_0)} \qquad (22)$$

5 Experimental Results

5.1 Synthetic Data - Forward Moving Camera

The LS algorithm has been tested with synthetic data to verify its correctness and its effectiveness in the presence of additive white Gaussian noise. In the experiment, a scene point of interest initially has coordinates (10 mm, 20 mm, 200 mm) with respect to the camera. The camera is moving along its optical axis at 50 mm/s. A set of disparities is generated by assuming a pinhole camera model with a focal length of 16mm. These disparities are then fed into the LS algorithm. The simulation results are presented in figures 4, 5 and 6. In figure 4, the scaled depth and its error are plotted against the number of disparities used. As expected, more disparities improve the accuracy of the estimation and the estimate converges to the true value after about 20 disparities. Similar conclusions can be made for the optical flow estimate. In figures 5 and 6, the x and y components of the optical flow and their errors are plotted respectively.

The previous experiment acts as a control experiment and the only error associated with the disparities is quantization error. In the second experiment, zero-mean additive white gaussian noise with σ equal to 0.04 is introduced to perturb the synthetic disparity data to observe the effect of random noise on the LS algorithm. In general, the convergence rate is somewhat slower and the algorithm requires more disparities to perform within a certain error allowance. Figures 7 through 9 show the simulation results.

5.2 Real data - Laterally Moving Camera

5.2.1 Poster Experiment

In this experiment, the image sequence is captured by a camera moving laterally in front of a flat poster which is at a constant (known) depth from the camera. The images are captured by a SONY^{TM} XC-57 CCD[6] camera. The camera moves parallel to the poster at an increment of 0.875 mm between each frame taken and the sequence consists of 70 frames (figure 10). Since all the feature points have the same depth from the camera, they will have the same optical flow throughout the image plane. By examining the statistics of the optical flow in the image plane and comparing it with ground truth (the true optical flow), we can determine the *accuracy* (how close it is to ground truth) and the *repeatability* (how repeatable the reading is throughout all feature points, i.e., a measure of the dispersion of the data) of the flow extraction algorithm. We measure accuracy as the average of the optical flow values extracted; we measure repeatability as the root mean square error of the flow.

[6] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily best for the purpose.

One of the difficulties that has to be confronted is to obtain ground truth reliably. This involves the measurement of the depth of the poster and the focal length of the camera. We utilized Tsai's *monoview coplanar* method [25] for camera calibration. His method requires a calibrated pattern with known dimensions. The output of the algorithm includes the focal length of the camera, as well as the homogeneous matrix describing the geometric relationship between the camera frame and the calibrated object frame in space. From this homogeneous matrix, we can deduce the relative orientation and distance between the two frames, and hence the absolute range of the calibration object with respect to the camera frame. With the known absolute depth of the poster[7], the focal length of the lens, and the known motion of the camera, the true optical flow can be deduced and compared with the output of the flow extraction algorithm.

The image sequence is fed into our flow extraction algorithm [34]. Our main goal is to demonstrate the effectiveness of the algorithm and the LS procedure. A correlating window of size 7×1 is employed in a temporal cross correlation algorithm. The mean and root mean square error of the optical flow are calculated across the image. Figure 11a shows the mean of the extracted optical flow value versus the true optical value derived from the calibrated results. The flow converges to the true value as more disparities are used in the LS framework. This reflects the accuracy of the flow extraction algorithm. In addition, as a comparison, another graph is plotted in figure 11a. This is the mean of the optical flow obtained by using only one disparity measurement and without using the LS procedure. The mean without using the LS procedure exhibits somewhat random and unpredictable behaviors while the mean obtained from the LS procedure approaches the true value in a consistent manner. Figure 11b show the calculated root mean square error of the flow as a function of the number of disparities used in the LS framework. It, again, decreases as the number of disparities increases which demonstrates the improved repeatability. Once again, the root mean square error of the optical flow obtained from using only one disparity measurement is plotted in figure 11b for comparison. It does not have a very definite trend as opposed to the one obtained from the LS module which decreases steadily towards zero. In addition, the error from the LS procedure is smaller than the error obtained without using the LS estimation except for the first 5 disparities. This shows the improvement of the LS procedure in the flow estimation process.

5.2.2 Poster and Can Experiment

In this experiment, a can is placed in front of the poster. The image sequence consists of 50 frames (figure 12) with the camera moving laterally in front of the scene. The correlating window is 7x1 in the temporal cross correlation algorithm and the number of disparities used for the LS module is 15 for every feature point. Each image is filtered with a Gaussian function to eliminate the high frequency noise. The resulting optical flow is displayed in three formats, one in terms of intensity-encoded pictures (figure 13b), the other in a 2D vector field (figure 13c), and a 3D plot (figure 13d). The presence of the can is characterized by different optical flow as shown in figures 13b and c.

[7] The poster was put on top of the calibration plate after camera calibration.

Figure 13b is the x-component of the optical flow encoded in intensity with the lighter region representing larger optical flow. The presence of the can in front of the flat poster is obviously detected. This can also be seen from figure 13c which is the 2D vector field of the scene at frame 3. The boundary of the can is not sharp and can be explained by the fact that the correlating window covers two regions (can and background) with different depths. The SSD graph evaluated at the boundary can be characterized by *bimodality*, i.e., it has two local minimum points [12]. This has an adverse effect on the matching algorithm since the bimodality might confuse the matching criterion resulting in false matches. Some researchers have suggested an adaptive windowing approach [29] to alleviate this problem. An adaptive window changes its size to cover only those regions with similar depths when performing matching.

5.2.3 Tree Sequence Experiment - Outdoor Scene

The tree sequence is a real outdoor image sequence provided to us by SRI. It has 128 frames and each frame is captured by incrementing the camera 1.2cm between frames. A set of spatiotemporal disparities is first extracted by our temporal cross correlation algorithm [34] and is then fed into the LS procedure for optical flow estimation. A 7x1 correlating window is employed in the temporal cross correlation [34]. A total of 10 disparities per feature point are used in the LS procedure for optical flow estimation. One original frame of the sequence (frame 3) is displayed in figure 14a and the optical flow obtained at this instance of time is encoded in intensity values and displayed in figure 14b.

5.3 A Comparison between Different Estimation Schemes

If the camera velocity component along the optical axis is known, it is possible to recover absolute depth from time-to-contact. However, it is also possible to directly estimate absolute depth if complete camera motion and some internal camera parameters are known. In this section, a comparison is made using synthetic data, between the absolute depth obtained from (19) and that from (21). Based on the results, the depth obtained from (21), i.e., the result obtained using more prior information in the estimation process, is more accurate than that obtained using less prior information.

A scene point has coordinates (10, 20, 200) mm initially relative to the camera which is translating at (10, 20, 50) mm/s in space. A set of spatiotemporal disparities is generated using perspective projection. A focal length of 16 mm is assumed. The disparities, corrupted by additive Gaussian noise with sigma equal to 0.001, are used in both estimation schemes. Absolute depths are independently recovered using (19) and (21). The estimation errors are plotted using overlays in figure 15. It can be seen that the estimate obtained by utilizing prior information is more accurate.

As expected, the estimate from (21) degrades if there are significant errors associated with the prior information. Errors of 0.0625% in the focal length, 0.1% in the x-component, and 0.1% in the y-component are introduced. The error of the estimate based on erroneous prior information is plotted in figure 15. The result agrees with the intuition that less accurate estimates are obtained when using less information.

5.4 Real Data - Forward Looking Camera

Optical flow induced by a forward moving camera is notoriously difficult to extract [5,9]. An image sequence with 100 frames (figures 16a and b) for a forward looking camera is captured to test the algorithm. A flat bull's eye pattern is placed in front of a forward moving camera orthogonal to the velocity vector of the camera. It is captured by the same experimental setup as described in the previous sections. The disparities are first calculated with our temporal cross correlation algorithm and then fed into the LS module. Since the optical flow is extremely small in the vicinity of the FOE, the temporal disparity is extremely large and is hard to estimate especially when the pair of pixels are far away from one another. In order to circumvent this problem, the set of temporal profiles chosen for temporal cross correlation is obtained by linear interpolation of the regular rectangular grids to limit the pixel separation distances. Interpolation between pixels to get the temporal profiles might not be necessary if the resolution of the chip of the CCD camera is very high around the FOE. A set of 20 disparities is used to calculate optical flow for each feature point.

The magnitude of the optical flow field at frame 3 is encoded in intensity and is plotted in figure 16d. A 2D vector field is also plotted in figure 16e. It agrees with our intuition that the optical flow is smaller (dimmer) in the vicinity of the FOE, and is larger (brighter) when far way from the FOE.

The algorithm has also been tested on another forward looking camera sequence obtained from NASA Ames (courtesy of Dr. Sridhar Banavar). It has 151 frames with objects moving no more than a pixel per frame. The first and the last frame of the sequence are shown in figures 17a and 17b. The third frame, for which the optical flow and scaled depth are evaluated, is shown in figure 17c. The scaled depth and 2D optical flow field are also shown in figures 17d and 17e respectively.

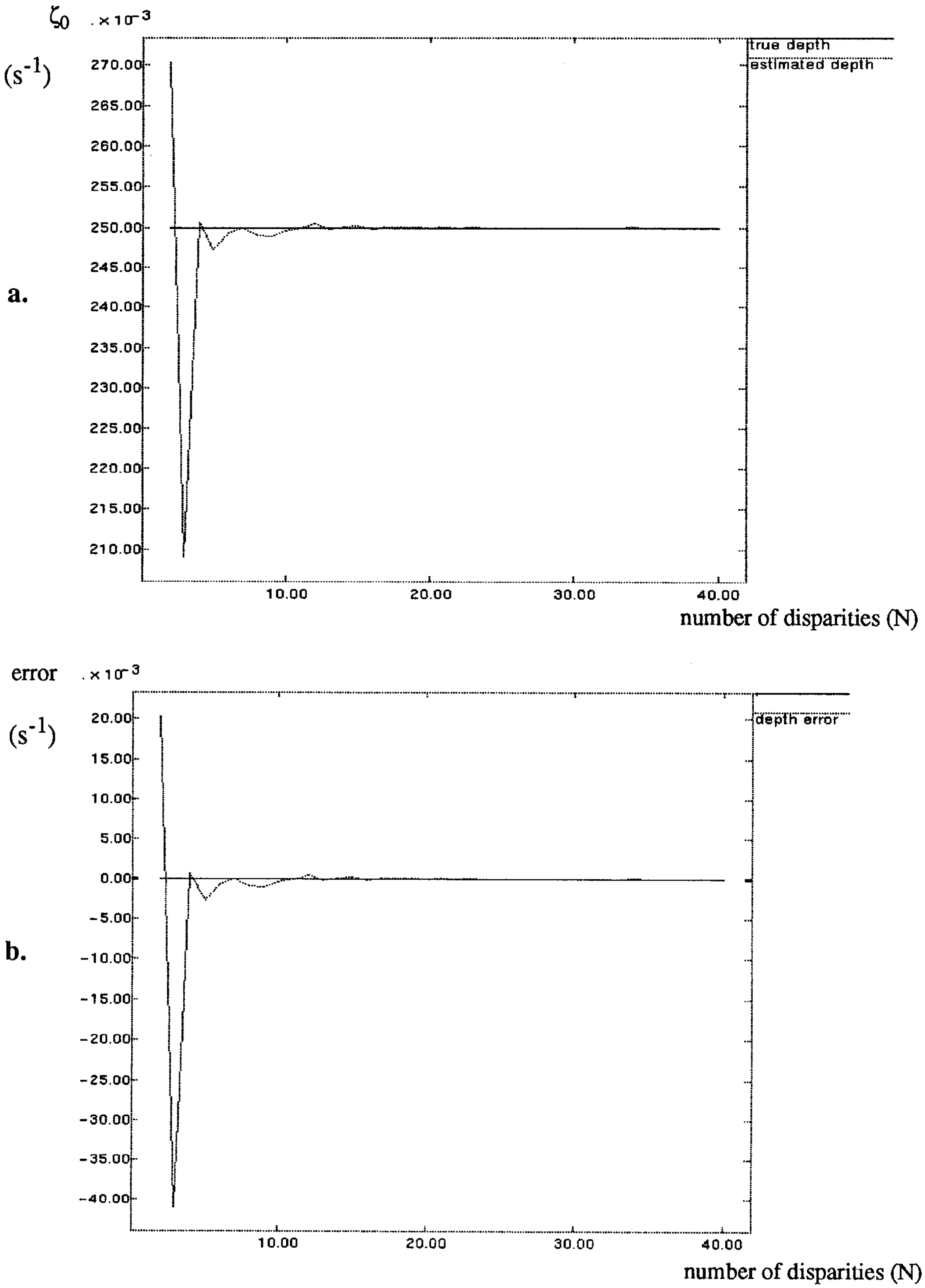

FIGURE 4. Synthetic data: Depth and error as a function of the number of disparities for a forward looking camera.

a) The scaled depth converges to the true value after about 20 disparity inputs.

b) The deviation of the depth estimate from the true depth decays as the number of disparities used in the LS algorithm increases.

