MODEL BASED ESTIMATION OF IMAGE DEPTH AND DISPLACEMENT

NAGW-1333

by

Kevin T. Damour

Rensselaer Polytechnic Institute
Electrical, Computer, and Systems Engineering Department
Troy, New York 12180-3590

May 1992

CIRSSE REPORT #113

Table Of Contents

List of Tables

List of Figures

Acknowledgment

I would like to acknowledge the help and friendship provided by the professors, staff, and colleagues during my doctoral studies here at RPI. In particular I wish to thank Professor Howard Kaufman for his thoughtful suggestions, supervision, and direction provided for the research project and for his invaluable help in the preparation of this thesis. I consider myself most fortunate to have had a mentor and friend such as he.

Particular thanks is given to Professors Holmes, Kelley, Roysam, and Woods who served on my committee and provided many helpful suggestions and guidance in support of this thesis.

I would like to acknowledge grant support from NASA under NAGW-1333, from NSF under grant MIP-9013247, and travel support from NATO Collaborative Research Program grant CRG-900834.

I would like to thank Professors George Saridis and Alan Desrochers for the direction and organization they provided for the Center for Intelligent Robotic Systems for Space Exploration. Additionally, thanks to Betty Lawson, Marilyn Mamone, and Denise Elwell who always had helpful suggestions and provided logistical support for my research.

Thanks to Professor Stewart of the Computer Science Department of RPI for providing me access to his vision facilities and for the opportunity to discuss topics of mutual interest with his graduate students.

Finally, thanks to my parents, Paul and Nancy, for all your love, support, and encouragement over the years. Thanks B.

Abstract

Passive depth and displacement map determinations have become an important part of computer vision processing. Applications that make use of this type of information include autonomous navigation, robotic assembly, image sequence compression, structure identification, and 3-D motion estimation. With the reliance of such systems on visual image characteristics, a need to overcome image degradations, such as from random image-capture noise, motion, and quantization effects, is clearly necessary. Many depth and displacement estimation algorithms also introduce additional distortions due to the gradient operations performed on the noisy intensity images. These degradations can limit the accuracy and reliability of the displacement or depth information extracted from such sequences.

Recognizing the previously stated conditions, a new method to model and estimate a restored depth or displacement field is presented. Once a model has been established, the fields can be filtered using currently established multidimensional algorithms. In particular, the reduced order model Kalman filter (ROMKF), which has been shown to be an effective tool in the reduction of image intensity distortions, was applied to the computed displacement fields. Results of the application of this model show significant improvements on the restored field. Previous attempts at restoring the depth or displacement fields assumed homogeneous characteristics which resulted in the smoothing of discontinuities. In these situations, edges were lost. This thesis provides an adaptive model parameter selection method

that maintains sharp edge boundaries in the restored field. This has been successfully applied to images representative of robotic scenarios.

In order to accommodate image sequences, the standard 2-D ROMKF model is extended into three dimensions by the incorporation of a deterministic component based on previously restored fields. The inclusion of past depth and displacement fields allows a means of incorporating the temporal information into the restoration process. A summary on the conditions that indicate which type of filtering should be applied to a field is provided.

CHAPTER 1

1. Introduction to Research Topic

1.1. Problem Definition and Motivation

Passive depth and displacement field estimations have become an important part of computer vision and image processing. Applications include autonomous navigation, vision assisted robotic assembly of structures, image sequence compression, structure identification, and 3-D position and motion of objects. With the reliance of such systems on visual characteristics, a need to overcome image degradations or distortions, such as from random image-capture noise, motion blurring, and quantization effects, is clearly required. Many algorithms, used in the estimation of displacement and depth fields, introduce additional distortions due to the gradient operations applied to the input intensity images. These degradations can limit the accuracy and reliability of the displacement or depth information extracted from such images. With these observations in mind, a new method to estimate a restored depth or displacement field is presented.

Specifically, this thesis is concerned with the application of a *model-based* approach to the estimation of depth and displacement maps from image sequences or stereo image pairs. Once a model has been developed, the fields can be filtered using established multidimensional algorithms. A model-based Kalman type estimator is presented for spatio-temporal filtering of noise and degradations in the depth and displacement fields. Of particular interest is the estimation of displacement in general continuous fields and fields with rigid objects of known shape and dimensions. Results

show that the reduced order model Kalman filter (ROMKF) [6] is an effective procedure to reduce distortions in the estimated fields.

In order to accommodate image sequences, the standard 2-D ROMKF model is extended into three dimensions by the incorporation of a deterministic component based on previously restored fields. The inclusion of past depth and displacement fields allows a method of incorporating the temporal information into the restoration process. A summary of the conditions that indicate which type of filtering (i.e., spatial homogeneous, spatial multiple-model, or spatio-temporal) should be applied to a field is provided.

