ABSTRACT

Title of Thesis: A Parallelogram Four-Frame Model
in 3-D Motion Analysis and Structure Recovery
Using Unified Optical Flow Field Approach

Zhichun Lin, Master of Science in Electrical Engineering, 1991
Department of Electrical and Computer Engineering

Thesis directed by: Dr. Nirwan Ansari
Assistant Professor
Department of Electrical and Computer Engineering

In this paper, a new concept of the unified optical flow field (UOFF) for 3-D motion analysis from a stereo image sequence is implemented, which is an extension of the fundamental temporal optical flow formulations developed by Horn and Schunck. A parallelogram four-frame model is established and a spatial optical flow is also introduced.

In order to alleviate the problem of the strict requirement of two identical sensors in common stereo imagery, a new imaging system is presented in this paper that needs only one camera, two plane mirrors and two switches to generate a phase shifted binocular sequence of images. It is shown that the structure and motion of the object surface can be reconstructed from the parallelogram four-frame image sequence set if the information of the settings of the experiment, including the focal length, the distance between two mirrors and the intersecting angle between two observing directions are known. Detailed computer simulations are presented and analyzed to illustrate the algorithm discussed.

A Parallelogram Four-Frame Model in 3-D Motion and Structure Recovery Using Unified Optical Flow Field Approach

by

Zhichun Lin

Thesis submitted to the Faculty of the Graduate School of the New Jersey Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering

1991

APPROVAL SHEET

Title of Thesis: A Parallelogram Four-Frame Model
in 3-D Motion Analysis and Structure Recovery
Using Unified Optical Flow Field Approach

Candidate: Zhichun Lin
Master of Science in Electrical Engineering, 1991

Thesis and Abstract Approved by the Examining Committee:

Dr. Nirwan Ansari, Advisor Assistant Professor Department of Electrical and Computer Engineering	Date
Dr. Edwin S.H. Hou Assistant Professor Department of Electrical and Computer Engineering	Date
Dr. Y.Q. Shi Assistant Professor Department of Electrical and Computer Engineering	Date
Dr. C.Q. Shu Research Associate Department of Electrical and Computer Engineering	Date

New Jersey Institute of Technology, Newark, New Jersey.

VITA

Zhichun Lin

Education

1990-1991	New Jersey Institute of Technology	MSEE
1985-1989	Shanghai Jiao Tong University	BSEE

Dedicated to

My Dear Parents

&

My Lovely Sister

ACKNOWLEDGEMENT

I am very grateful to my graduate advisor Dr. Nirwan Ansari whose inspirations and guidance benefited me significantly; without his support this work could not have been finished.

I would also like to express my special thanks to Dr. C. Q. Shu, Dr. Edwin Hou and Dr. Y. Q. Shi. In addition to providing me of the essential formulations of motion reconstruction using U.O.F.F, Dr. Shu helped me a lot from the very beginning of this thesis and was always willing to answer my questions. Dr. Hou helped me find out a bug in my computer program. Dr. Y. Q. Shi provided me with some very helpful papers on Optical Flow. Without their generous support, I cannot possibly complete the thesis.

I would also like to thank my friends and classmates at NJIT, who have given me their assistance and encouragement in many ways.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Research on estimation of surface structure and motion of objects from image sequences has come to play a dominant role within the computer vision community over the last decade. There are basically two different categories of approaches to recover the surface structure of objects and the relative motion between objects and observers: the optical flow field approach[1],[2] and the feature correspondence approach [3],[4].

It is generally agreed that the feature extraction and correspondence approach is difficult, and so far only partial solutions suitable for simplistic situations have been developed[5]. Although some of the current techniques are classified as the multiple optical flow field approach to stereo imagery[6],[7], since they still need to implement the correspondence between features in the pair of stereo images, they are essentially some hybrid of the optical flow approach and the feature correspondence approach. Therefore, they inevitably inherit the similar problem in feature correspondence approach.

In this paper, a new concept of the unified optical flow field (UOFF), originally proposed by Shu and Shi[2], for motion analysis from a stereo image sequence is implemented. It is an extension of the fundamental optical flow formulations developed by Horn and Schunck[1]. Instead of the original *two-frame*

model for *temporal* optical flow computation, a *four-frame model* is established and a *spatial* optical flow is also introduced[2].

In stereo imagery, two cameras are usually needed during the imaging. In order to make the derivation easier, most approaches proposed require the optical characteristics of the two cameras to be identical. However, in practice, this requirement is somehow too strict to be achieved. This makes these approaches difficult to be realized. To overcome the obstacle brought by this requirement, we present here a new imaging system which needs only one camera, two plane mirrors and two switches to generate a phase shifted binocular sequence of images. The four consecutive images of the sequence form a parallelogram four-frame model.

To recover the structure and motion of the object surface, an image interpolation technique employing optical flow is applied to the model to produce a pair of stereo images. In addition, two assumptions are made. One is the invariance of object brightness not only for the time variation but also for the space variation. The other is the smoothness constraint which assumes that neighboring points on the objects have similar velocities and the velocity field of the brightness patterns in the image varies smoothly almost everywhere. More detailed mathematical derivation of the above assumptions will be given later in following chapters.

In Chapter 2, we will review an iterative algorithm to determine *temporal* optical flow. In Chapter 3, the unified optical flow field is introduced. The *four-frame experimental model* is set up. It will be shown that the structure and motion of the object surface can be reconstructed from the four-frame image sequence set if the information of the settings of the experiment including the focal length, the distance between two mirrors and the intersecting angle

between two observing directions are known. In Chapter 4, detailed simulations are presented and analyzed to illustrate the algorithm discussed in Chapter 2 and 3. Conclusion along with discussion is drawn in Chapter 5.

Chapter 2

Optical Flow

2.1 Introduction

Optical flow is the distribution of apparent velocities of movement of brightness pattern in an image. Since optical flow can arise from relative motion of the objects and the viewers, it can provide important information about the spatial arrangement of the objects viewed, and the rate of change of this arrangement.

In order to reconstruct the 3-D spatial and temporal information of an observed object from its optical flow vectors, a good algorithm is needed to determine the optical flow of an observed object by the two frames of image captured by a camera at two different instants. Horn and Schunck[1] proposed a method in early 80's to determine the optical flow by the brightness distribution of all the pixels in the image plane. Since then, this method has been very frequently cited.

2.2 Determine Optical Flow

To simplify the computation, it is necessary to assume a restricted problem domain. To avoid variations in brightness due to shading effects, it is initially

assumed that the surface being imaged is flat. It is further assumed that the incident illumination is uniform across the surface, and reflectance varies smoothly and has no spatial discontinuities. The latter condition assures that the image brightness is differentiable. Situations where objects occlude one another partially are excluded because discontinuities in reflectance are found at object boundaries.

In the simple situation described above, the motions of the brightness patterns in the image are determined directly by the motions of the corresponding points on the surface of the object. Computing the velocities of points on the object is a matter of simple geometry once the optical flow is known.

There are two basic constraints: One is that the brightness of a particular point in the pattern is constant, i.e.,

$$\frac{dE}{dt} = 0, \tag{2.1}$$

where E is the brightness function $E(x, y, t)$ as a function of the spatial variables x, y, and the temporal variable t.

Using the chain rule for differentiation it can be seen that

$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0. \tag{2.2}$$

Let

$$u = \frac{dx}{dt} \quad and \quad v = \frac{dy}{dt}. \tag{2.3}$$

The following single linear equation with two unknowns u and v is thus obtained,

$$E_x u + E_y v + E_t = 0. \tag{2.4}$$

The other constraint is the smoothness constraint: Neighboring points on the objects have similar velocities, and the velocity field of the brightness patterns in

the image varies smoothly almost everywhere. One way to express the additional constraint is to minimize the sum of the squares of the Laplacians of the x- and y-components of the flow. The Laplacians of u and v are defined by

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad and \quad \nabla^2 v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}. \tag{2.5}$$

Here the square of the magnitude of the gradient will be used as a smoothness measure. For convenience, assume that the images are sampled on a square grid at regular intervals. Let the measured brightness be $E_{i,j,k}$ at the intersection of the i th row and j th column in the k th image frame. Note that this introduces quantization errors.

The differentiation operation is approximated by,

$$\begin{aligned} E_x \approx & \frac{1}{4}\left\{E_{i,j+1,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i+1,j,k}\right\} \\ & + \frac{1}{4}\left\{E_{i,j+1,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i+1,j,k+1}\right\}, \end{aligned} \tag{2.6}$$

$$\begin{aligned} E_y \approx & \frac{1}{4}\left\{E_{i+1,j,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i,j+1,k}\right\} \\ & + \frac{1}{4}\left\{E_{i+1,j,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i,j+1,k+1}\right\}, \end{aligned} \tag{2.7}$$

$$\begin{aligned} E_t \approx & \frac{1}{4}\left\{E_{i,j,k+1} - E_{i,j,k} + E_{i+1,j,k+1} - E_{i+1,j,k}\right\} \\ & + \frac{1}{4}\left\{E_{i,j+1,k+1} - E_{i,j+1,k} + E_{i+1,j+1,k+1} - E_{i+1,j+1,k}\right\}. \end{aligned} \tag{2.8}$$

Figure 2.1 shows the relationship in space and time between these measurements. Here the unit of length is the grid spacing interval in each image frame and the unit of time is the image frame sampling period.

The approximations of the Laplacians of u and v take the following forms

$$\nabla^2 u \approx K(\bar{u}_{i,j,k} - u_{i,j,k}) \quad and \quad \nabla^2 v \approx K(\bar{v}_{i,j,k} - v_{i,j,k}), \tag{2.9}$$

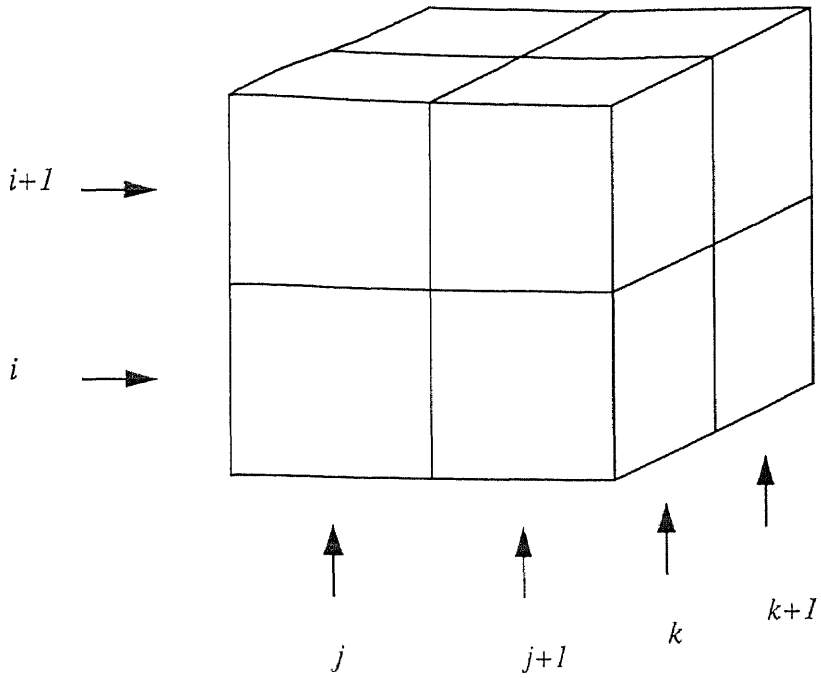

Figure 2.1: The three partial derivatives of image brightness at the center of the cube are each estimated from the average of first differences along four parallel edges of the cube. Here the column index j corresponds to the x direction in the image, the row index i to the y direction, while k lies in the time direction.

where the local averages $\bar{u}$ and $\bar{v}$ are defined as follows[8],

$$\bar{u}_{i,j,k} = \frac{1}{6}\{u_{i-1,j,k} + u_{i,j+1,k} + u_{i+1,j,k} + u_{i,j-1,k}\}$$

$$+ \frac{1}{12}\{u_{i-1,j-1,k} + u_{i-1,j+1,k} + u_{i+1,j+1,k} + u_{i+1,j-1,k}\}, \tag{2.10}$$

$$\bar{v}_{i,j,k} = \frac{1}{6}\{v_{i-1,j,k} + v_{i,j+1,k} + v_{i+1,j,k} + v_{i,j-1,k}\}$$

$$+ \frac{1}{12}\{v_{i-1,j-1,k} + v_{i-1,j+1,k} + v_{i+1,j+1,k} + v_{i+1,j-1,k}\}. \tag{2.11}$$

The proportionality factor K equals 3 if the average is computed as shown, and we again assume that the unit of length equals the grid spacing interval.