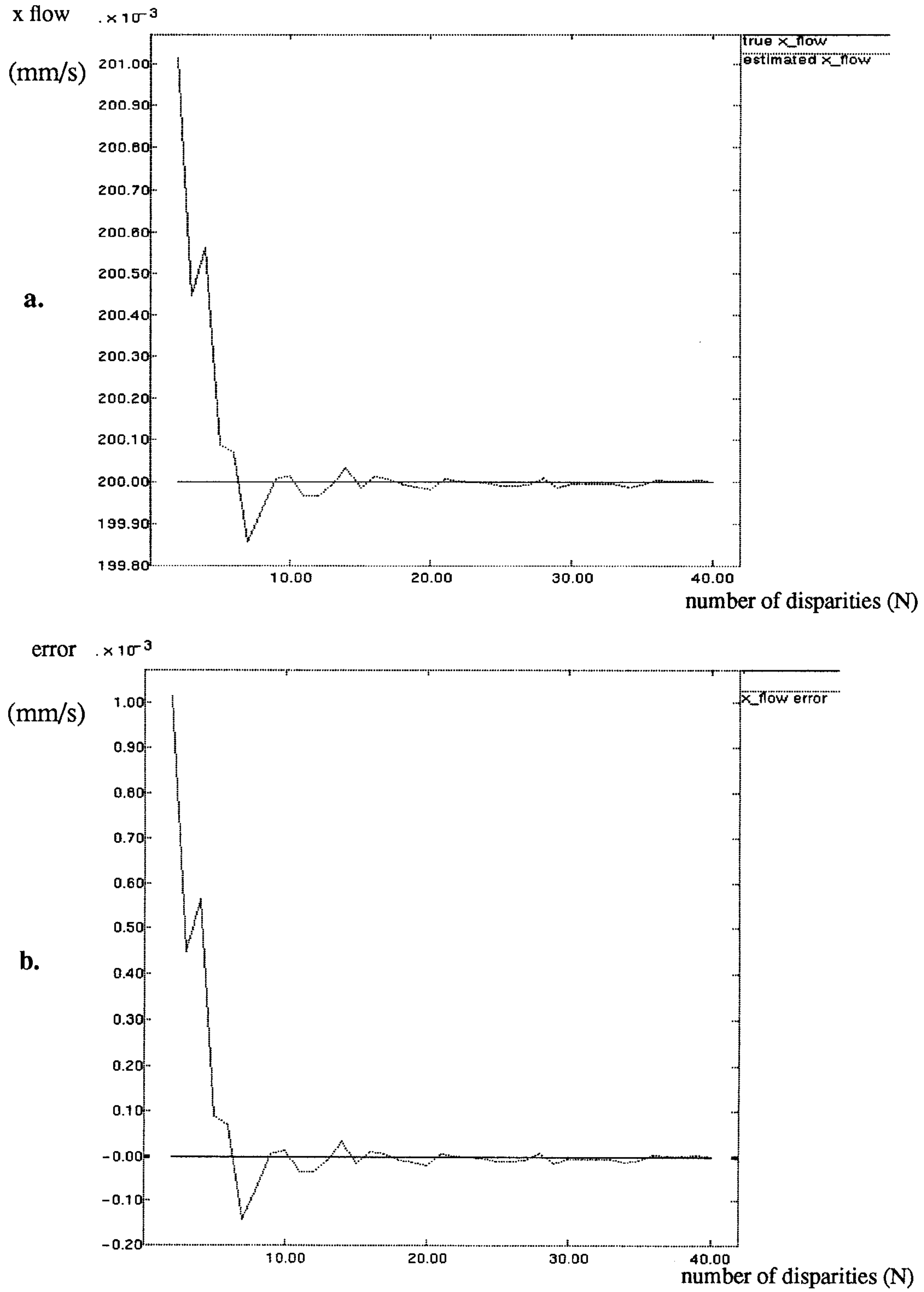

FIGURE 5. Synthetic data: Optical flow and error as a function of the number of disparities for a forward looking camera.

a) x-component of the optical flow as a function of the number of disparities used in the LS algorithm.

b) The deviation of the estimate from the true optical flow value.

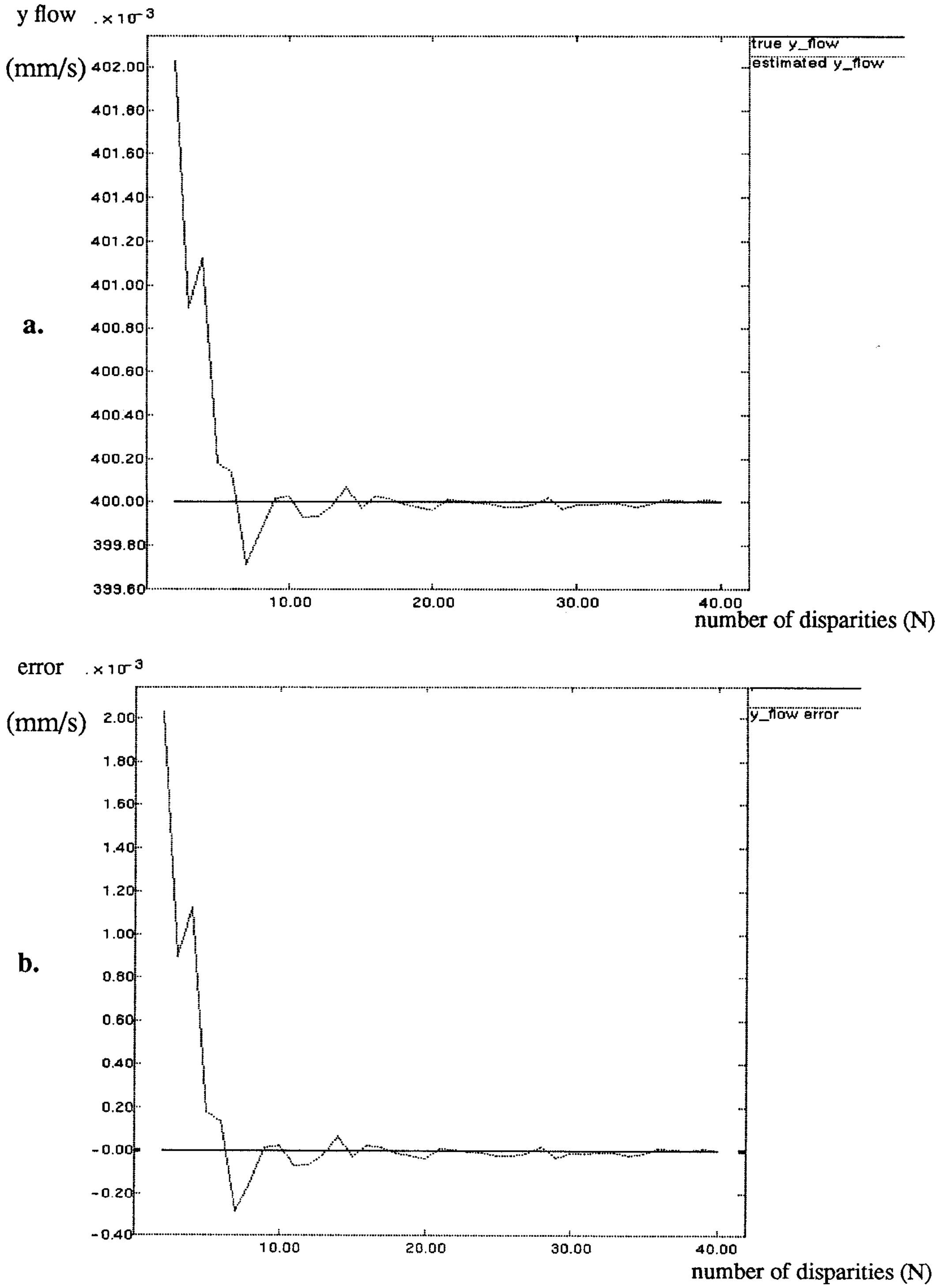

FIGURE 6. Synthetic data: Optical flow and error as a function of the number of disparities for a forward looking camera.

a) y-component of the optical flow as a function of the number of disparities used in the LS algorithm.

b) The deviation of the estimate from the true optical flow value.

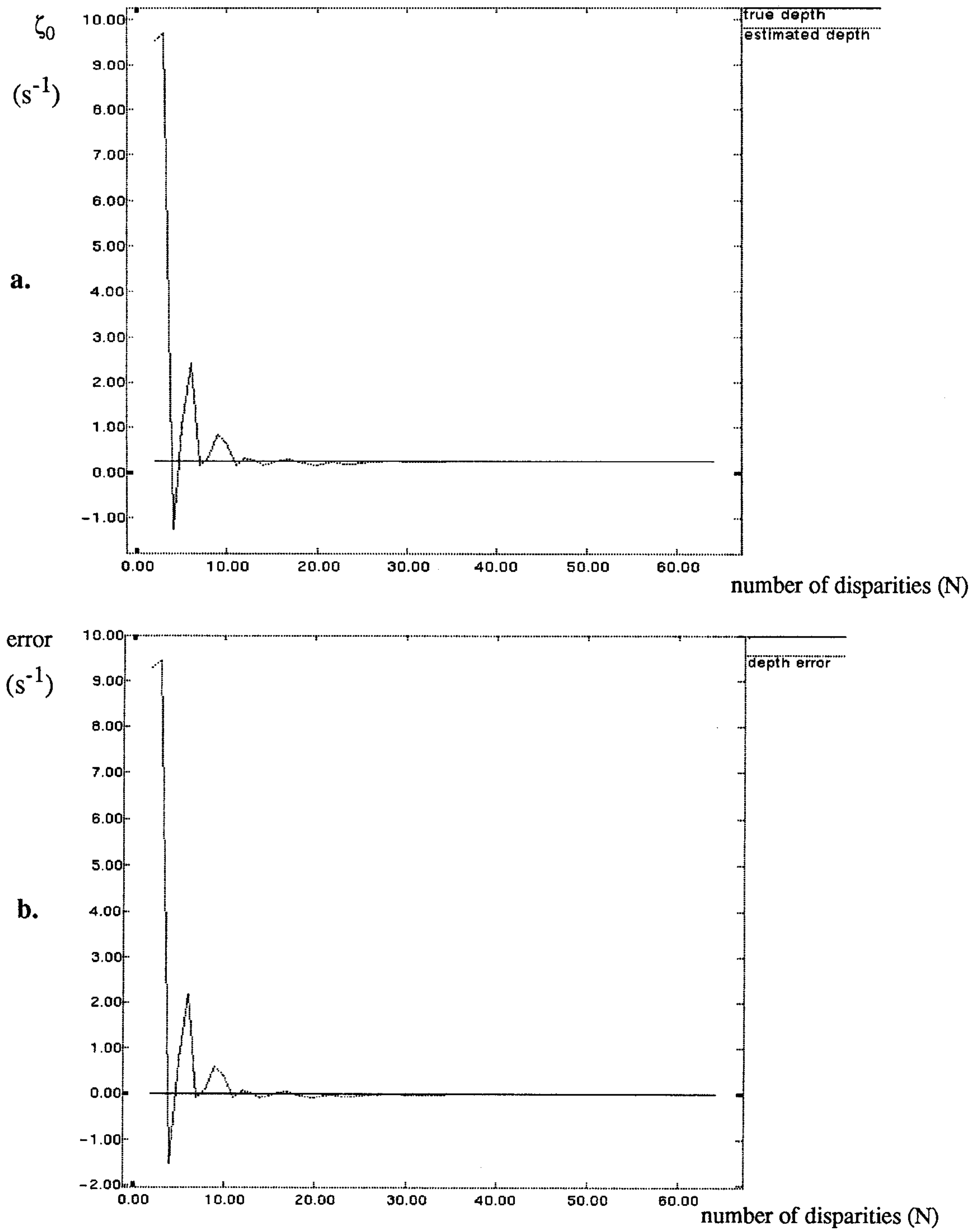

FIGURE 7. Noise corrupted version: The synthetic data are corrupted by additive white Gaussian noise with σ=.04:

a) The scaled depth converges to the true value after using about 20 disparities.

b) The deviation of the depth estimate from the true depth decays as the number of disparities used in the LS algorithm increases.

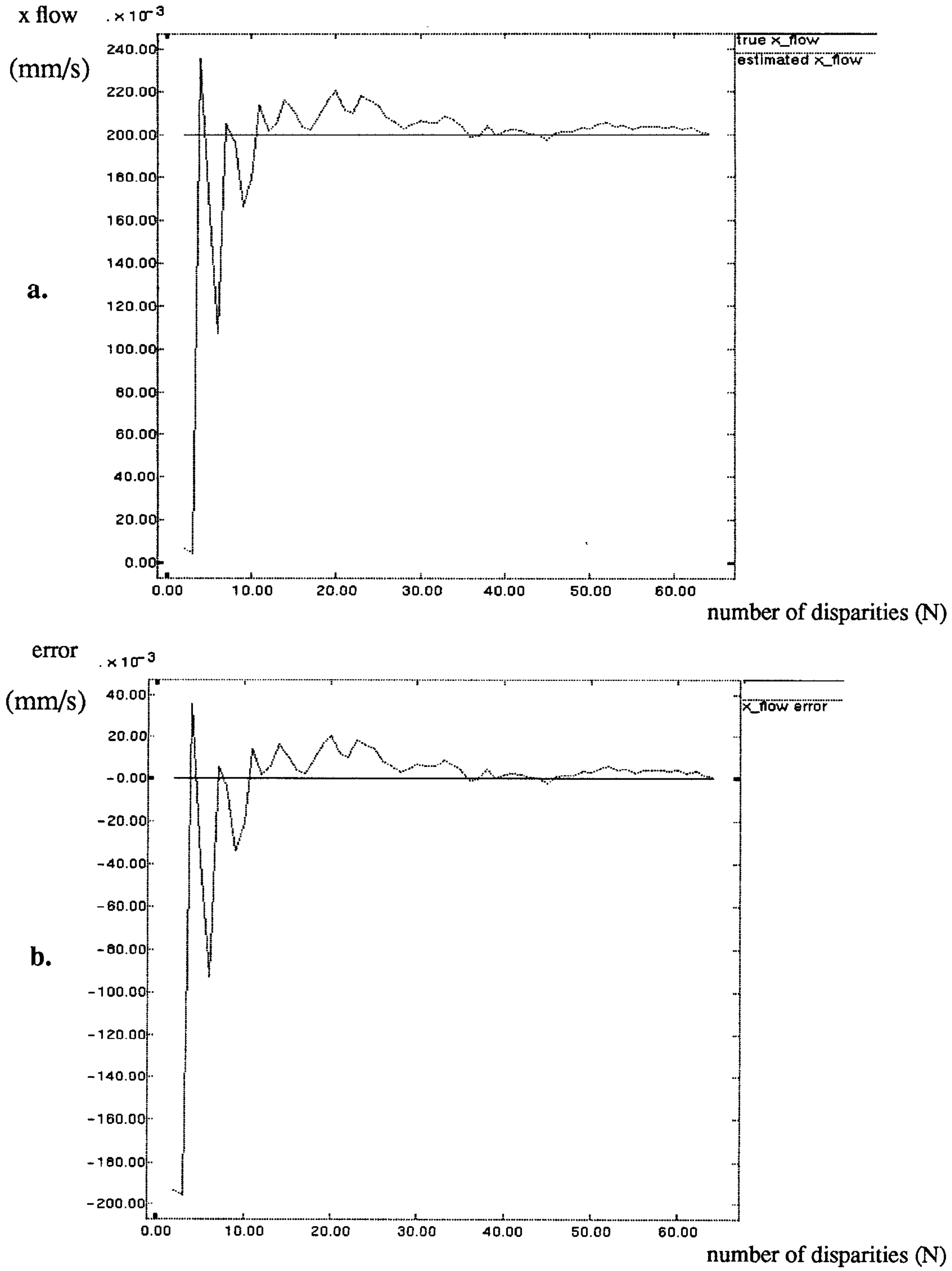

FIGURE 8. Noise corrupted version: The synthetic data are corrupted by additive white Gaussian noise with σ=.04:

a) x-component of the optical flow as a function of the number of disparities used in the LS algorithm.

b) The deviation of the estimate from the true optical flow value.

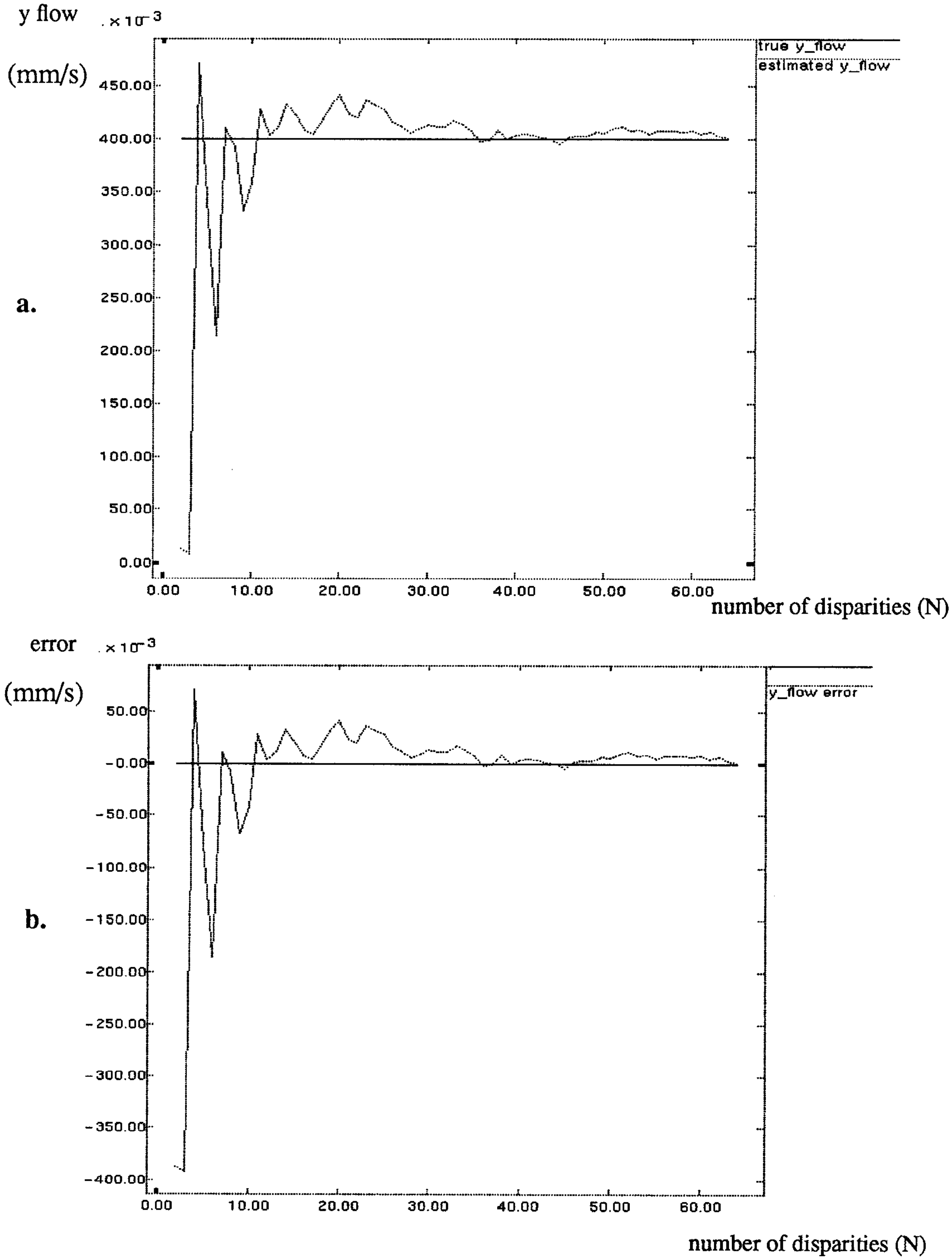

FIGURE 9. Noise corrupted version: The synthetic data are corrupted by additive white Gaussian noise with σ=.04:

a) y-component of the optical flow as a function of the number of disparities used in the LS algorithm.

b) The deviation of the estimate from the true optical flow value.

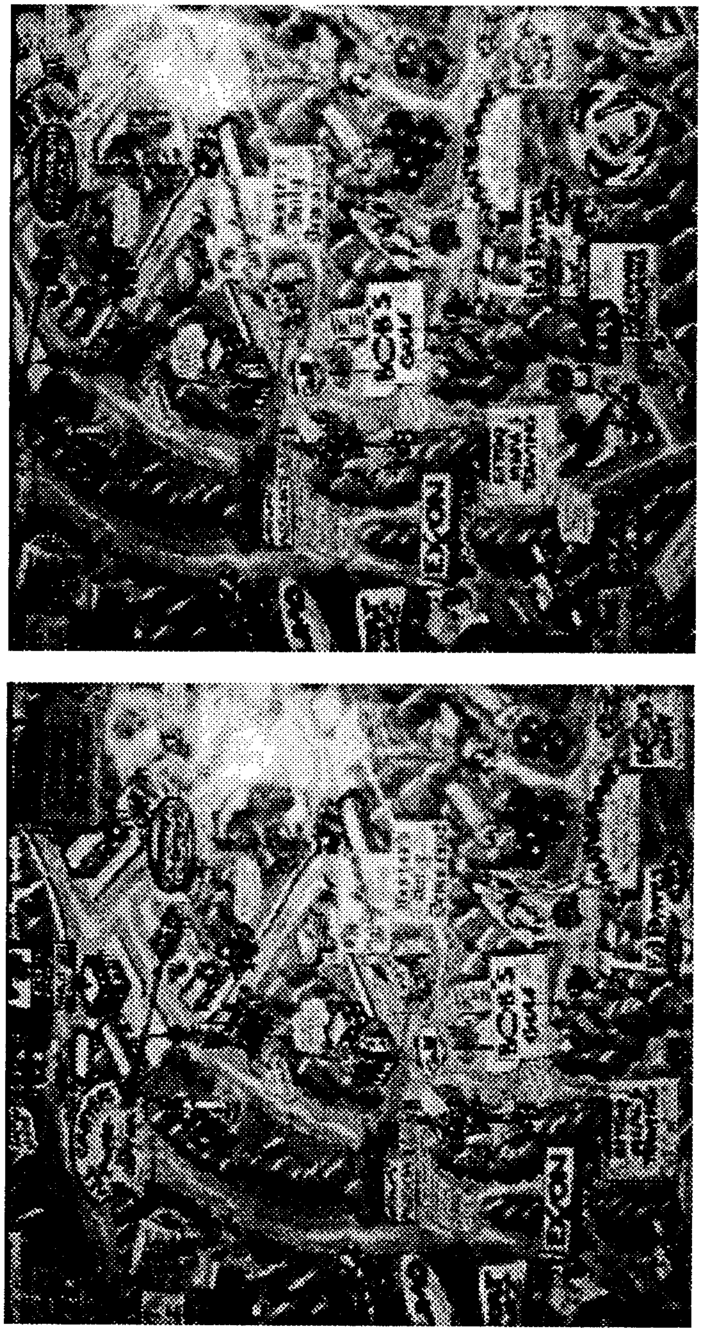

FIGURE 10. The first and the last frame of the poster sequence which consists of 70 frames.