The total error to be minimized is

$$\epsilon_0^2 = \int\int (\alpha_0^2 {\epsilon_c}^2 + {\epsilon_b}^2)\, dx\, dy. \tag{2.12}$$

where ϵ_b, the sum of the errors in the equation for the rate of change of image brightness is equal to

$$\epsilon_b = E_x u + E_y v + E_t, \tag{2.13}$$

and ϵ_c, the measure of the departure from smoothness in the velocity flow, is equal to

$${\epsilon_c}^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2. \tag{2.14}$$

Here, α_0^2 is a weighing factor empirically chosen according to the magnitude of the expected noise. Using the calculus of variation to minimize Equation (2.12), we obtain

$$\begin{aligned} {E_x}^2 u + E_x E_y v &= \alpha_0^2 \nabla^2 u - E_x E_t. \\ E_x E_y u + {E_y}^2 v &= \alpha_0^2 \nabla^2 v - E_y E_t. \end{aligned} \tag{2.15}$$

Introducing the Laplacian approximations into the above equations and letting $\alpha = 3\alpha_0$,

$$(\alpha^2 + {E_x}^2)u + E_x E_y v = (\alpha^2 \bar{u} - E_x E_t),$$

$$E_x E_y u + (\alpha^2 + {E_y}^2)v = (\alpha^2 \bar{v} - E_y E_t). \tag{2.16}$$

Solving for u and v we find that

$$(\alpha^2 + {E_x}^2 + {E_y}^2)u = +(\alpha^2 + {E_y}^2)\bar{u} - E_x E_y \bar{v} - E_x E_t,$$

$$(\alpha^2 + {E_x}^2 + {E_y}^2)v = -E_x E_y \bar{u} + (\alpha^2 + {E_x}^2)\bar{v} - E_y E_t. \tag{2.17}$$

These equations can be written in the alternate form

$$(\alpha^2 + {E_x}^2 + {E_y}^2)(u - \bar{u}) = -E_x \left[E_x \bar{u} + E_y \bar{v} + E_t\right],$$

$$(\alpha^2 + {E_x}^2 + {E_y}^2)(v - \bar{v}) = -E_y \left[E_x \bar{u} + E_y \bar{v} + E_t\right]. \tag{2.18}$$

Now we have a pair of equations for each point in the image. If we solve them simultaneously by one of the standard methods such as Gauss-Jordan elimination, it would be very costly because the corresponding matrix is sparse and very large. Thus we choose the Gauss-Seidel method as an iterative approach[9]. Now we can compute a new set of velocity estimates (u^{n+1}, v^{n+1}) from the estimated derivatives and the average of the previous velocity estimates (u^n, v^n) by

$$u^{n+1} = \bar{u}^n - E_x \left[E_x \bar{u}^n + E_y \bar{v}^n + E_t\right] / (\alpha^2 + {E_x}^2 + {E_y}^2),$$

$$v^{n+1} = \bar{v}^n - E_y \left[E_x \bar{u}^n + E_y \bar{v}^n + E_t\right] / (\alpha^2 + {E_x}^2 + {E_y}^2). \tag{2.19}$$

At the edge of the image, some of the points needed to compute the local average of velocity lie outside the image. Here we simply copy velocities from adjacent points further in.

2.3 Experimental Results

To examine the accuracy of the algorithm suggested in the above section experimentally, some computer simulation have been done. First, an image setting

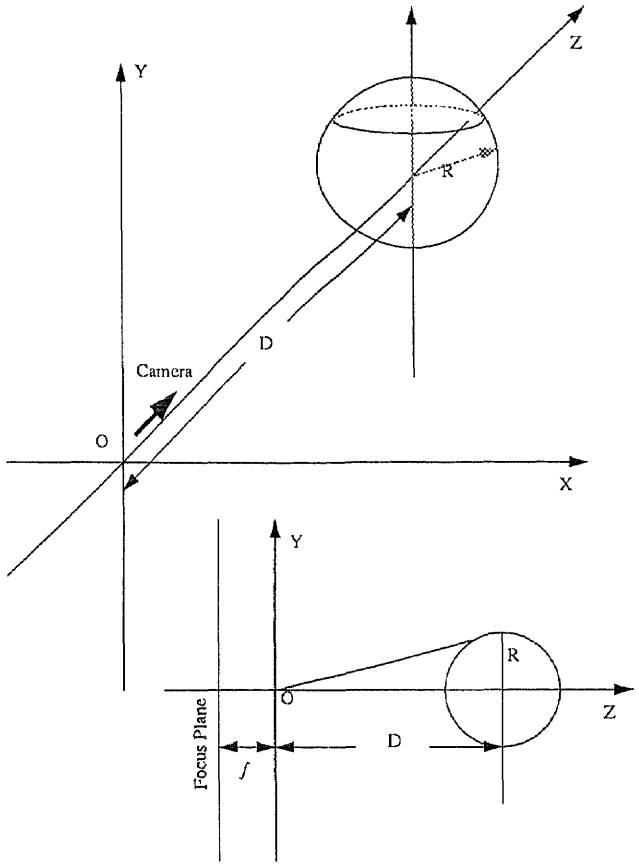

Figure 2.2: An image setting in the experiment to compute the *temporal* optical flow distribution on the surface of a rotating sphere.

is given as shown in Figure 2.2. We assume that the camera is located upon the origin, with its optical axis pointing to the positive-z direction. The focal length of the camera is denoted as f. The center of the rotating sphere is on the z-axis with a distance of D from the origin. Its radius is denoted as R.

In order to minimize the experimental noise and error, we do not perform the experiment *"physically,"* i.e., we do not *really* take pictures of a *real* rotating sphere by a *real* camera. Instead, a computer simulation is made. Given all the information about the image setting, including D, f, R, and the brightness function of the sphere surface, we can determine the corresponding image patterns which are supposed to be taken by cameras. So, in this experiment, we take two frames of image of the observed sphere, one at time T and the other one at time $T + \Delta T$, after some motion of the sphere. Using the algorithm given in Section 2.2, we get the iterative results of the optical flow for each pixel of the image. Furthermore, knowing all the spatial information of the experiment layout, we can easily obtain the optical flow distribution on the sphere surface theoretically. This theoretical solution can be regarded as the actual solution. It will be convenient for us to determine the efficiency of Horn's method by comparing the two results. To visualize the results, we display the optical flows by *a needle diagram*, as shown in Figure 2.3. The optical flow vector of each pixel in the image is represented by a little arrow whose length is proportional to the magnitude of the optical flow of this pixel, and whose direction is the same as that of the optical flow.

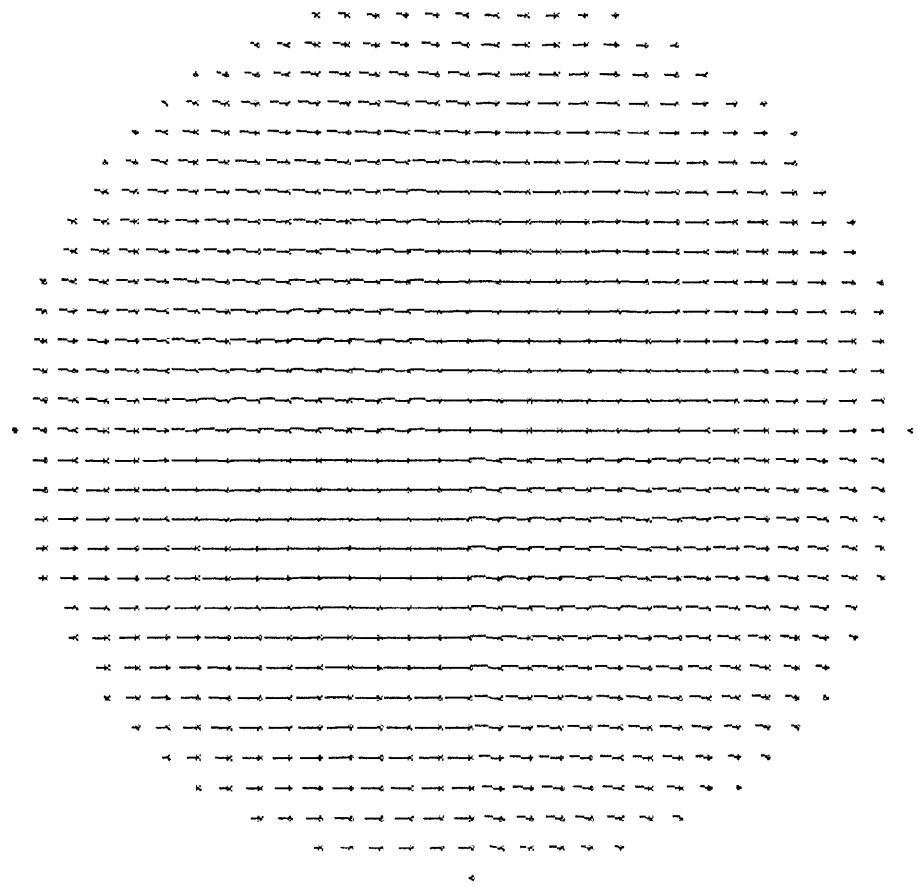

Figure 2.3: Needle diagram displaying optical flow distribution patterns

Chapter 3

Unified Optical Flow Field (UOFF) and 3-D Reconstruction

3.1 Introduction

In some sense, a camera taking pictures of an object is like a human eye observing an object. Just like one can hardly tell the distance of the object by only one eye, it is also impossible for us to reconstruct the 3-D spatial information with just one single view. Intuitively, we can guess that if we introduce one more camera into the imaging system, it would be possible to recover the 3-D information of the observed object, just like one can tell both the motion and distance of objects by viewing with two eyes. This intuition of *stereo imagery* leads to a novel concept of *Unified Optical Flow Field* proposed by Shu and Shi[2]. We can tell from the name that this is a generalized extension of the original optical flow method developed by Horn[1]. Instead of using the temporal optical flow field only to detect the $x - y$ motion of the object, we establish an additional spatial optical flow field to gain more information about the spatial location of each point in the space.

Usually, in stereo imagery, two identical cameras are utilized to take pairs of stereo images, thus forming a sequence of stereo images. However, in practice, it is very difficult, sometimes even impossible, to make the optical characteristics of two cameras to be identical. To overcome this difficulty, a new experimental arrangement is presented in this thesis. Instead of two identical cameras, only one camera is needed, along with two plane mirrors and two switches. A special phase shifted binocular images from the sequence form a parallelogram model. Before we proceed to the discussion on the parallelogram four-frame model, it is necessary to introduce a dynamic image interpolation algorithm in order to produce a pair of stereo images.

3.2 Dynamic Image Interpolation

A dynamic image interpolation technique is developed in this section. In Figure 3.1, images (m) and (n) are taken by a camera at t_m and t_n instants, respectively. The associated brightness functions are designated by $E(x, y, t_m)$ and $E(x, y, t_n)$, respectively, where (x, y) are used to represent the pixel coordinates on images (m) and (n).

Images (m) and (n) can be viewed as two images chosen from a monocular image sequence. Define

$$\Delta t_1 \triangleq t_n - t_m, \tag{3.1}$$

$$\Delta x_1 \triangleq x_n - x_m, \tag{3.2}$$

$$\Delta y_1 \triangleq y_n - y_m. \tag{3.3}$$

In Equations (3.2) - (3.3), (x_m, y_m) and (x_n, y_n) are representing the two pixels in images (m) and (n) such that they are related to the same world point in the 3-D space, respectively. It is assumed that the brightness of a particular point

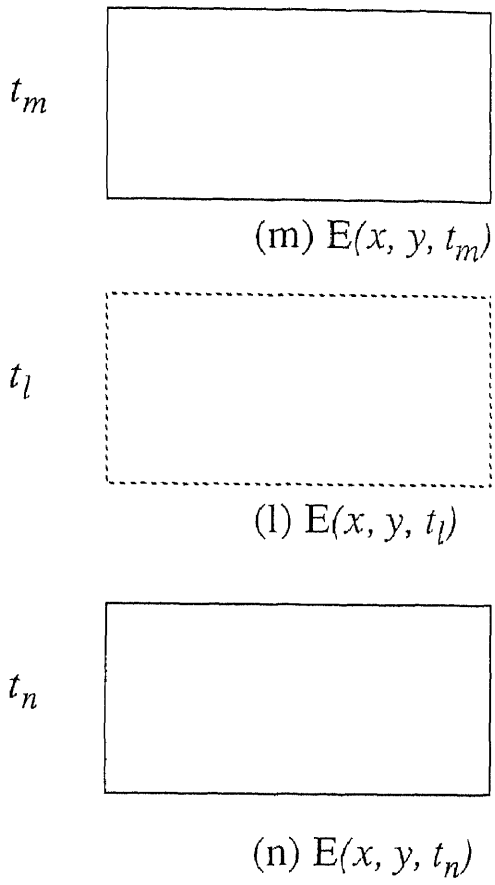

Figure 3.1: Dynamic Image Interpolation

in the pattern is constant, *i.e.*,

$$E(x_m, y_m, t_m) = E(x_n, y_n, t_n). \tag{3.4}$$

From Equation (3.1) - (3.3) it follows that

$$E(x_n, y_n, t_n) = E(x_m + \Delta x_1, y_m + \Delta y_1, t_m + \Delta t_1). \tag{3.5}$$

Expanding the right-hand side of the above equation in the Taylor series leads to

$$E(x_n, y_n, t_n) = E(x_m, y_m, t_m) + \frac{\partial E(x, y, t_m)}{\partial x}\Delta x_1 + \frac{\partial E(x, y, t_m)}{\partial y}\Delta y_1 + \frac{\partial E(x, y, t_m)}{\partial t}\Delta t_1 + \varepsilon \tag{3.6}$$

where ε represents second and higher order terms of Δx_1, Δy_1 and Δt_1. With $E_{mx} \triangleq \frac{\partial E(x,y,t_m)}{\partial x}$, $E_{my} \triangleq \frac{\partial E(x,y,t_m)}{\partial y}$ and $E_{mt} \triangleq \frac{\partial E(x,y,t_m)}{\partial t}$, the following relationship is obtained from Equations (3.4) and (3.6),

$$E_{mx}\Delta x_1 + E_{my}\Delta y_1 + E_{mt}\Delta t_1 + \varepsilon = 0 \tag{3.7}$$

Dividing both sides of Equation (3.7) by Δt_1, and let $\Delta t_1 \to 0$, we have

$$E_{mx}u + E_{my}v + E_{mt} = 0, \tag{3.8}$$

where

$$u \triangleq \lim_{\Delta t_1 \to 0} \frac{\Delta x_1}{\Delta t_1}, \tag{3.9}$$

$$v \triangleq \lim_{\Delta t_1 \to 0} \frac{\Delta y_1}{\Delta t_1}. \tag{3.10}$$

That is, u and v are the velocities of pixel (x, y) in image (m) moving along the x and y directions, respectively.

Under the smoothness constraint, following the derivation in [1] the next two equations are obtained

$$(E_{mx})^2 u + E_{mx}E_{my}v = \alpha\nabla^2 u - E_{mx}E_{mt}, \tag{3.11}$$

$$(E_{my})^2 v + E_{mx}E_{my}u = \alpha\nabla^2 v - E_{my}E_{mt}, \tag{3.12}$$

where α is a constant. It is noted that partial derivatives E_{mx}, E_{my} and E_{mt} can be determined by using image data $E(x, y, t_m)$ and $E(x, y, t_n)$[1]. Now having two equations, the two unknowns u and v can be solved by using the iterative algorithm developed in [1].

It is noted that there is a dashed line rectangle in Figure 3.1 denoted by image(l). It can be considered as a virtual image generated at t_l instant by the same camera which generates images (m) and (n). Its brightness pattern is represented by $E(x, y, t_l)$. It will be shown that image (l) can be estimated from images (m) and (n) by using an interpolation technique.

Again, the brightness invariance of two pixels, (x_l, y_l) and (x_m, y_m), located in images (m) and (l), respectively, relating to the same world point in the 3-D space is assumed, *i.e.*,

$$E(x_l, y_l, t_l) = E(x_m, y_m, t_m). \tag{3.13}$$

In the following discussion it is assumed that

$$t_l = (1 - \lambda)t_m + \lambda t_n, \qquad \lambda \in [0, 1]. \tag{3.14}$$

That is, t_l is a linear combination of t_m and t_n. When λ varies between 0 and 1 continuously, t_l will vary between t_m and t_n continuously. It can be seen that

$$t_l - t_m = \lambda\Delta t_1 \tag{3.15}$$

Define

$$\Delta x_{lm} \triangleq x_l - x_m, \tag{3.16}$$

$$\Delta y_{lm} \triangleq y_l - y_m. \tag{3.17}$$

Equation (3.13) becomes

$$E(x_l, y_l, t_l) = E(x_l - \Delta x_{lm}, y_l - \Delta y_{lm}, t_m). \tag{3.18}$$

With u and v defined in Equations (3.9) & (3.10), we obtain

$$E(x_l, y_l, t_l) \approx E(x_l - u\lambda\Delta t_1, y_l - v\lambda\Delta t_1, t_m). \tag{3.19}$$

Since u and v have been determined and Δt_1 is known, $E(x_l, y_l, t_l)$ can then be estimated via the use of Equation (3.19). Applying Equation (3.19) to pixels on image (l), pixel by pixel, results in an estimated image (l). The error between the interpolated image (l) and the "true" image, (the one taken at t_l instant if it is possible), depends on the time difference $t_l - t_m$, *i.e.*, $\lambda\Delta t_1$.

3.3 A Parallelogram Four-Frame Model

In this section a new imaging arrangement is presented. In Figure 3.2, the imaging geometry is shown. Here S is a camera, M_1 and M_2 are two plane mirrors with an angle β in between, SW_1 and SW_2 are two switches used to control the timing of taking images. The optical center of S is denoted by O. The optical axis of S is denoted by OZ. The OZ axis and the normal of M_1 are coplanar. In this thesis, the case in which OZ and the normals of M_1 and M_2 are coplanar is discussed. Axis OX is perpendicular to OZ such that OX, OZ and the normal of M_1 are coplanar. Axis OY is perpendicular to OXZ plane, and is not drawn in Figure 3.2.

From geometric optics, it is known that the light reflected by M_2 then by M_1 can arrive at camera S generating an image that is equivalent to the image "generated" by a "right" camera with the optical center and the optical axis being O^R and O^RZ^R, respectively, as shown in Figure 3.2. The angle between

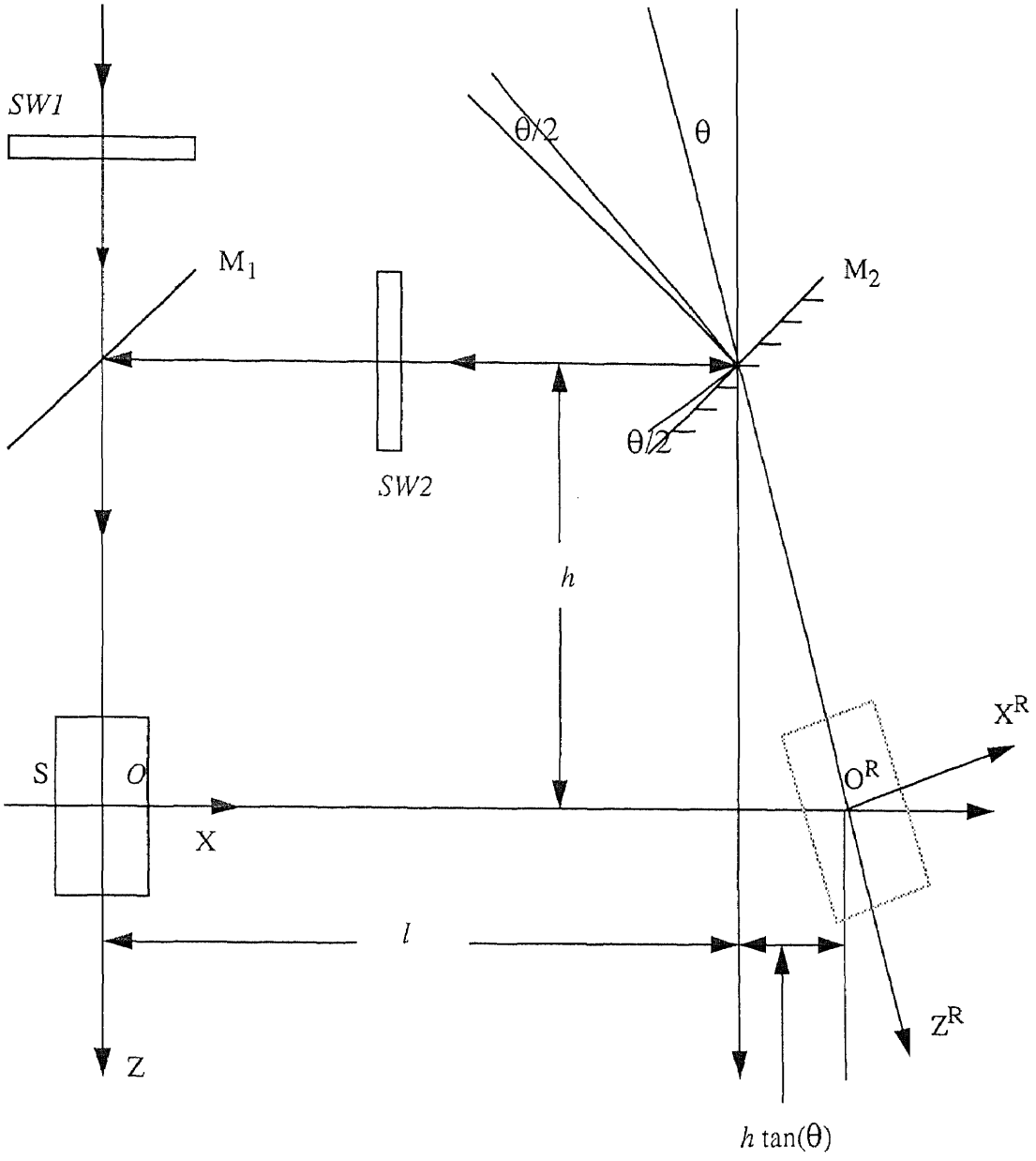

Figure 3.2: Imaging geometry using parallelogram four-frame model

two optical axes is 2β. Another Cartesian coordinate system is established with O^RX^R axis on the OXZ plane, and O^RY^R axis perpendicular to $O^RX^RZ^R$ plane. Again O^RY^R is not drawn in Figure 3.2. The whole system depicted in Figure 3.2 works in such a manner that the light transmitted through M_1, and the light reflected by M_2 and then by M_1 reach S alternately. Therefore, it generates a particular binocular image sequence, *i.e.*, a phase shifted binocular image sequence. A closer look is taken next at these four consecutive images in this binocular image sequence.