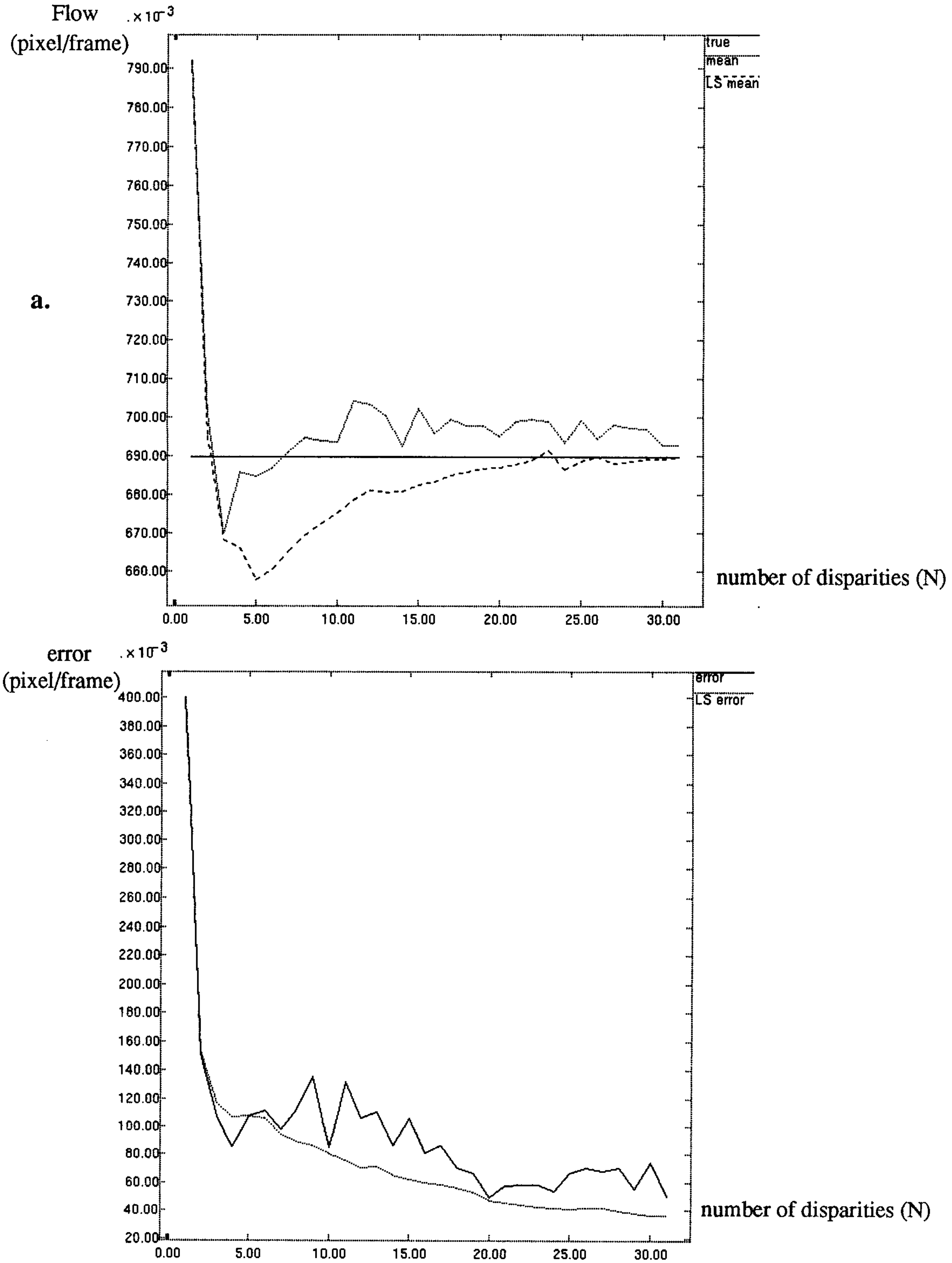

FIGURE 11. Poster experiment: a) The optical flow at frame 3 from the LS algorithm vs. the true flow value from the calibration algorithm. The extracted flow approaches the ideal flow value as the number of disparities used in the LS module increases. The flow estimated without using LS procedure is also included for comparison. b) The root mean square error of the extracted flow at frame 3 decreases as the number of disparities used increases. The error of the flow estimate obtained without using LS procedure is also higher than that obtained using LS procedure.

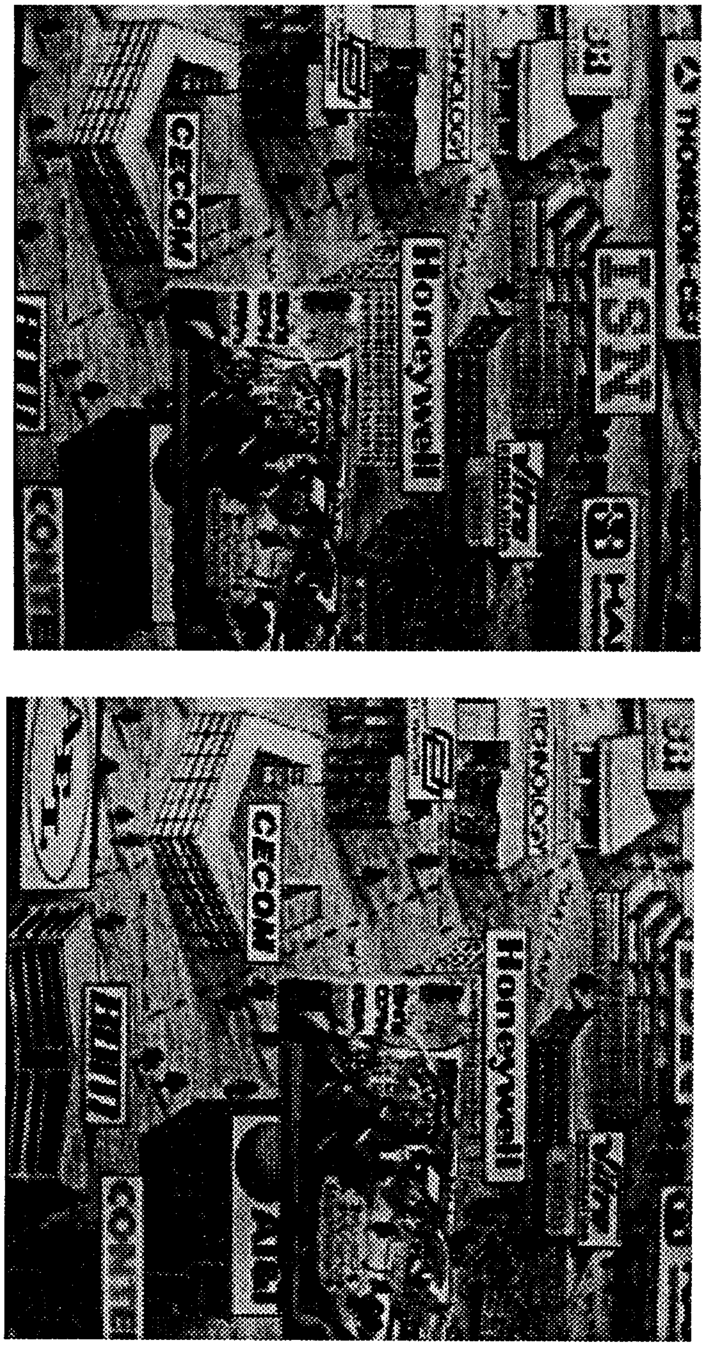

FIGURE 12. The first and the last frame of the can and poster sequence which consists of 50 frames. The can has been covered with textured paper.

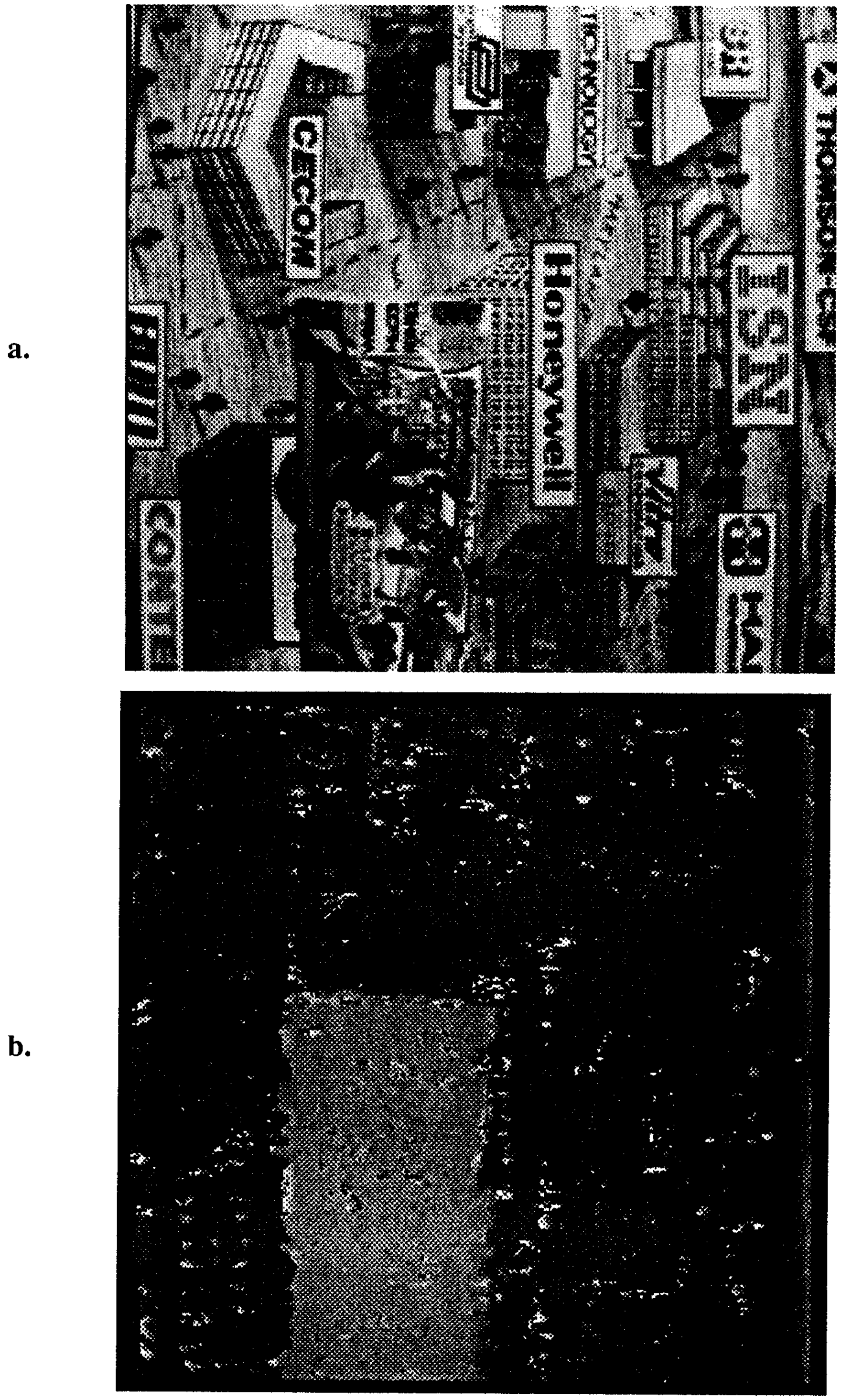

FIGURE 13. Poster and can experiment: a) Frame 3 of the sequence; the optical flow is evaluated at this instance of time. Note that the can is hardly visible in front of the poster. b) The depth of the image at time instant 3. Depth is encoded in intensity values. Lighter intensity represents smaller depth from the camera. The algorithm can readily perform segmentation and distinguishes the presence of the can in front of the poster.

c.