In Figure 3.3, a parallelogram four-frame model is shown. There t_1, t_2, t_3 and t_4 represent four different instants in time domain. In order to simplify the derivations, it is assumed that

$$t_3 - t_1 = t_4 - t_2. \tag{3.20}$$

Let

$$\Delta t \triangleq t_4 - t_2, \tag{3.21}$$

and hence

$$t_2 = t_1 + (1 - \lambda)\Delta t, \tag{3.22}$$

$$t_3 = (1 - \lambda)\Delta t_2 + \lambda t_4, \tag{3.23}$$

where $\lambda \in [0, 1]$. At t_1 and t_3 instants, the switches $SW1$ and $SW2$ are controlled so that only the light transmitted through plane mirror M_1 can arrive at camera S generating the brightness patterns $E^L(x^L, y^L, t_1)$ and $E^R(x^R, y^R, t_4)$, respectively. Therefore, images (a), (b), (c) and (d) taken by camera S at four different instants are equivalent to two images, (a) and (c), taken by the "left" camera at t_1 and t_3 instants, respectively, and two images, (b) and (d), obtained by the "right" camera at t_2 and t_4 instants, respectively. These four images are referred to as a parallelogram four-frame model.

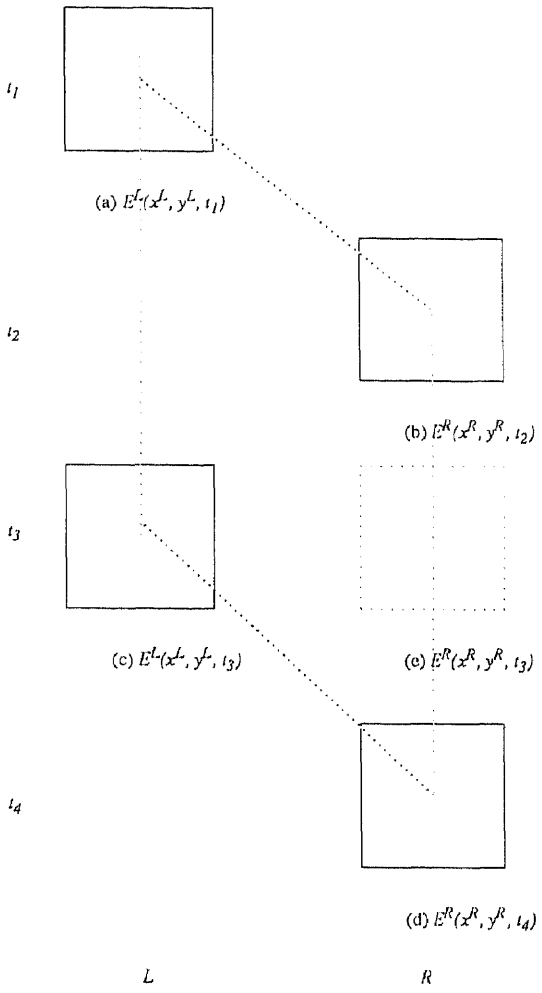

Figure 3.3: A parallelogram four-frame model

3.4 3-D Motion Estimation

Given the parallelogram four-frame model a method is developed to estimate 3-D motion using the unified optical flow field. This can be illustrated through the following six procedures.

Procedure 1. Images (a) and (c) can be viewed as two images chosen from a monocular image sequence generated by the "left" camera. Define

$$\Delta x^L = x_3^L - x_1^L, \tag{3.24}$$

$$\Delta y^L = y_3^L - y_1^L. \tag{3.25}$$

In Equations (3.24) - (3.25), (x_3^L, y_3^L) and (x_1^L, y_1^L) represent the two pixels in images (c) and (a), respectively, such that they are related to the same world point in 3-D space. It is assumed that the brightness of a particular point in the pattern is constant[1], *i.e.*,

$$E^L(x_3^L, y_3^L, t_3) = E^L(x_1^L, y_1^L, t_1). \tag{3.26}$$

From Equations (3.24) - (3.25), it follows that

$$E^L(x_1^L, y_1^L, t_1) = E^L(x_3^L - \Delta x^L, y_3^L - \Delta y^L, t_3 - \Delta t). \tag{3.27}$$

Expanding the right-hand side of the above equation in the Taylor series yields

$$E^L(x_1^L, y_1^L, t_1) = E^L(x_3^L, y_3^L, t_3) + \frac{\partial E^L(x^L, y^L, t_3)}{\partial x^L}(-\Delta x^L) + \frac{\partial E^L(x^L, y^L, t_3)}{\partial y^L}(-\Delta y^L) + \frac{\partial E^L(x^L, y^L, t_3)}{\partial t}(-\Delta t) + \epsilon, \tag{3.28}$$

where ϵ contains second and higher order terms in Δx^L, Δy^L and Δt. With $E_{3x}^L \triangleq \frac{\partial E^L(x^L, y^L, t_3)}{\partial x^L}$, $E_{3y}^L \triangleq \frac{\partial E^L(x^L, y^L, t_3)}{\partial y^L}$ and $E_{3t}^L \triangleq \frac{\partial E^L(x^L, y^L, t_3)}{\partial t}$, we obtained the following using Equations (3.26) - (3.28)

$$E_{3x}^L(-\Delta x^L) + E_{3y}^L(-\Delta y^L) + E_{3t}^L(-\Delta t) + \epsilon = 0. \tag{3.29}$$

Dividing both sides of Equation (3.29) by $-\Delta t$, and let $\Delta t \to 0$, we have

$$E_{3x}^L u_3^L + E_{3y}^L v_3^L + E_{3t}^L = 0, \tag{3.30}$$

where

$$u_3^L \triangleq \lim_{\Delta t \to 0} \frac{\Delta x^L}{\Delta t}, \tag{3.31}$$

$$v_3^L \triangleq \lim_{\Delta t \to 0} \frac{\Delta y^L}{\Delta t}. \tag{3.32}$$

This means that u_3^L and v_3^L are the velocities of pixel (x^L, y^L) in image (c) moving along the x and y directions, respectively.

Under the smoothness constraint, following the derivation in [1] the next two equations are obtained

$$(E_{3x})^2 u_3^L + E_{3x}E_{3y}v_3^L = \alpha_1 \nabla^2 u_3^L - E_{3x}E_{3t}, \tag{3.33}$$

$$(E_{3y})^2 v_3^L + E_{3x}E_{3y}u_3^L = \alpha_1 \nabla^2 v_3^L - E_{3y}E_{3t}, \tag{3.34}$$

where α_1 is a constant. The two unknowns u_3^L and v_3^L can thus be solved by using the iterative algorithm developed in [1].

Procedure 2. A close look at images (b), (d) and (e) in Figure 3.3 reveals that image (e), a rectangle drawn with dashed line in Figure 3.3, can be estimated from images (b) and (c), by using the dynamic image interpolation technique developed in Section 3.2. Images (m) and (n) in Section 3.2 correspond to images (b) and (d) here, respectively. The counterpart of estimated image (l) is image (e). That is

$$E^R(x^R, y^R, t_3) = E^R(x^R - u_2^R \lambda \Delta t, y^R - v_2^R \lambda \Delta t, t_2), \tag{3.35}$$

where the superscript R stands for the "right" camera. The subscript 2 represents the pixel velocities evaluated at t_2 instant. Image (e), *i.e.*, the brightness

pattern $E^R(x^R, y^R, t_3)$ can be considered as a virtual image generated by the "right" camera at t_3 instant. In other words, at this point, a pair of stereo images (c) and (e), that is, $E^L(x^L, y^L, t_3)$ and $E^R(x^R, y^R, t_3)$ are available.

Procedure 3. Similar to Procedure 1, applying Horn and Schunck's approach[1] to image (b) and estimated image (e), the following equation is derived

$$E^R_{3x} u^R_3 + E^R_{3y} v^R_3 + E_{3t} = 0, \tag{3.36}$$

where the superscript R and the subscripts 3, x, y, t have the same meanings as defined previously. With the smoothness constraint u^R_3 and v^R_3 can be solved by an iterative algorithm [1] derived from the following two equations,

$$(E^R_{3x})^2 u^R_3 + E^R_{3x} E^R_{3y} v^R_3 = \alpha_1 \nabla^2 u^R_3 - E_{3x} E_{3t}, \tag{3.37}$$

$$(E^R_{3y})^2 v^R_3 + E^R_{3x} E^R_{3y} u^R_3 = \alpha_1 \nabla^2 v^R_3 - E_{3y} E_{3t}. \tag{3.38}$$

In the above two equations, α_1 is a constant. It is observed that the time interval between t_2 (when image (b) is taken) and t_3 (when image (e) is supposed to be "generated") is $\lambda \Delta t$.

Procedure 4. Following the procedure presented in [2], the optical flow for a spatial sequence of two images is developed. The image sequences, from image (c) to image (e), can be viewed as "spatial" sequence of images - the time is "frozen" and the camera moves in the spatial domain. One way to describe the camera movement in space is fixing the left camera and moving the right camera. The displacement of the virtual right camera from the left camera is controlled by plane mirror M_2 as shown in Figure 2. A measure of the variation of the right camera position with respect to the left camera position is denoted by δs. It is defined by

$$\delta s \triangleq \sqrt{\tilde{x}^2 + \chi^2 \theta^2},$$

where $\tilde{x}$ is the x coordinate of the origin O^R of the coordinate system $O^R - X^R Y^R Z^R$ in the coordinate system $O - XYZ$, θ is the angle between OZ and $O^R Z^R$, and χ is a characteristic length according to imaging setting.

Define

$$\delta x \triangleq x^R - x^L, \qquad \delta y \triangleq y^R - y^L, \tag{3.39}$$

where (x^R, y^R) and (x^L, y^L) are coordinates of pixels in images (e) and (c) such that these two pixels are related to the same world point in 3-D space. Therefore, δx and δy are the horizontal and vertical coordinate differences of these two pixels, respectively.

Define

$$u^s \triangleq \lim_{\delta s \to 0} \frac{\delta x}{\delta s}, \tag{3.40}$$

$$v^s \triangleq \lim_{\delta s \to 0} \frac{\delta y}{\delta s}. \tag{3.41}$$

Thus, u^s and v^s are considered as the spatial variation rates of δx and δy with respect to δs, respectively. These two quantities generated from the spatial sequence of images can be viewed as the counterparts of u_3^L and v_3^L (or u_3^R and v_3^R) generated from the temporal sequence of images.

Similarly, the spatial optical flow field characterized by the following two equations[2] is established by using the brightness space-invariant assumption and the smoothness constraint:

$$(E_{3x}^L)^2 u^s + E_{3x}^L E_{3y}^L v^s = \alpha_2^2 \nabla^2 u^s - E_{3x}^L E_s^L, \tag{3.42}$$

$$(E_{3y}^L)^2 v^s + E_{3x}^L E_{3y}^L u^s = \alpha_2^2 \nabla^2 v^s - E_{3y}^L E_s^L, \tag{3.43}$$

where α_2 is a constant and $E_s^L \triangleq \frac{\partial E^L}{\partial s}$,

$$\frac{\partial E^L}{\partial s} \approx \frac{E^R(x^L, y^L, t_3) - E^L(x^L, y^L, t_3)}{\partial s}. \tag{3.44}$$

This is similar to the estimation of $\frac{\partial E}{\partial x}$ and $\frac{\partial E}{\partial y}$ in [1].

Procedure 5. According to [2], the unified optical flow field consisting of the six field quantities, u^R, v^R, u^L, v^L, u^s and v^s, is established. These field quantities are related by Equations (3.33) - (3.34), (3.37) - (3.38), (3.42) - (3.43). They form a set of equations for the unified optical flow field and are rewritten below.

$$(E_{3x}^L)^2 u_3^L + E_{3x}^L E_{3y}^L v_3^L = \alpha_1^2 \nabla^2 u_3^L - E_{3x}^L E_{3t}^L, \tag{3.45}$$

$$(E_{3y}^L)^2 v_3^L + E_{3x}^L E_{3y}^L u_3^L = \alpha_1^2 \nabla^2 v_3^L - E_{3y}^L E_{3t}^L, \tag{3.46}$$

$$(E_{3x}^R)^2 u_3^R + E_{3x}^R E_{3y}^R v_3^R = \alpha_1^2 \nabla^2 u_3^R - E_{3x}^R E_{3t}^R, \tag{3.47}$$

$$(E_{3y}^R)^2 v_3^R + E_{3x}^R E_{3y}^R u_3^R = \alpha_1^2 \nabla^2 v_3^R - E_{3y}^R E_{3t}^R, \tag{3.48}$$

$$(E_{3x}^L)^2 u^s + E_{3x}^L E_{3y}^L v^s = \alpha_2^2 \nabla^2 u^s - E_{3x}^L E_s^L, \tag{3.49}$$

$$(E_{3y}^L)^2 v^s + E_{3x}^L E_{3y}^L u^s = \alpha_2^2 \nabla^2 v^L - E_{3y}^L E_s^L, \tag{3.50}$$

It is noted that Equations (3.45) - (3.46) are a pair of equations from which u_3^L and v_3^L can be solved. Similarly, Equations (3.47) - (3.48) and Equations (3.49) - (3.50) are pairs associated with u_3^R and v_3^R, and u^s and v^s, respectively. Since the format of these pairs of equations are the same as the ones in [1], the iterative algorithm proposed in [1] can be employed here to determine (u_3^L, v_3^L), (u_3^R, v_3^R) and (u^s, v^s). It can be shown in the next procedure that these six field quantities contain sufficient information to recover the motion in 3-D space.

Procedure 6. In [2] a set of eight equations for analysis of 3-D motion field based on the unified optical flow field has been developed. In order to employ the result to estimate 3-D motion for the new imaging system discussed in this thesis, a close look is taken at the two imaging settings: that presented

in [2] and that discussed in this thesis. It can be seen that two settings are basically the same. The two optical axes, presented in [2] and in this thesis, are both coplanar. The center of the right plane mirror M_2 is restricted to translate along the OO^R axis. The M_2 can be rotated around the OY axis. In both cases characteristic length that can be chosen according to imaging setting is denoted by χ. The difference is that the distance between two lens center of the two cameras denoted by l in [2] becomes $l + h\ tan(2\beta)$ here.

It is therefore clear that all the eight equations and the remarks made in [2] are applicable here except that l is replaced by by $l + htan(2\beta)$.

In summary, the set of equations used for reconstructing the six 3-D motion parameters is listed below.

$$Z = -\frac{flsin(2\beta)}{u^s\sqrt{[l + h\ tan(2\beta)]^2 + 4\chi^2\beta^2 sin(2\beta) + 2f(1 - cos(2\beta))}}, \tag{3.51}$$

$$X = \frac{x^L Z}{f}, \tag{3.52}$$

$$Y = \frac{y^L Z}{f}, \tag{3.53}$$

$$x^R = u^s\sqrt{[l + h\ tan(2\beta)]^2 + 4\chi^2\beta^2} + x^L, \tag{3.54}$$

$$y^R = u^s\sqrt{[l + h\ tan(2\beta)]^2 + 4\chi^2\beta^2} + y^L, \tag{3.55}$$

$$\dot{Z} = \frac{\left(u^R - u^L cos(2\beta)\right) Z^2}{lf\ cos(2\beta)}, \tag{3.56}$$

$$\dot{X} = \frac{u^L Z}{f} + \frac{\dot{Z}X}{Z}, \tag{3.57}$$

$$\dot{Y} = \frac{v^L Z}{f} + \frac{\dot{Z}Y}{Z}, \tag{3.58}$$

It can be seen that in Equations (3.51) - (3.58) all quantities on the right-hand side are available because they are either the imaging setting parameters, the unified optical flow field quantities, or the 3-D coordinates and their derivatives

which have been solved by the preceding equations. Quantities x^R and y^R on the left-hand side of Equations (3.54) - (3.55) are used to determine $u^R(x^R, y^R, t_3)$ which should be utilized in Equation (3.56) to calculate $\dot{Z}$. All quantities on the left-hand side of Equations (3.51) - (3.53), (3.56) - (3.58), being the motion parameters: position and velocity in 3-D space, and therefore can be solved directly.

In Chapter 4, a detailed simulation will be given to verify the accuracy of this method.

Chapter 4

Simulation Experiments

4.1 Imaging Setting

A series of experimental simulations will be presented in this chapter to demonstrate the unified optical flow field theory discussed in Chapter 3. Figure 4.1 shows the imaging setting for the simulation. Note that this setting is similar to the one given in Chapter 2, as shown in Figure 2.2. The difference here is that "two" cameras are used in this experiment (a real one on the left side and a virtual one on the right side). The object to be observed is a sphere with radius R. Its center is along the OZ axis. The distance between the center of the sphere O_1 and the origin of the coordinate system O is denoted by D. Because of the scheme of the imaging system, the left and right cameras are identical in their optical characteristics, such as focal length f. The optical axis of the left camera is aligned with the OZ axis. The optical axis of the right camera is $O^R O_1$, which intersects OZ-axis with an angle θ. Note that the intersecting angle between two plane mirrors M_1 and M_2 is $\frac{\theta}{2}$. The simulated sphere surface can be represented by the following equation,

$$X^2 + Y^2 + (Z - D)^2 = R^2. \tag{4.1}$$

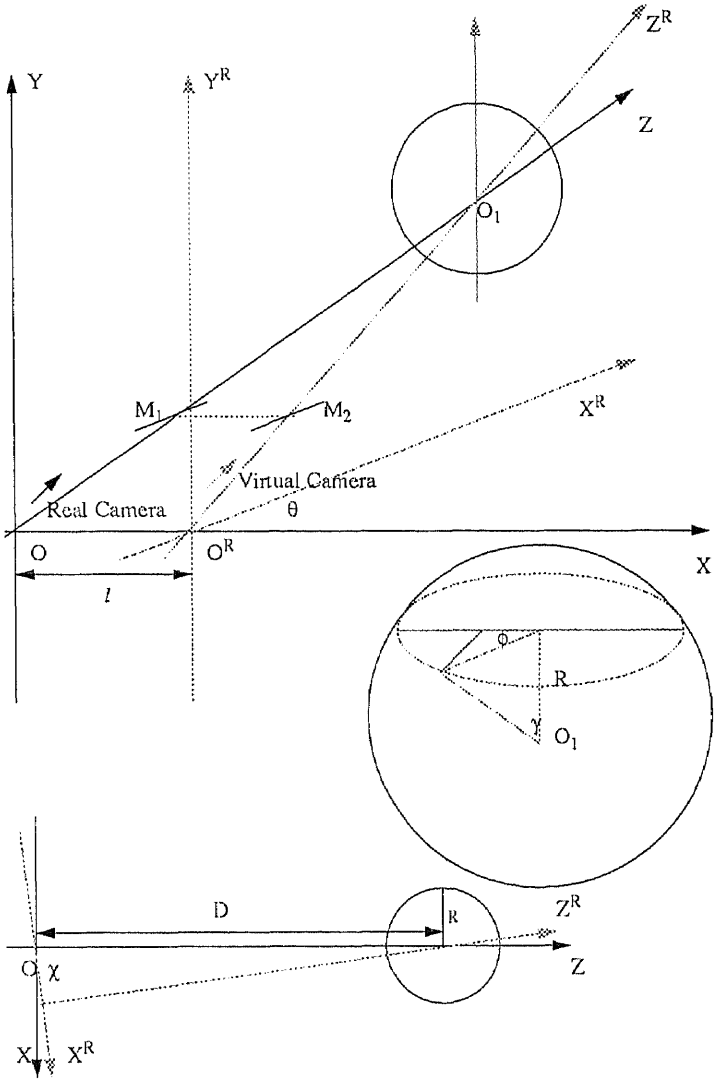

Figure 4.1: Simulated imaging geometry

The next step is to choose a brightness function for the surface of the sphere. It can be expressed by $E(X, Y, Z)$, where X, Y and Z are the coordinates of the point under observation. However, from Equation (4.1), we can see that variable Z can be treated as a function of variables X and Y. Therefore, we may let $E(X, Y, Z) = E(X, Y)$. Actually, every point on a plane can be uniquely determined by two independent variables. Here we choose a polar coordinate system in determining the position of the point on the sphere. Hence, we can express the brightness function as $E(X, Y) = E(\phi, \gamma)$, where ϕ and γ are two angles denoted in Figure 4.1. To make the simulation condition conformable with the assumptions and constraints of the method for determining optical flows, a sinusoidal function of angles ϕ and γ is chosen as the brightness function of the sphere surface.

$$E(\phi, \gamma) = K_1(1 + sin(K_2\phi))(1 + sin(K_3\gamma)), \tag{4.2}$$

where K_1, K_2 and K_3 are constants.

It can be seen that if $0 < K_1 < 64$, the value of the brightness function $E(\phi, \gamma)$ will range from 0 to 255. Furthermore, in order to avoid abrupt changes of brightness on the sphere surface, K_2 is chosen to be a integer such that an integer multiples of the period of $sin(K_2\theta)$ equals to 360°. This is also applicable in choosing the value of K_3.