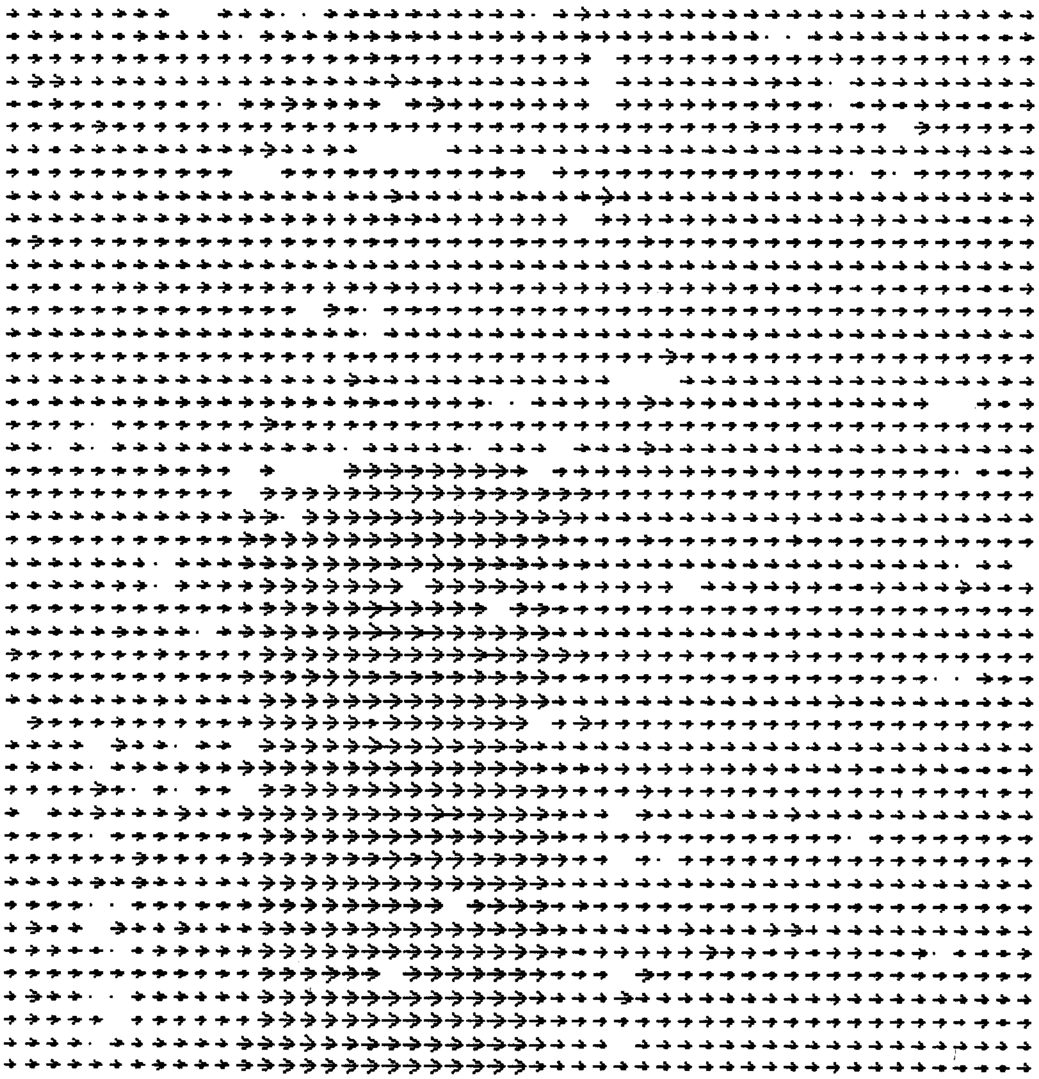

FIGURE 13. Poster and can experiment: c) The 2D optical flow field evaluated at frame 3. The can, which is closer to the camera, has larger optical flow and can be easily seen

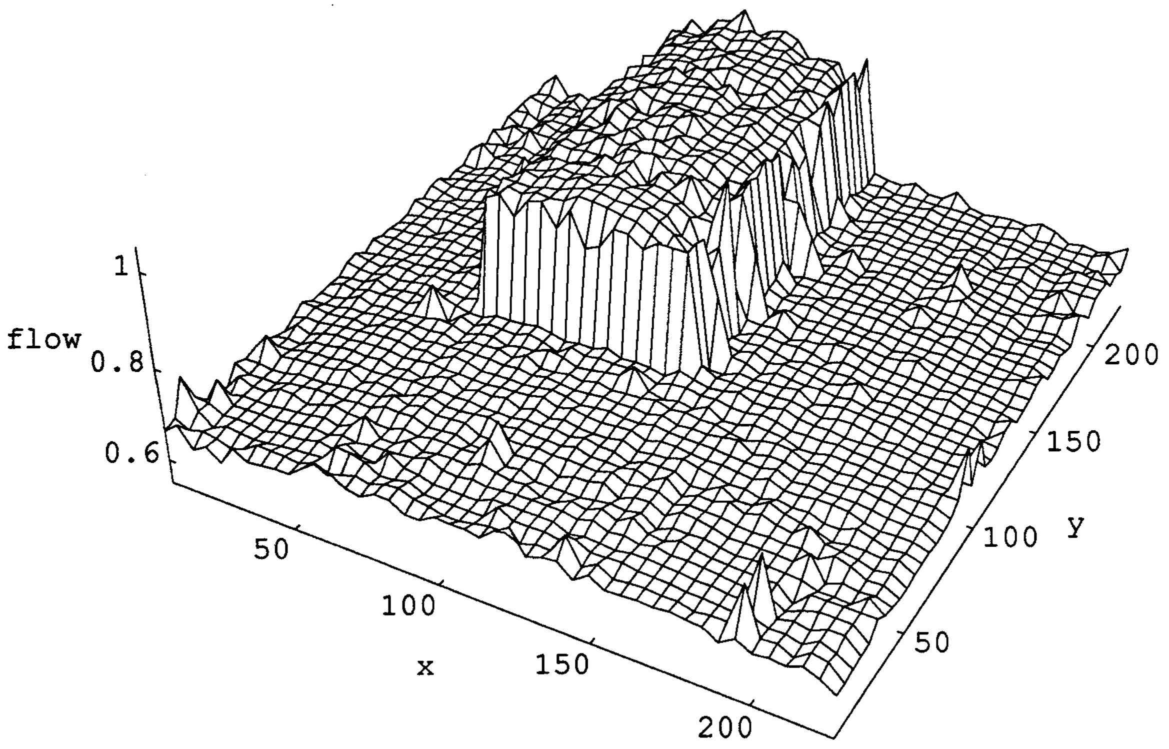

FIGURE 13. Poster and Can: d) Perspective plot of the flow field evaluated at frame 3. A 3x 3 median filter is employed on the flow field. The viewpoint has also been altered to enhance the visibility.

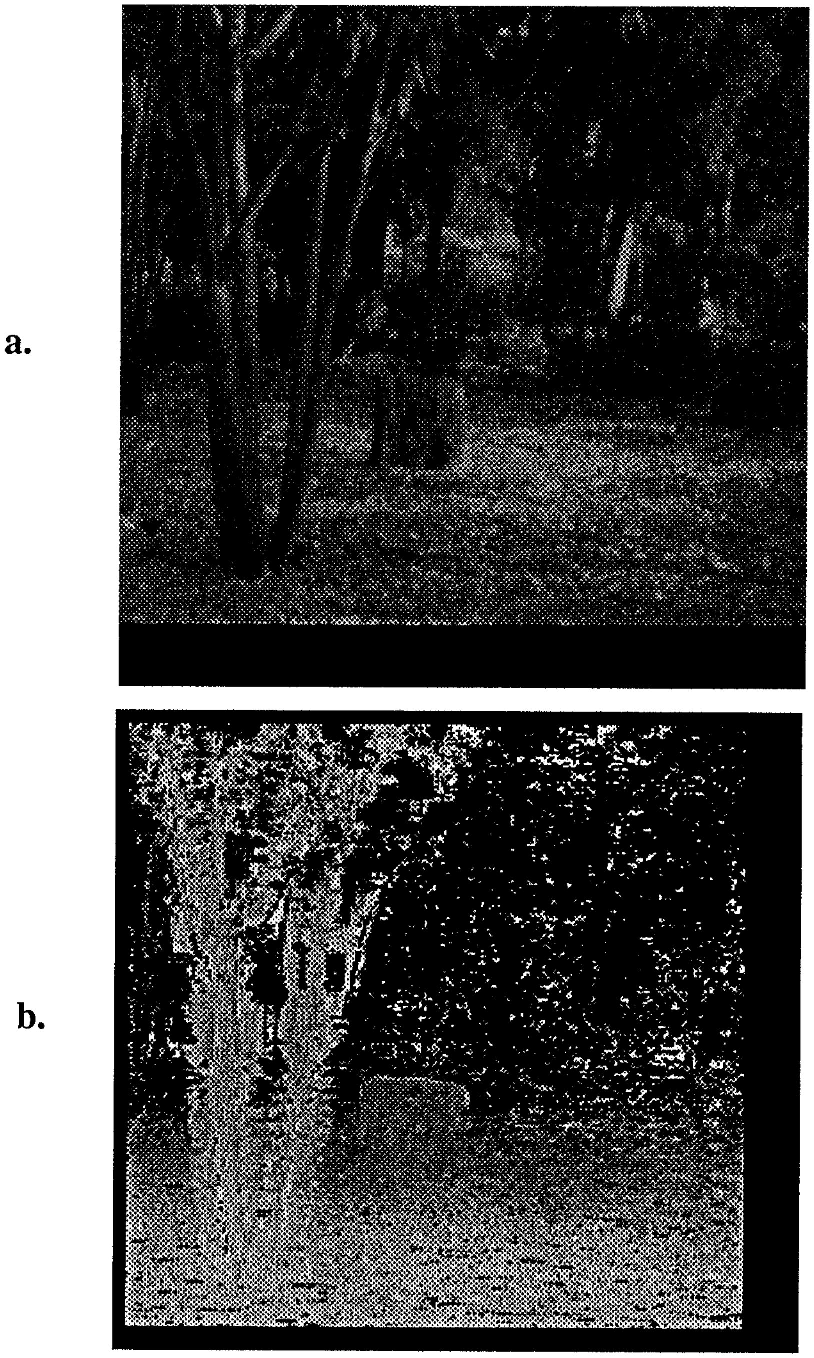

FIGURE 14. Tree Sequence:
a) The original image of the sequence at frame 3.
b) The optical flow extracted from LS algorithm encoded in intensity values.