4.2 Image Capturing

We need to obtain four frames of images of the sphere using the imaging setting as shown in Figure 3.1. Two methods are considered.

In the first method, we have to scan over the surface of the sphere, and project every point back onto the image frame; the brightness value of the point

on the sphere will be assigned to the corresponding pixel in the image frame. However, a scheme is needed to *scan* over the sphere. If the spatial intervals between sampled points are too big, some pixels may be *missed* in the projected image. On the other hand, if the intervals are too small, a severe *overlapping* of projections will arise; this is neither necessary nor efficient.

The other method is described below. For each pixel in the $SIZE \times SIZE$ image frame, a projection line is drawn from it to the sphere, through the origin O. If the line never intersects the sphere, the pixel is not a projection of any point on the sphere surface, and therefore it is taken as a background pixel. On the other hand, if it intersects, the brightness value of the point on the sphere surface where the projection line intersects is assigned to this pixel in the image. In this way, every pixel in the image can be explored . A detailed algorithm is given below.

Procedure $\{projection\}$

begin

$unit := \frac{2fR}{D \times SIZE}$;

for $i := 0$ to $SIZE - 1$ do begin

$i_1 := \left(i - \frac{SIZE}{2} + \frac{1}{2}\right) unit$;

for $j := 0$ to $SIZE - 1$ do begin

$j_1 := \left(j - \frac{SIZE}{2} + \frac{1}{2}\right) unit$;

$delta := D^2(i_1^2 + j_1^2)^2 - (f^2 + i_1^2 + j_1^2)\left(D^2\left(i_1^2 + j_1^2\right) - R^2 f^2\right)$;

if $delta >= 0$ then begin

$Z := \frac{D\left(i_1^2 + j_1^2\right) + \sqrt{delta}}{f^2}$;

$X := -\frac{Z\, i_1}{f}$;

$Y := -\frac{Z\, j_1}{f}$;

```
                phi := atan(X/(D-Z));
                gamma := acos(Y/R);
                image[i][j] := Brightness_function(phi, gamma);
            end
            else begin
                image[i][j] := BACKGROUND;
            end
        end
    end
end    {projection}
```

In the above algorithm, the *Brightness_function* varies for different images. Assume that the sphere is spinning with uniform angular speed around its own axis which is parallel to Y axis. It is clockwise if it is observed from the top. For example, if the sphere rotates from instant T to instant $T + \Delta T$ for an angle of ψ, and if the brightness function detected by the left camera at instant T is $f(\phi, \gamma)$, then the brightness function detected by the right camera at instant $T + \Delta T$ will be $f((\phi - \psi + \theta), \gamma)$. Similarly, at instant $T + 2\Delta T$, the brightness function for the left camera will be $f((\phi - 2\psi), \gamma)$, and at instant $T + 3\Delta T$, the brightness fuction for the right camera will be $f((\phi - 3\psi + \theta), \gamma)$. The above four brightness functions are the brightness functions $E^L(x^L, y^L, t_1)$, $E^R(x^R, y^R, t_2)$, $E^L(x^L, y^L, t_3)$ and $E^R(x^R, y^R, t_4)$, respectively, as shown in Figure 3.3. Using the interpolation technique discussed in Chapter 3, the brightness function $E^R(x^R, y^R, t_3)$ can be obtained.

Having obtained the four frames of images, the Gauss-Seidel method discussed in Chapter 2 can be applied to obtain the unified optical flows iteratively.

The motion of the object can be recovered by using the eight equations given in Chapter 3 (3.51 - 3.58). Next, we shall discuss the experimental results.

4.3 Analysis of the Experimental Results

Each experiment performed follows the procedures below.

1. Choose imaging setting parameters. They include:

   - f: focus length of the camera
   - D: distance between the sphere center and the origin O
   - θ: intersecting angle of two viewing directions
   - R: sphere radius
   - $\phi/\Delta T$: the angular velocity at which the sphere is rotated uniformly and clockwise around its axis.

2. Simulate four frames of images which correspond to $E^L(x, y, t_1)$, $E^R(x, y, t_2)$ $E^L(x, y, t_3)$ and $E^R(x, y, t_4)$ as shown in Figure 3.3, respectively. Then interpolate an image which corresponds to $E^L(x, y, t_2)$ in Figure 3.3.

3. In image frame $E^L(x, y, t_1)$ shown in Figure 3.3, decide which pixel is to be examined.

4. Geometrically determine the actual values of 3-D location and motion parameters of the corresponding point, including X, Y, Z, $\dot{X}$, $\dot{Y}$ and $\dot{Z}$, on the sphere surface.

5. Use unified optical flow method to recover the corresponding 3-D location and motion parameters.

6. Examine error of each parameter.

4.3.1 Experiment 1

The parameters of the imaging setting are shown in Table 4.1.

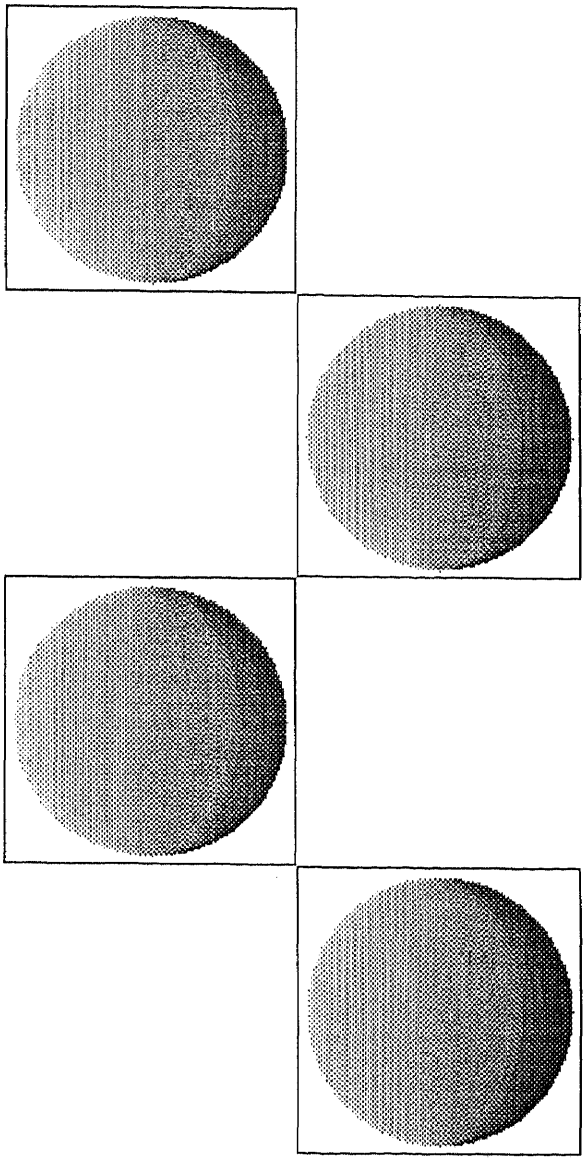

Figure 4.2: A set of simulated images obtained corresponding to the parallelogram four-frame model

f	D	θ	R	$\phi/\Delta T$
10.0	1000.0	0.04°	0.0001	3°

Table 4.1: Imaging setting parameters in Experiment 1

First, we examine a particular point on the sphere surface. The actual and the recovered values of the 3-D location and motion parameters of this point are shown in Table 4.2.

	Actual Value	Recovered Value	Percentage of Error
X	0.000040000	0.000040000	0%
Y	- 0.000056667	- 0.000056667	0%
Z	999.999927966	1000.000049797	0%
$\dot{X}$	0.000003715	0.000003715	0%
$\dot{Y}$	0.000000000	0.000000000	0%
$\dot{Z}$	-0.000002192	-0.000002194	0.09%

Table 4.2: Comparison of the actual and the recovered location and motion parameters in Experiment 1

The values obtained by the simulation are compatible to the values obtained mathematically. Note that the recovered 3-D motion parameter $\dot{Z}$ is very close to the reference value. The recovered X, Y, $\dot{X}$ and $\dot{Y}$ and the reference $\dot{X}$ and $\dot{Y}$ are so close that they are essentially the same within the given precision range.

Table 4.3 shows the error bounds of each parameter per pixel over the whole image.

Further analysis is needed to show how the parameters of the imaging setting affect the accuracy of the recovery. These effects are demonstrated by the subsequent experiments.

Parameter	X	Y	Z
Error range ρ	ρ <0.0001%	ρ <0.0001%	ρ <0.0001%
Parameter	$\dot{X}$	$\dot{Y}$	$\dot{Z}$
Error range ρ	ρ <0.0001%	ρ <0.0001%	0.05< ρ <0.1%

Table 4.3: Overall error performance of recovery in Experiment 1

4.3.2 Experiment 2

The parameters of the imaging setting are shown in Table 4.4.

f	D	θ	R	$\phi/\Delta T$
10.0	1000.0	0.04°	4.0	3°

Table 4.4: Imaging setting parameters in Experiment 2

Similar to Experiment 1, we first examine one particular point on the sphere surface. The actual and the recovered values of the 3-D location and motion parameters of this point are shown in Table 4.5.

	Actual Value	Recovered Value	Error range ρ
X	1.595377513	1.595379994	ρ <0.0002%
Y	- 2.260118143	- 2.260121658	ρ <0.0002%
Z	997.110945384	997.112496179	ρ <0.0002%
$\dot{X}$	0.151270543	0.155975030	ρ =3.0162%
$\dot{Y}$	0.000000000	0.009878195	
$\dot{Z}$	-0.083533770	-4.445481432	>1000%

Table 4.5: Comparison of the actual and the recovered location and motion parameters in Experiment 2

Note that the imaging setting in this experiment is the same as in the previous one except that the radius of the sphere R is increased to 4.0. In this case, the location parameters recovered agree well with the theoretical ones.

However, the recovered motion parameters, in particular, Z, are not acceptable as compared to the theoretical values.

Table 4.6 shows the error bounds of each parameter per pixel over the whole image.

Parameter	X	Y	Z
Error range ρ	ρ <0.0002%	ρ <0.0002%	ρ <0.0002%
Parameter	X	Y	Z
Error range ρ	ρ <17.500%	ρ <17.500%	ρ >1000.0%

Table 4.6: Overall error performance of recovery in Experiment 2

From the dramatic change of the accuracy of recovery due to the change of the scale of the observed object, one may guess that this is brought about by the presumed *far field* condition, to which this experiment violates. Since all the derivations conducted in the paper do not employ any measurement unit, the *far field* condition can be satisfied by either increasing the scale of the observing distance, which is represented by value D, or by decreasing the scale of the observed object, which is represented by value R. The following experiment will demonstrate this conjecture.

4.3.3 Experiment 3

The parameters of the imaging setting are shown in Table 4.7.

f	D	θ	R	$\phi/\Delta T$
10.0	Varying	0.04°	Varying	3°

Table 4.7: Imaging setting parameters in Experiment 3

Here, all parameters but D and R are fixed. We are varying D and R in order to examine the far-field condition.

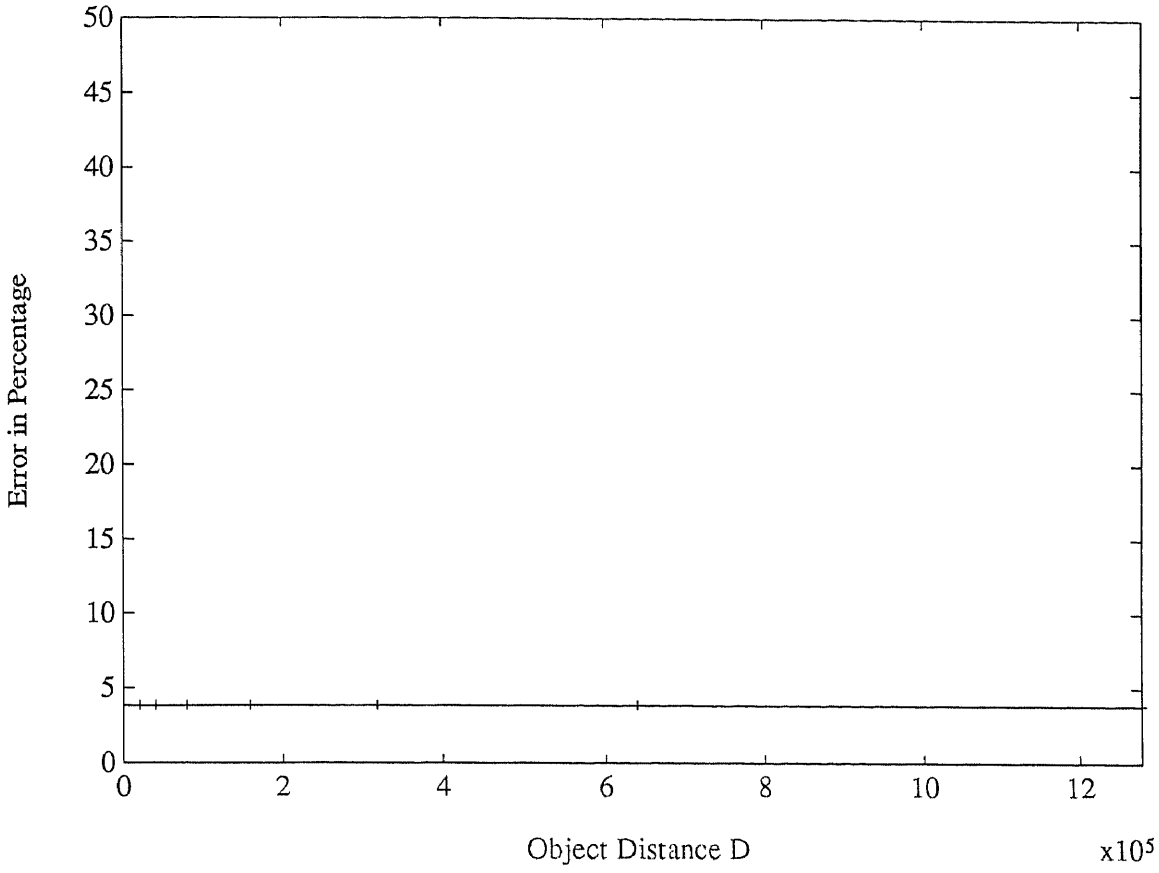

Figure 4.3: Error in percentage in recovering $\dot{Z}$ vs. object distance D with D/R fixed.

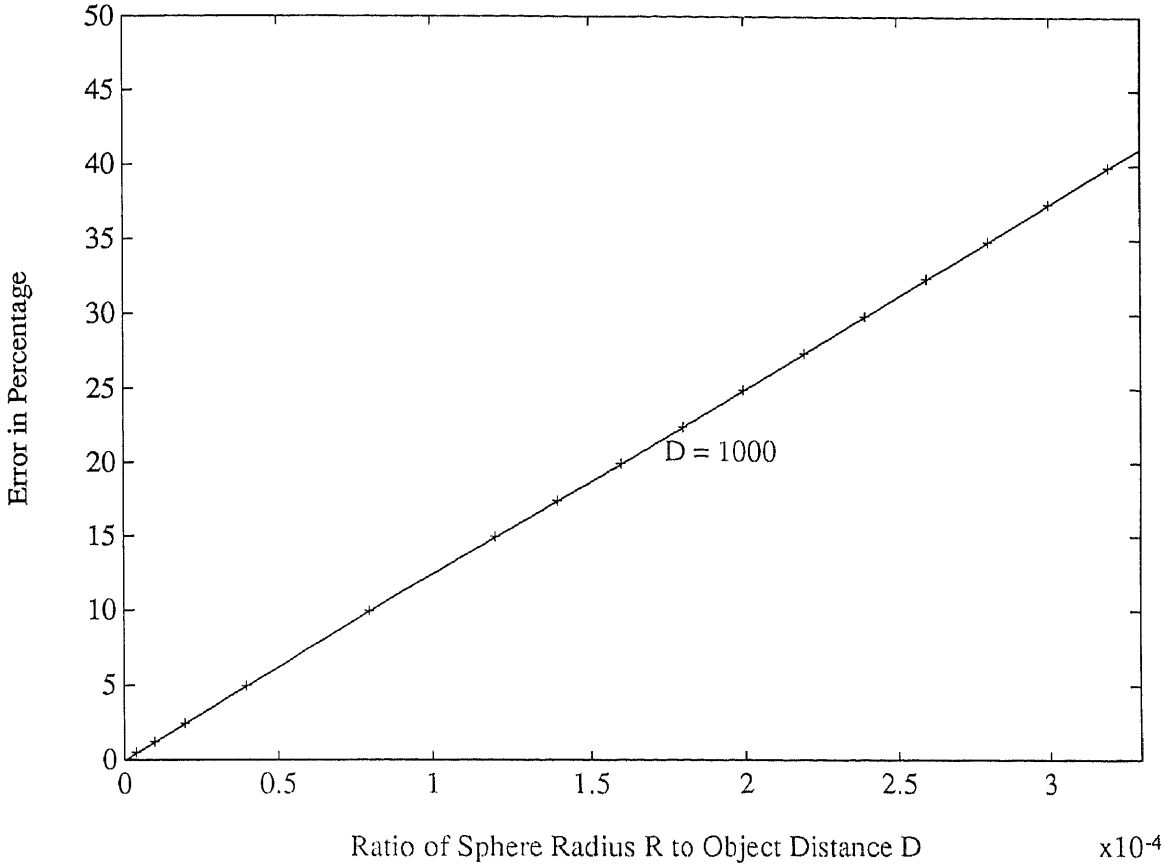

Figure 4.4: Error in percentage in recovering $\dot{Z}$ vs. ratio R/D with D fixed at 1000.0.

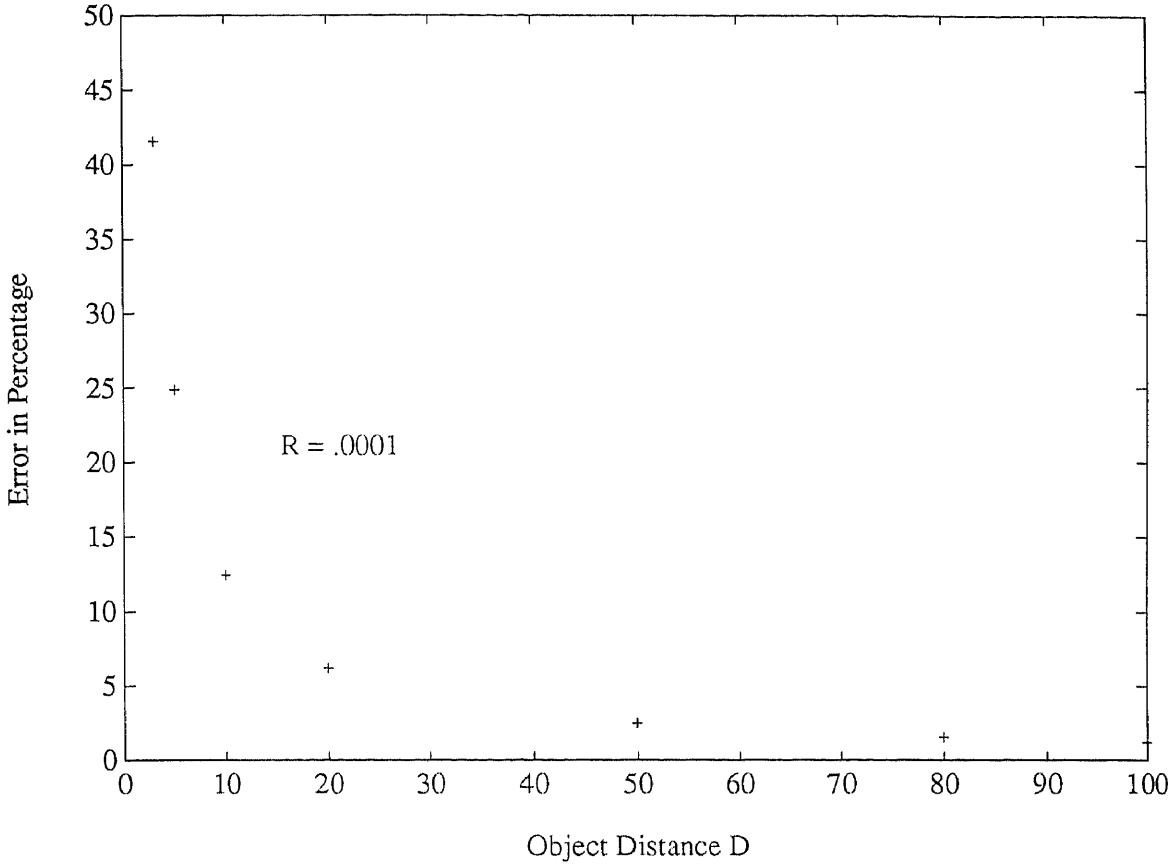

Figure 4.5: Error in percentage in recovering $\dot{Z}$ vs. object distance D with R fixed at 0.0001.

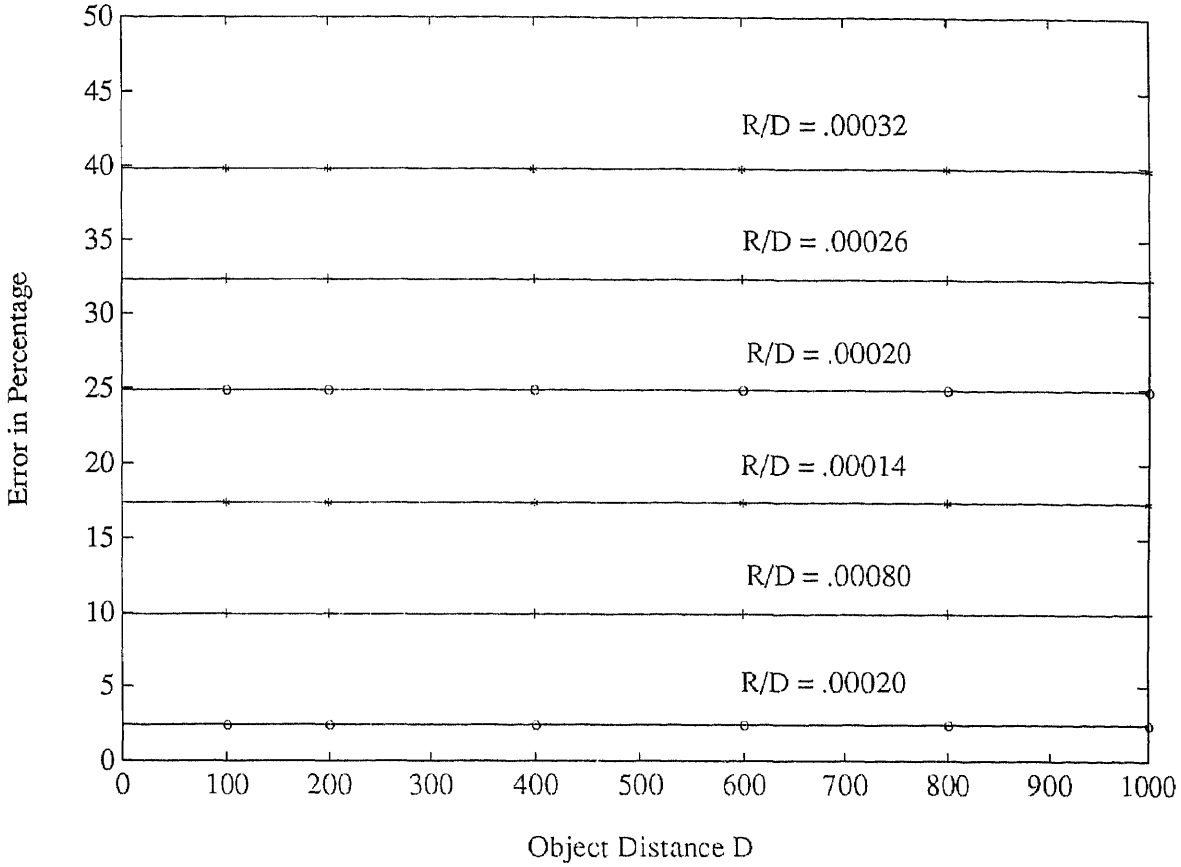

Figure 4.6: Error in percentage in recovering $\dot{Z}$ vs. object distance D for various values of R/D.