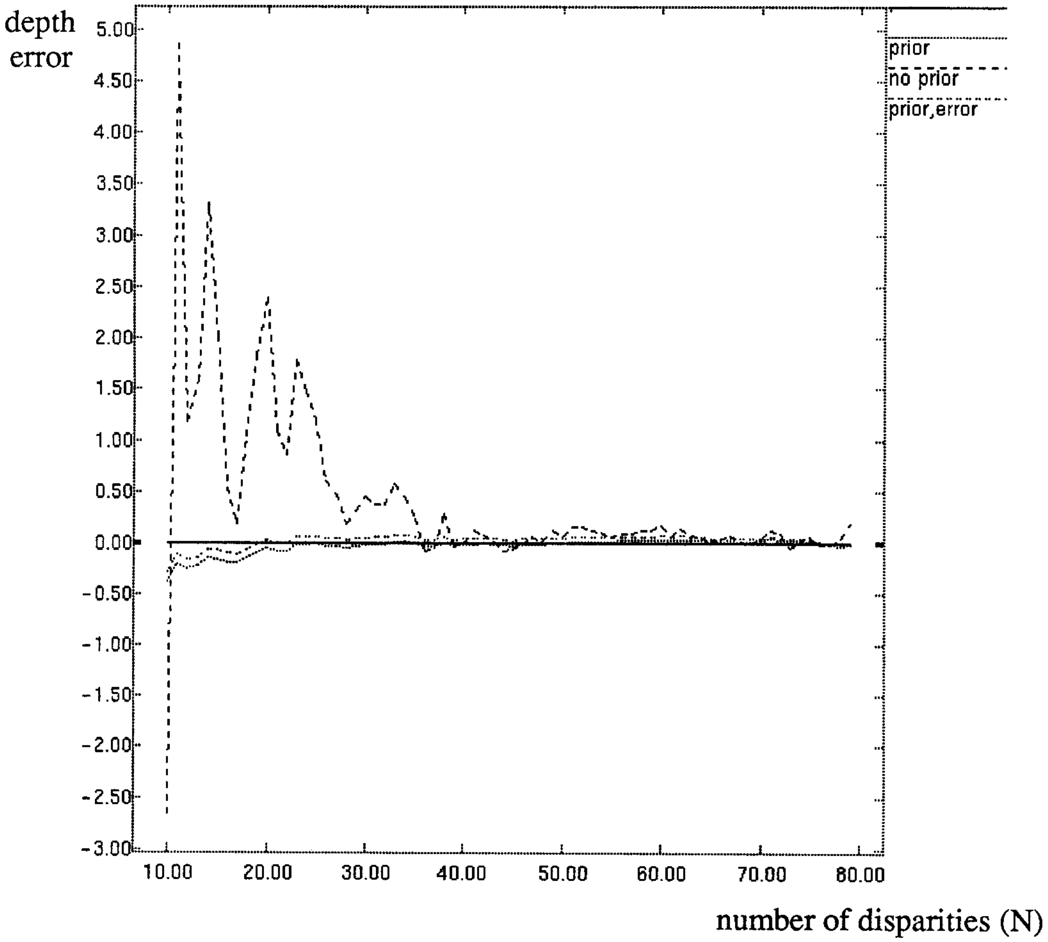

FIGURE 15. The estimation error of depth vs. the number of disparities used in a synthetic data experiment. Without using any prior information, the estimate is less accurate than that from an estimation procedure which uses prior information. The estimate, however, degrades when the prior information is corrupted with uncertainties.

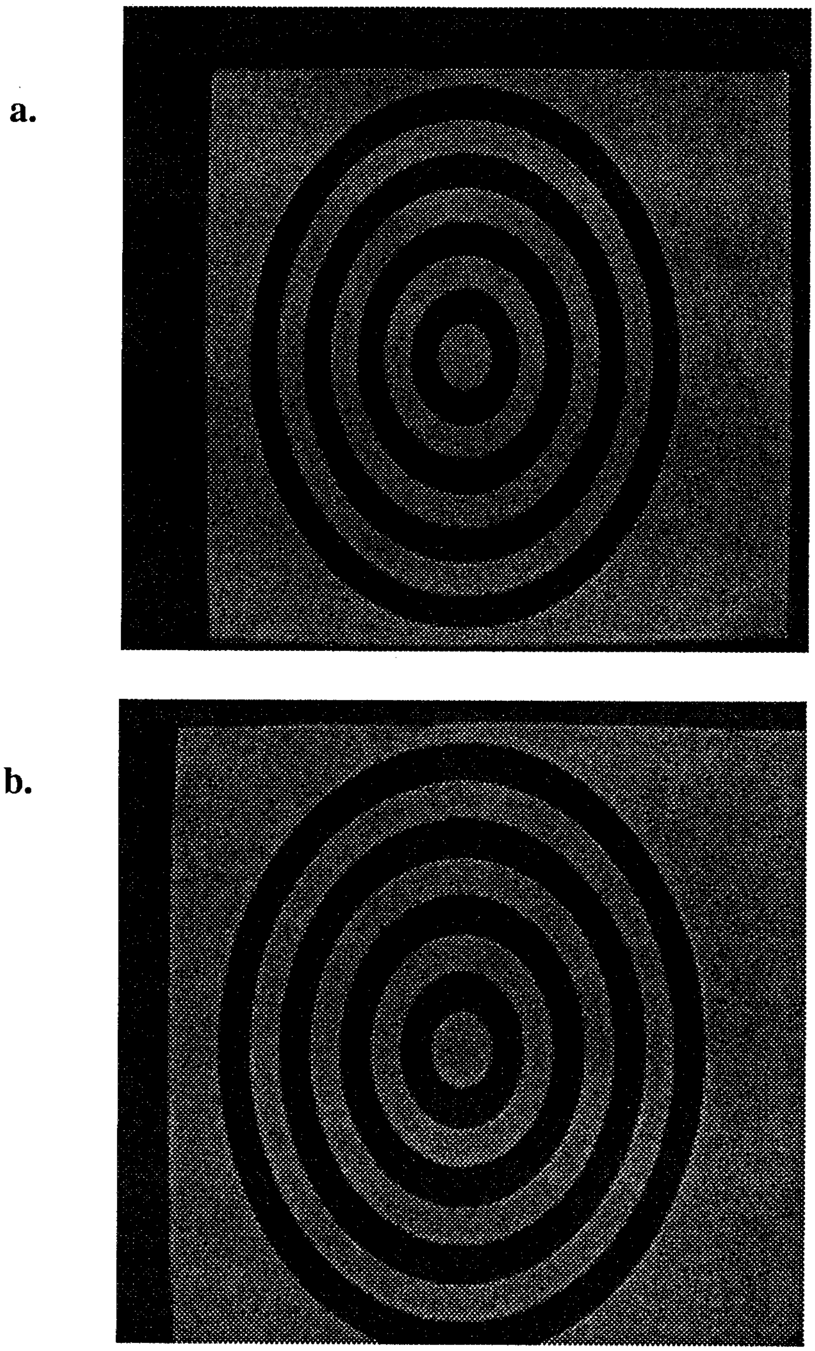

FIGURE 16. Bull's eye: The first (a) and the last (b) frame of the forward looking camera sequence - Bull's eye pattern. The pattern is flat and is perpendicular to the velocity vector of the camera.

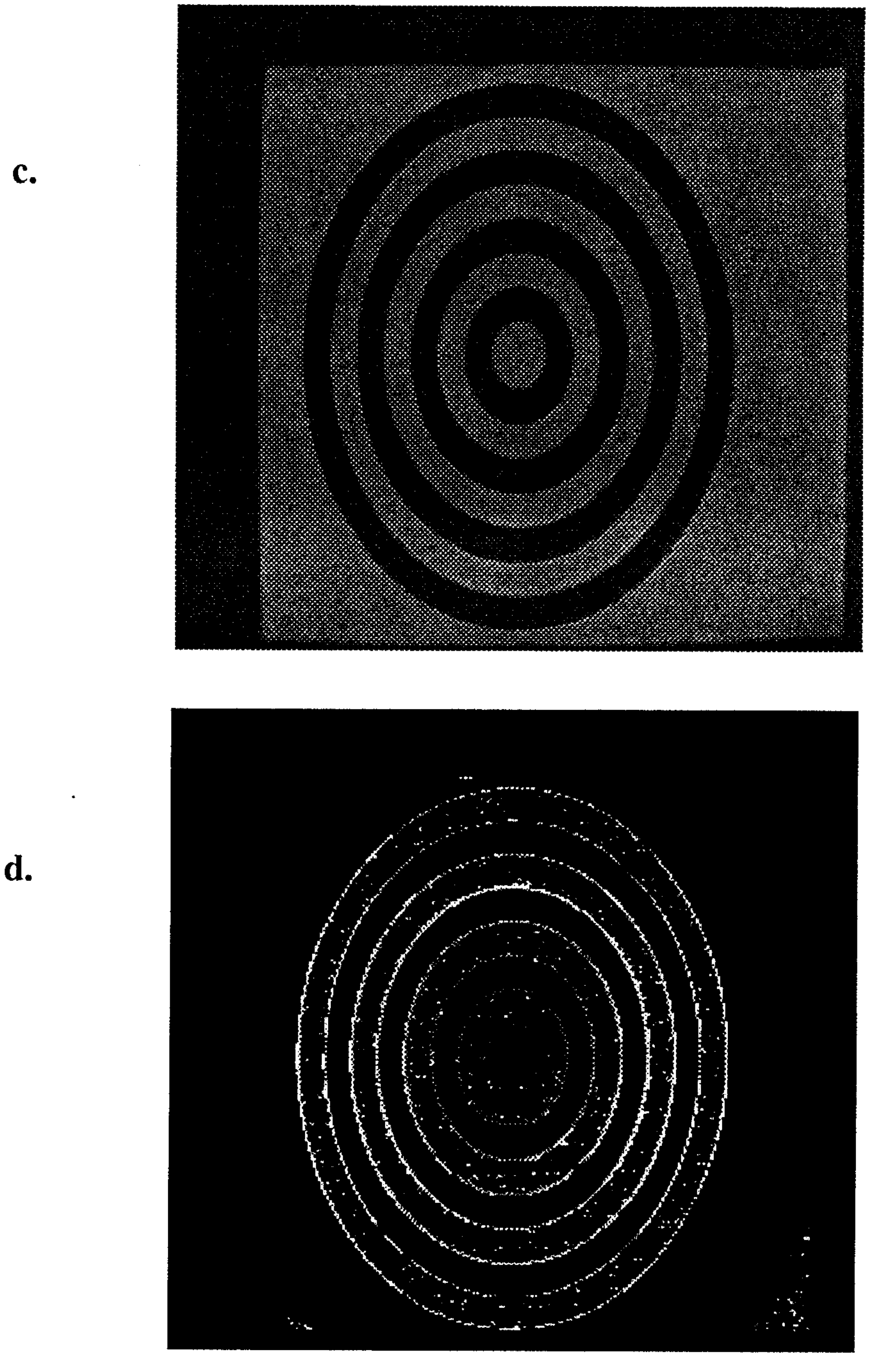

FIGURE 16. Bull's eye (con't):
c) The third frame of the sequence.
d) The magnitude of the optical flow is intensity-encoded. Note that the optical flow estimate is getting smaller (dimmer) when close to the FOE.

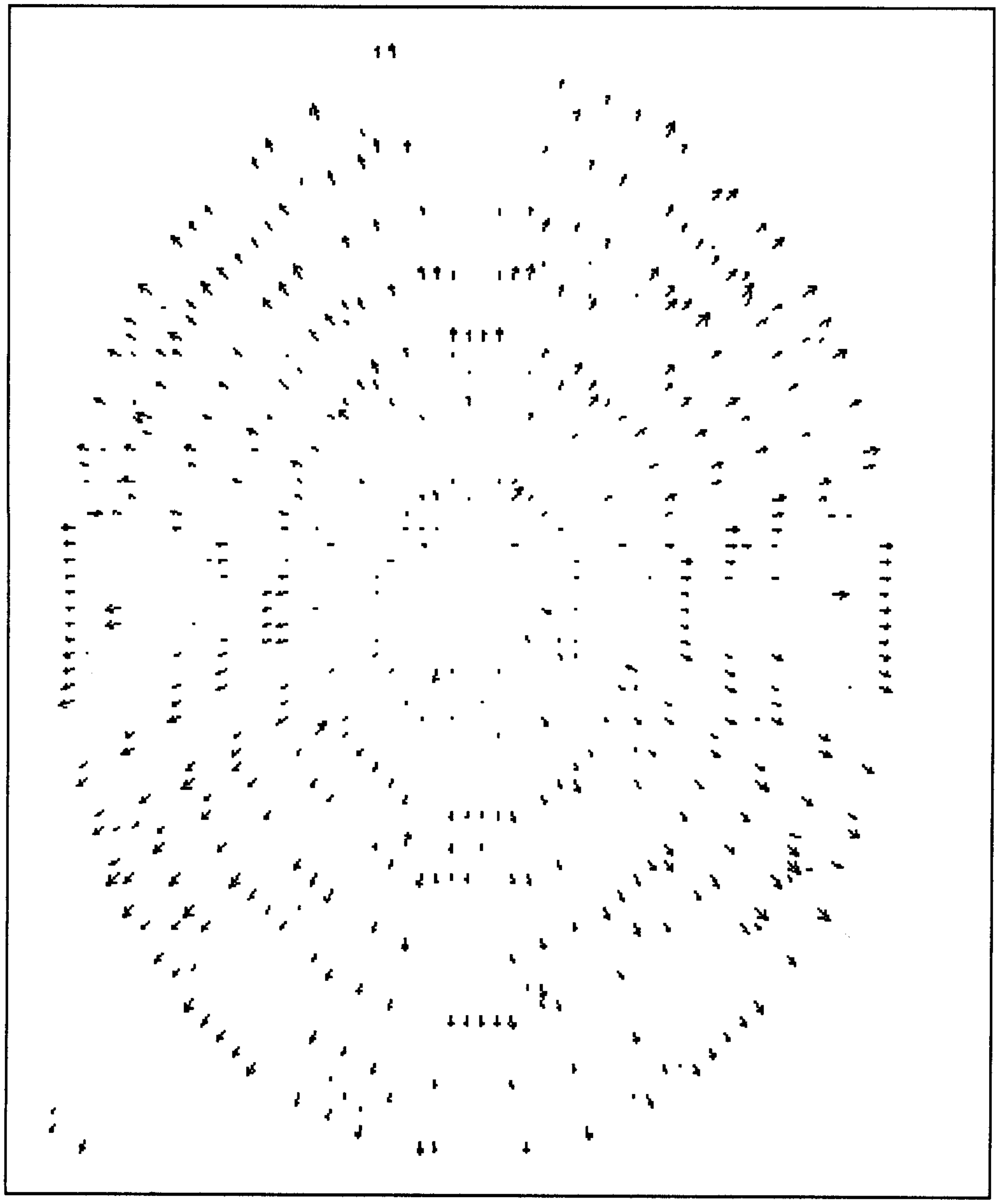

FIGURE 17. Bull's eye (con't): e) The 2D optical flow field evaluated at frame 3.

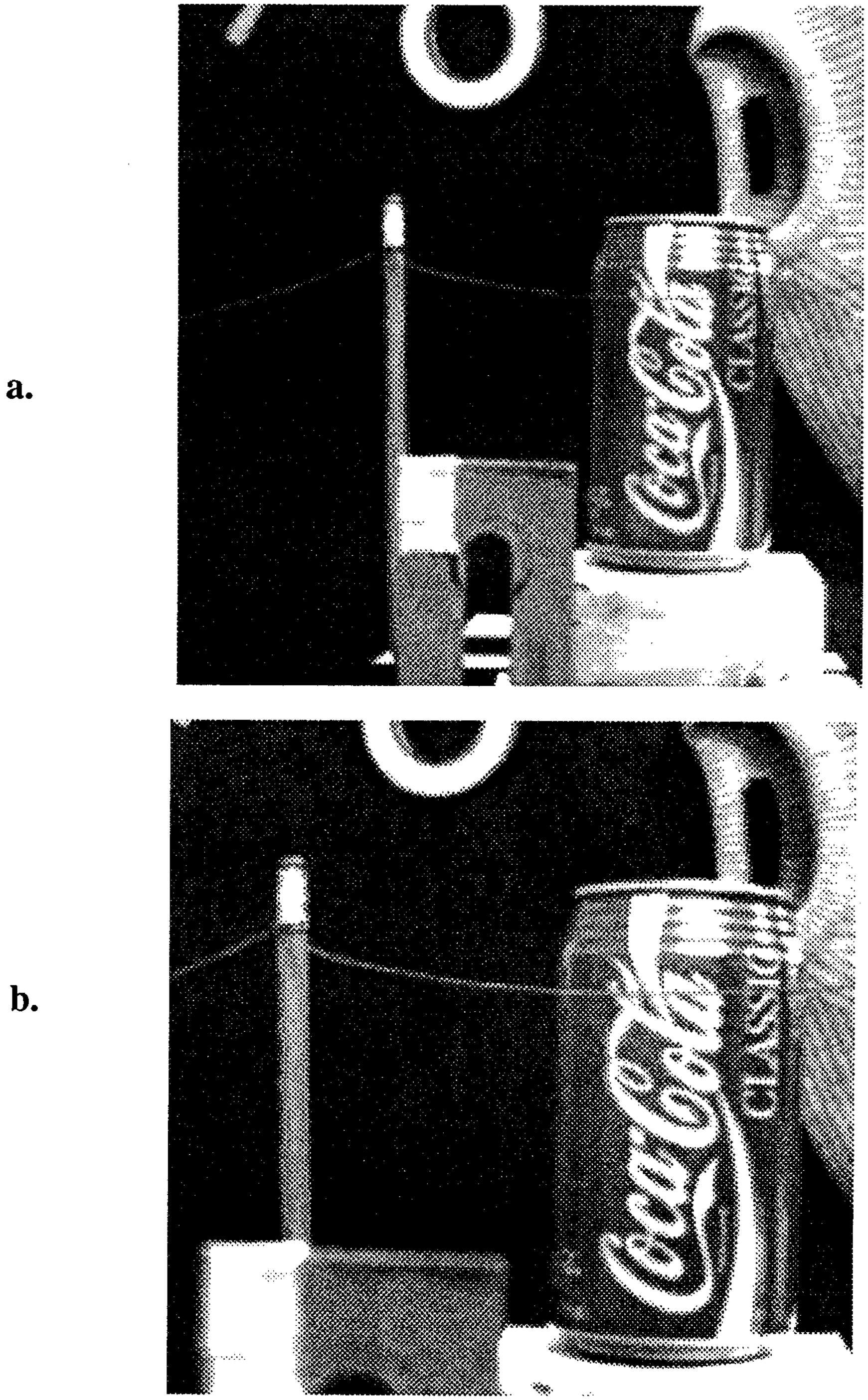

FIGURE 17. Forward looking camera from NASA: a) The first frame of the forward looking camera sequence which consists of 151 frames. b) The last frame of the sequence.

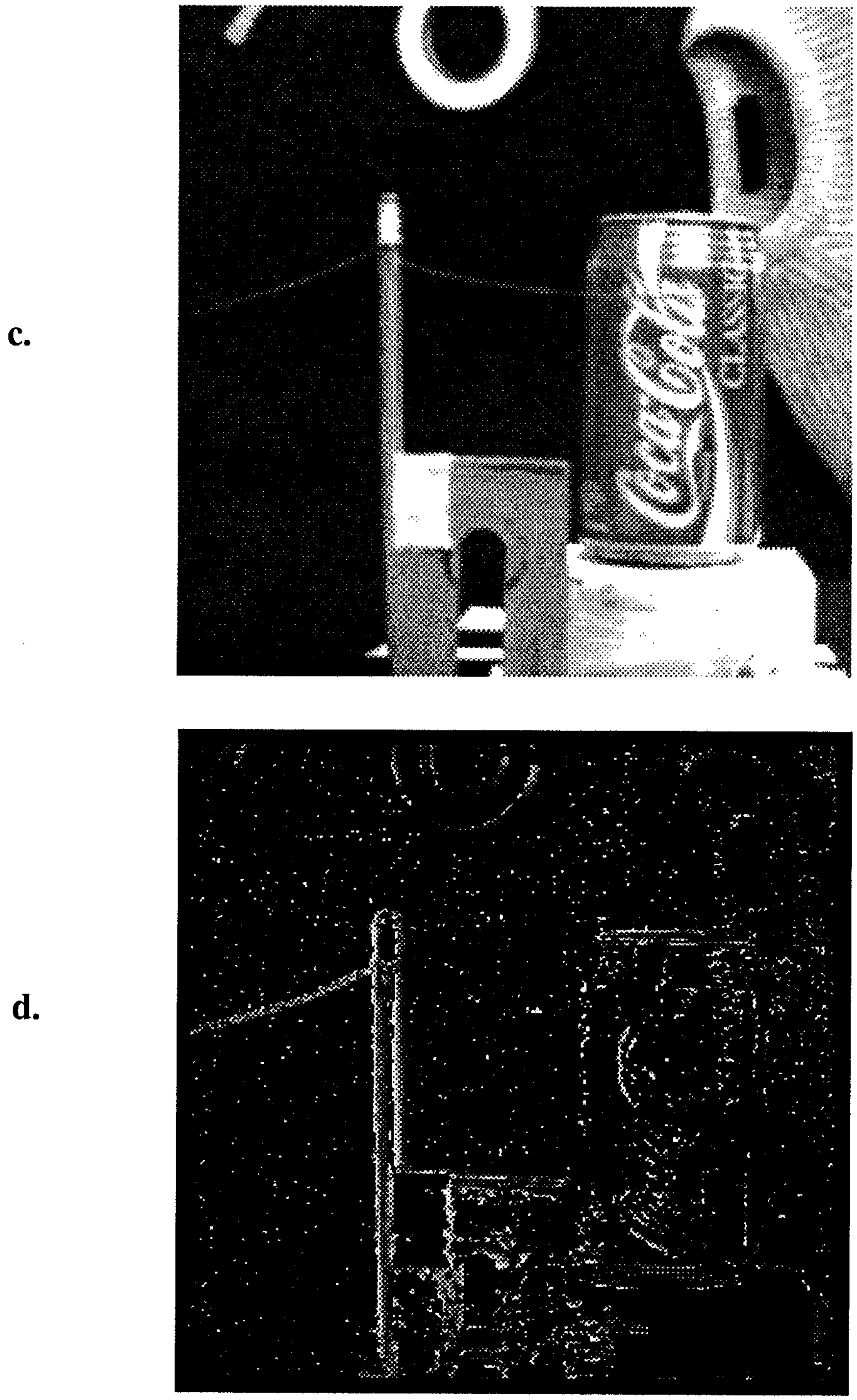

FIGURE 17. NASA Sequence (con't):
c) The third frame of the sequence.
d) The scaled depth map of the scene at the instance of frame 3. Scaled depth has been encoded in intensity for display purpose. Brighter points represent smaller depth values from the camera frame and vice versa.

e.

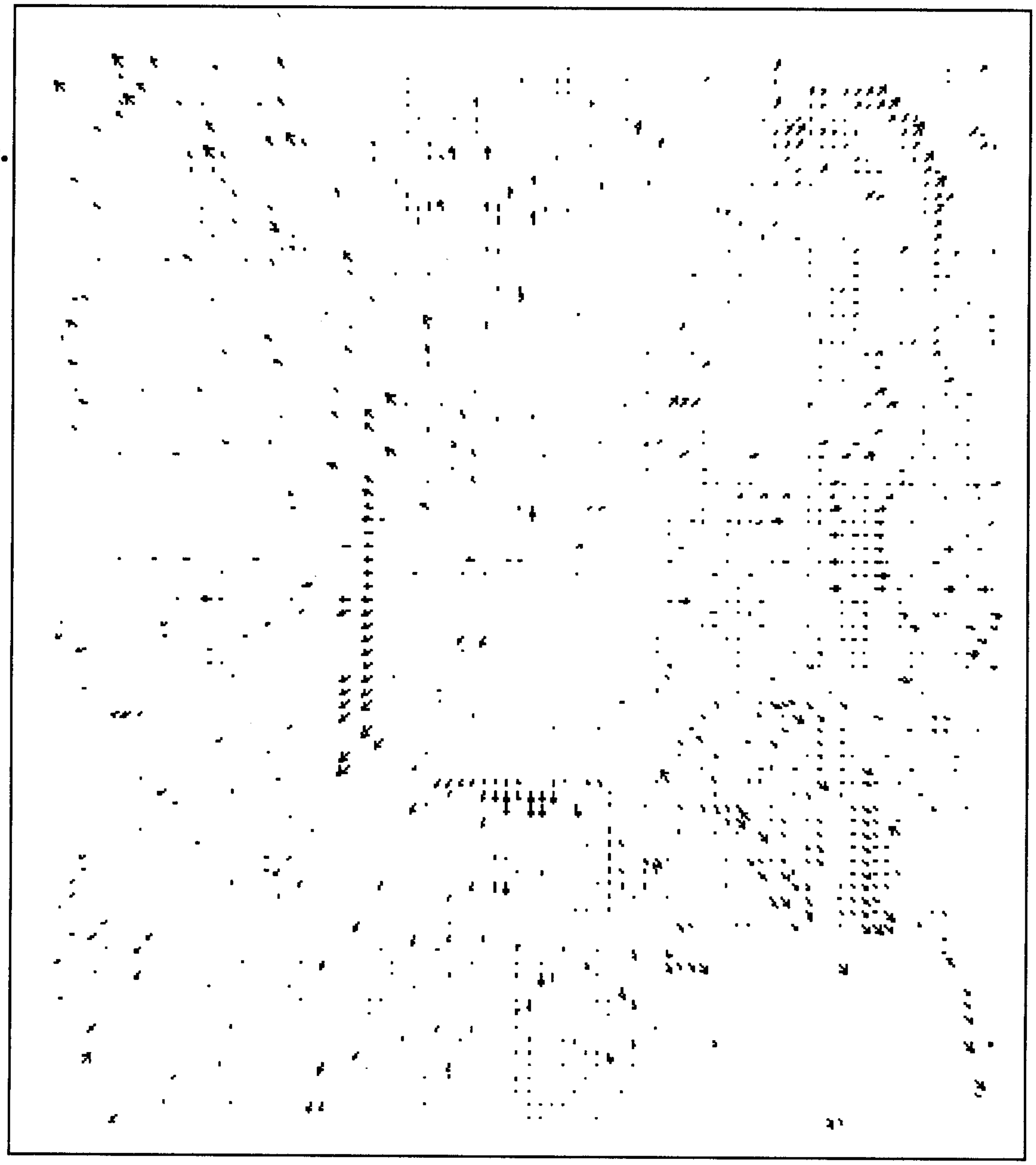

FIGURE 17. NASA Sequence (con't): e) The optical flow field at frame 3 displayed in 2D vector format.