Following conclusions are drawn from these simulations.

- Figure 4.3 shows that the error in recovering motion parameter $\dot{Z}$ remains almost the same when R and D vary proportionally, *i.e.*, the ratio R/D is constant.

- Figure 4.4 shows that the error in recovering $\dot{Z}$ increases linearly with respect to the ratio R/D in the range $0 < R/D < .00035$.

- Figure 4.5 shows that the error in recovering $\dot{Z}$ decreases inversely proportional to the object distance D. Here, the sphere radius R is fixed.

- Figure 4.6 shows that the error in recovering $\dot{Z}$ increases with respect to R/D, regardless of the scale of D.

These figures show that the error depends on the ratio R/D which is a measure of the far-field condition.

4.3.4 Experiment 4

The sphere surface can be reconstructed by scanning every pixel in the captured image, and by recovering the 3-D coordinates of its corresponding point on the sphere surface.

Table 4.8 shows the parameters of the imaging setting.

f	D	θ	R	$\phi/\Delta T$
10.0	1000.0	0.04°	0.0001	3°

Table 4.8: Imaging setting parameters in Experiment 4

We employ the following two measures to evaluate the reconstruction algorithm.

- The mean absolute error:

$$\varepsilon = \frac{\sum_{i=0}^{SIZE-1} \sum_{j=0}^{SIZE-1} |Z_{rec} - Z_{acu}|}{SIZE \times SIZE}$$

where Z_{rec} is the reconstructed value of coordinate Z, and Z_{acu} is the accurate value of it. In this experiment, $\varepsilon = .00122$

- The mean square error:

$$\sigma_\varepsilon{}^2 = \frac{\sum_{i=0}^{SIZE-1} \sum_{j=0}^{SIZE-1} (Z_{rec} - Z_{acu})^2}{SIZE \times SIZE}$$

In this experiment, $\sigma_\varepsilon = .00122$

The error in the depth recovery is very small compared with the actual depth which is about 1000.0 for each pixel on the sphere surface.

To show that the recovered plane *is* a sphere surface, a graphic approach is used to visualize the simulation results. Four images are shown in Figure 4.7. Images (a), (b) and (c) show the three projections on three different planes YOZ, ZOX and XOY from three directions XO, YO and ZO, respectively.

Here the brightness of the images does not reflect the actual brightness value of the corresponding point on the object sphere. Instead , the brightness is utilized to indicate the depth of the object Z, i.e., the further a point on the sphere is away from the the camera, the darker the corresponding pixel in the projection is. Note that the perimeter of image (a) is a circle, and the perimeters of image (b) and (c) are semicircles.

Furthermore, image (d) shows a pattern of several concentric circles. Each circle is an equidistance circle indicating that the reconstructed surface corresponding to the points on the circle is of the same depth. The depth corresponding to each circle increases from the center to the outer circles. This agrees with the surface of a sphere.

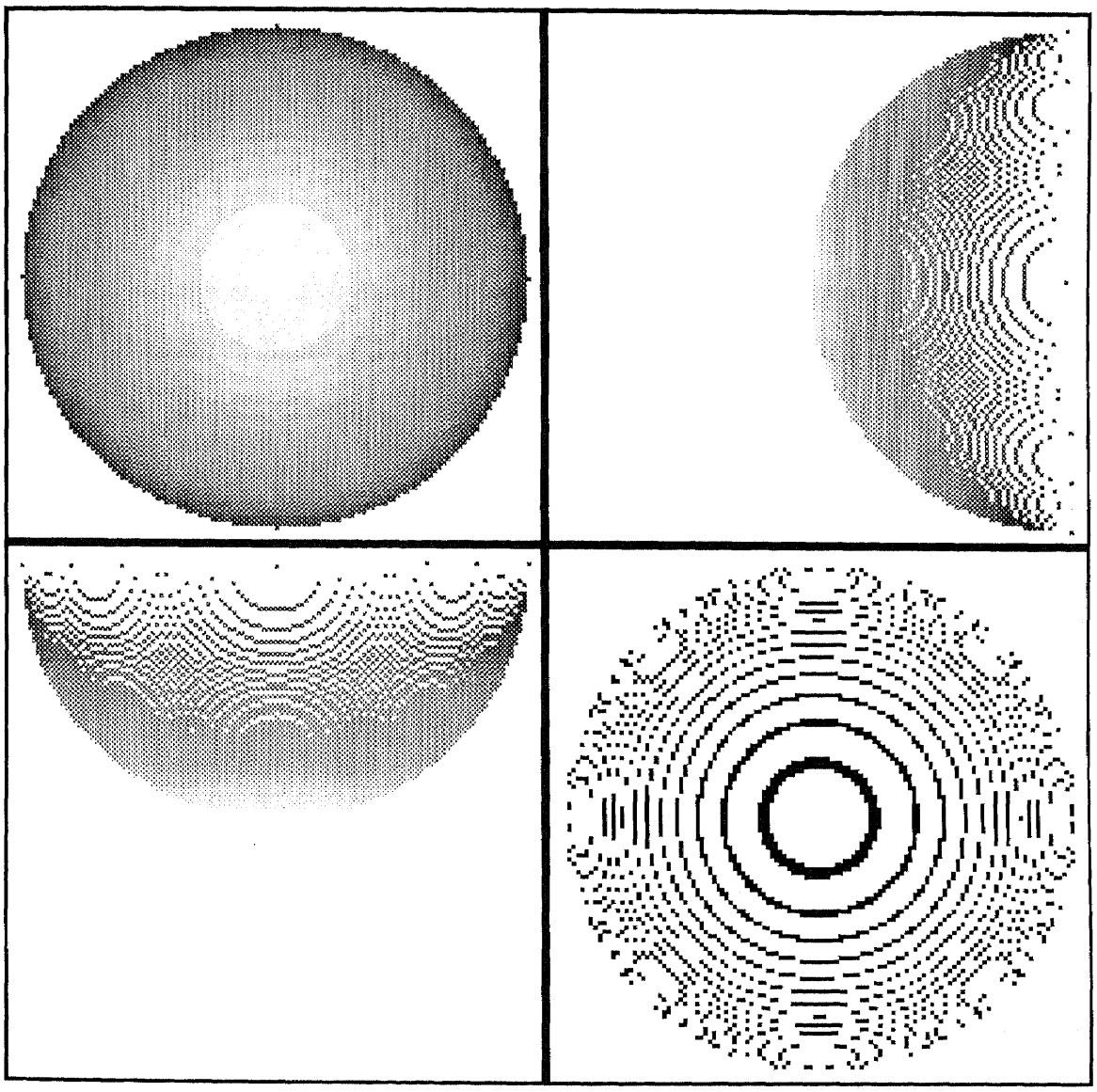

Figure 4.7: Projections from different directions of the reconstructed sphere

Chapter 5

Conclusions

From all the analysis we conduct and the experimental results we obtain in Chapter 2, 3 and 4, the following conclusions are drawn:

1. A new imaging system, consisting of only one sensor, two plane mirrors and two switches, is implemented in this paper. Based on the unified optical flow field, a parallelogram four-frame model is established along with a set of formulation to recover the surface structure and 3-D motion being proposed.

2. An image interpolation technique is used in this paper by employing the optical flow concept and scheme. Based on two "extreme" images taken at two instants, any "intermediate" image between two instants can be estimated.

3. The stereo imaging scheme presented in this thesis requires only one sensor(or camera). Such scheme eliminates the highly restricted requirement of two optically identical sensors in stereo imagery. Hence, it is expected to be more useful in practice.

4. Detailed experiments are conducted in Chapter 4. The accuracy of the

structure and motion recovery is analyzed in a quantitative way. Different aspects in the imaging setting which affect the accuracy of the overall recovery result are discussed. It is noticed that the accuracy of the recovery can be significantly affected if the far-field condition is not satisfied. Furthermore, we notice that the accuracy of the recovered parameters varies even for the same imaging setting. The accuracy of the recovered parameters falls into following three levels.

- Level 1: X, Y and Z.
- Level 2: $\dot{X}$ and $\dot{Y}$.
- Level 3: $\dot{Z}$.

From Level 1 to Level 3, the accuracy of recovery decreases rapidly.

However, it is shown through the simulations that if the imaging setting parameters are chosen properly, the error of recovery can be minimized.

5. In conclusion, the application of the new arrangement of the imaging system presented and the unified optical flow approach established to analyze structure and motion from a sequence of stereo images are very promising.

Bibliography

[1] B.K.P. Horn and B.G.Schunck, "Determining optical flow," *Artificial Intelligence* 17 (1981) pp.185-203

[2] C.Q. Shu and Y.Q. Shi, "A new approach to motion analysis from a sequence of stereo images," *Journal of Visual Communication and Image Representation.* (Submitted)

[3] J. Weng, T.S. Huang and N. Ahuja, "Motion and structure from two perspective views: algorithm, error analysis and error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 451-476, May 1989.

[4] Y.C. Kim and J.K. Aggarwal, "Determining object motion in a sequence of stereo images," *IEEE J. Robotics Automation*, vol. RA-3, no. 6, pp. 599-614, Dec. 1987.

[5] J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images - a review," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 917-935, August 1988.

[6] A. Mitiche, "On combining stereopsis and kineopsis for space perception," *Proceedings of the First Conf. on Artificial Intelligence*, (Denver, CO), pp. 156-160, Dec. 1984.

[7] H.H. Nagel, "Dynamic stereo vision in a robot feedback loop based on the evaluation of multiple interconnected displacement vector fields" *Proceedings of the International Symp. on Robotics Research*, pp. 200-206, 1985.

[8] Conte, S.D. and de Boor, C., *Elementary Numerical Analysis* (McGraw-Hill, New York, 1965, 1972).

[9] Hamming, R.W., *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1962)