6 Remarks

Under the LS framework, we have shown how to combine multiple noisy disparity measurements (from other correlation modules) to obtain the optimal depth, time-to-contact, and optical flow using a feature point trajectory model. We use different techniques that depend upon the availability of internal camera parameters and camera motion information. If no motion is known a priori, the LS estimation gives estimates of optical flow and time-to-contact (or scaled depth). If one component of the velocity is known, absolute depth can be calculated directly from time-to-contact without knowing any internal camera parameters. On the other hand, if complete camera motion knowledge is available through some independent inertial sensors along with internal camera parameters, the LS estimation algorithm is reformulated to directly calculate absolute depth, bypassing the calculation of optical flow. The framework is applicable to general camera translation and rotation (with the aid of the derotation module described in [32]) and can be implemented in a massively parallel manner at each pixel. Internal camera parameters can be estimated through camera calibration procedures such as the one described in [25].

We model the trajectory of feature points with optical flow and scaled depth being the model parameters. The LS framework minimizes the discrepancy between the disparity measurements and the model by searching for the appropriate parameters in the parameter space using the LS minimization method. The LS framework performs without knowing any motion a priori but reduces the dimension of the parameter space if partial motion is known in some special cases.

In this report, LS minimization is done in a batch mode. A sufficient number of disparity observations is recorded before invoking the procedure. In contrast, the procedure can also be implemented in a sequential/recursive setting. In that case, the objective function defined in (13) is minimized recursively whenever a new disparity observation arrives.

Currently, the LS algorithm combines all the disparity measurements with equal weight without taking into account the uncertainties associated with each individual measurement. In the future, the LS algorithm can be modified to weight reliable data more heavily than less reliable data in an intelligent fashion, e.g. by considering the confidence measure of each disparity measurement in the minimization procedure (weighted LS approach).

Acknowledgments

The authors would like to thank Dr. H. Harlyn Baker of SRI for contributing the sequence of outdoor images, Dr. Banavar Sridhar of NASA Ames for contributing the dense forward looking imagery and John Horst of NIST for preparing the 3D plot in this paper.

References

[1] Longuet-Higgins, H. C. and Prazdny, K., "The Interpretation of a Moving Retinal Image", Proceedings of the Royal Society of London B, 208: p.385-397, 1980.

[2] Bruss, A. R. and Horn, B, K. P., "Passive Navigation", Computer Vision, Graphics, and Image Processing, 21: p.3-20, 1983.

[3] Waxman, A. M. and Wohn, K., "Image Flow Theory: A Framework for 3D Inference from Image Flow Kinematics:, International Journal of Robotics Research, 4: p.95-108, 1985.

[4] Horn, B. K. P., "Robot Vision", MIT Press, 1985.

[5] Snyder, M. A., "Uncertainty Analysis of Image Measurements", Proceedings of Image Understanding Workshop, L.A., California, pp.681-693, 87.

[6] Bolles, R. C., Baker, H. H. and Marimont D. H., "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion", International Journal of Computer Vision, 1, 7-55, 1987.

[7] Albus, J. S. and Hong, T. S., "Motion, Depth, and Image Flow", Proceedings of IEEE Robotics and Automation Conference, Cincinnati, Ohio, May, 1990.

[8] Bhanu, B, Roberts, B. and Ming J., "Inertial Navigation Sensor Integrated Motion Analysis", Proceedings of Image Understanding Workshop, California, pp747-763, May, 1989.

[9] Matthies, L., Szeliski R. and Kanade, T., "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences", Technical Memorandum CMU-RI-TR-88-1, Carnegie Mellon University, Pittsburgh, PA 15213, January 1988.

[10] Horn, B. K. P. and Schunck, B. G., "Determining Optical Flow", Artificial Intelligence, 17, pp.185-203, 1981.

[11] Anandan, P., "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion", Proceedings of First International Conference on Computer Vision, pp.219-230, 1987.

[12] Singh, A., "An Estimation-Theoretic Framework for Image-Flow Computation", Proceedings of DARPA Image Understanding Workshop, September, 1990.

[13] Nagel, H.- H., "Displacement Vectors Derived from Second-order Intensity Variations in Image Sequences", Computer, Vision, Graphics and Image Processing, vol 21, pp85-117, 1983.

[14] Nagel, H- H., "On the Estimation of Optical Flow: Relations between Different Approaches and Some New Results", Artificial Intelligence, vol 33, pp299-324, 1987.

[15] Broida, T. J. and Chellappa, R., "Estimation of Object Motion Parameters from Noisy Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp90-99, 1986.

[16] Sridhar, B., Suorsa, H., and Hussien, B., "Passive Range Estimation for Rotorcraft Low Altitude Flight", Annual technical report of NASA-Ames, TR-103897, 1990.

[17] Young, G. S. J. and Chellappa, R., "3D Motion Estimation Using a Sequence of Noisy Stereo Images: Model, Estimation, and Uniqueness Results", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 12, pp735-759, 1990.

[18] Heeger, D. J. and Jepson, A., "Simple Method for Computing 3D Motion and Depth", Proceedings of Third International Conference on Pattern Recognition, pp96-100, 1990.

[19] Wong, R. Y. and Hall, E. L., "Sequential Hierarchical Scene Matching", IEEE Transactions on Computers, vol 27, 1978.

[20] Sethi, S. K. and Jain, R., "Finding Trajectories of Feature Points in a Monocular Image Sequence", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 9, no 1, pp56-73, 1987.

[21] Barnard, S. T. and Thompson, W. B., "Disparity Analysis of Images", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 2, pp333-340, 1980.

[22] Kearney, J. K., Thompson, W. B., and Boley, D. L., "Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 9, no 2, pp229-244, 1987.

[23] Heel, J., "Direct Dynamic Motion Vision", Proceedings of IEEE Conference on Robotics and Automation, Cincinnati, 1990.

[24] Okutomi, M. and Kanade, T., "A Multiple-Baseline Stereo", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp63-69, Maui, Hawaii, June, 1991.

[25] Tsai, R. Y., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, vol RA-3, no 4, pp323-344, August, 1987.

[26] Marr, D., "Vision", W. H. Freeman, San Francisco, 1982.

[27] Shahraray, B. and Brown, M. K., "Robust Depth Estimation from Optical Flow", Proceeding of Second International Conference on Computer Vision, pp641-650, 1988.

[28] Burger, W. and Bhanu, B., "Estimating 3D Egomotion from Perspective Image Sequences", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 12, no 11, November, 1990.

[29] Kanade, T. and Okutomi, M., "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment", Technical Memorandum CMU-CS-90-120, Carnegie Mellon University, Pittsburgh, PA 15213, April 1990.

[30] Chandrashekhar, S. and Chellappa, R., "Passive Navigation in a Partially Known Environment", Proceedings of IEEE Workshop on Visual Motion, New Jersey, pp2-7, 1991.

[31] Bolles, R. C., and Baker, H. H., "Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequences", Proceedings of the Third Workshop on Computer Vision: Representation and Control, Bellaire, MI, pp168-178, 1985.

[32] Lau, H., "Image Sequence Derotation by Warping", in preparation.

[33] Rangachar, R., Hong, T., and Herman, M., "Real-Time Differential Range Estimation Based on Time-space Imagery Using PIPE", Proceedings of SPIE Symposium on Real-time Image Processing II: Algorithms, Architectures, and Applications, vol. 1295, Orlando, FL, April 1990.

[34] Lau, H., Hong, T., and Herman, M., "A Hierarchical Model for the Determination of Optical Flow", in preparation.

[35] Nishihara, H. K., "Practical Real-time Imaging Stereo Matcher", Optical Engineering, vol. 23, no. 5, September, 1984.

[36] Subbarao, M., "Bounds on Time-to-Collision and Rotational Component from First-Order Derivatives of Image Flow", Computer Vision, Graphics, and Image Processing, 50, pp329-341, 1990.

[37] Raviv, D., "Invariants in Visual Motion", NIST Technical Report, NISTIR 4722, November, 1991.

[38] Nelson, R., and Aloimonos, J., "Using Flow Field Divergence for Obstacle Avoidance: Towards Qualitative Vision", Proceedings of Second International Conference on Computer Vision", Tampa, Florida, pp.188-196, 1988.

[39] Aloimonos, J., "Purposive and Qualitative Active Vision", Proceedings of DARPA Image Understanding Workshop, September, 1990.

[40] Raviv, D. and Herman, M., "Visual Servoing for Robot Vehicles Using Relevant 2D Image Cues", in preparation.

[41] Lucas, B. D. and Kanade, T., "An Iterative Image Registration Technique with An Application to Stereo Vision", Proceedings of DARPA Image Understanding Workshop, pp121-130, 1981.

[42] Simoncelli, E. P., Adelson, E. H., and Heeger, D. J., "Probability Distributions of Optical Flow", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp310-315, Maui, Hawaii, June, 1991.

NIST-114A
(REV. 3-90)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION OR REPORT NUMBER
NISTIR 4919

2. PERFORMING ORGANIZATION REPORT NUMBER

3. PUBLICATION DATE
SEPTEMBER 1992

4. TITLE AND SUBTITLE

Optimal Estimation of Optical Flow, Time-to-contact and Depth

5. AUTHOR(S)

Herbert Lau, Tsai-Hong Hong, and Martin Herman

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

10. SUPPLEMENTARY NOTES

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

Extraction of the optical flow field, time-to-contact[1] and depth from an image sequence is a vitally important problem in many contexts. Many robotic applications, e.g., obstacle avoidance, autonomous vehicles and the like, can benefit from this information. In this report, the problem is formulated under a least squares minimization framework. Based only upon a set of *spatiotemporal disparity*[2] measurements from a correlation algorithm, the algorithm in general recovers optimal estimates, in the sense of least squares, of both the *instantaneous* optical flow field and time-to-contact simultaneously. It does not utilize knowledge of camera translation or any internal camera parameters. Nevertheless, additional information can be recovered if prior information is incorporated into the algorithm. It is shown that if complete translational camera motion as well as some internal camera parameters are known, optimal estimates of *absolute depth* can directly be inferred from the disparity measurements. Closed-form solutions are obtained in all cases. The algorithm is simple, effective, and easy to implement, which enhances the possibility of hardware realization for real-time applications. Experimental results, both synthetic and real, are included to show the performance of the algorithm and to verify its robustness against random noise perturbation.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

closed-formed solutions, least squares estimation, optical flow, optimal estimates, spatiotemporal disparities, time-to-contact

13. AVAILABILITY

☒ UNLIMITED
☐ FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).
☐ ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.
☒ ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES
46

15. PRICE
A03

ELECTRONIC FORM