1.2. Contributions

The depth and displacement fields calculated from intensity images are crucial components in many computer assisted operations. When dealing with corrupted intensity images, the accuracy and reliability of the estimation of these fields are questionable. This thesis presents a method that deals with these situations and provides the following contributions:

- The presentation of a model to describe the underlying process for the depth and displacement field is given. Such modeling provides a method of obtaining more accurate and reliable field results than current non-model based estimation algorithms.

- The application of the ROMKF with a non-symmetric half plane support region is used to restore distorted depth and displacement fields. The ROMKF allows for a reduction in the order of the system state vector for filtering purposes, thus reducing the computational complexity of

model parameter estimation and filtering. The model parameters are estimated from the distorted fields.

• The use of adaptive parameter selection in the filtering procedure allows for changes in the underlying depth and displacement fields when dealing with discontinuity regions. The use of a homogeneous support model tends to smooth out edge content, whereas a multiple model approach allows for distortion reduction while maintaining clear discontinuities, sharp edges, in the restored fields.

• The extension of established general 2-D spatial modeling to 3-D by the incorporation of a deterministic temporal component, which is based on the results of the previously restored field, is presented. This allows for the processing of pairs of image sequences.

• The provision by which direct or indirect observations may be employed is provided in the selection process of the observation equation. Direct observation deals with the actually corrupted intensity images, while indirect observation deals with the distorted depth and displacement fields obtained through an external source, such as those provided through a stereo region matching algorithm.

• The use of prefiltering the intensity images prior to the depth and displacement estimation stage is shown to yield greater noise reduction in the restored field.

1.3. Terminology

Symbology:

$\mathbf{d}$ - true displacement 2-D vector from t to t + T

$\hat{\mathbf{d}}$ - estimated displacement 2-D vector from t to t + T

d - depth or displacement component, x or y

D - depth or displacement system state vector

I - intensity value

T- time between frames

X - 3-D location of a point in space

x - 2-D location of a point in an image

Constant Displaced Intensity relates intensity changes to the displacement fields by assuming that the intensity remains constant along the true displacement vector, $\mathbf{d} = (d_x, d_y)$:

$$I(x, y, t) = I(x + d_x, y + d_y, t + T) \tag{1}$$

Figure 1.1 graphically shows this assumption for an image sequence.

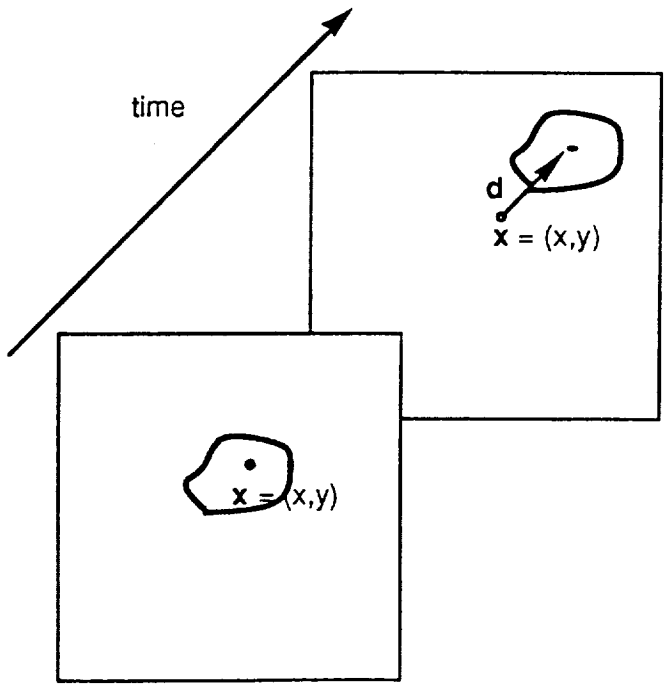

Figure 1.1 - Constant Displaced Intensity Representation

Depth Field is a dense 2-D field of scalar values representing the depth of the object imaged in each pixel location relative to a specified reference location. The term depth map is equivalent to depth field.

Dense Field is a field composed of a set of values for each location. In image fields, values are computed for each pixel, although some areas may be noted to contain no information. (See Sparse Field).

Displacement Field is the dense 2-D field of vectors describing the movement of intensity regions from one image to another. The 2-D field of vector values at each pixel represents the direction and distance of translation of intensity regions from one frame to another in image sequences. In specific stereo camera setups, one component of the vector value is zero, so the displacement field reduces to a 2-D field of scalar disparity measurements between the left and the right images. (See Section 2.2). The term displacement map can be used interchangeably with displacement field.

Displaced Frame Difference, abbreviated as dfd, is the intensity difference between a pixel and the past frame's pixel shifted by the estimated displacement vector, $\hat{\mathbf{d}} = (\hat{d}_x, \hat{d}_y)$. Using the assumption of constant pixel intensity between displaced frames, the displaced frame difference is equal to zero when the estimated displacement is identically equal to the true displacement:

$$\mathrm{dfd}(x, y; \hat{d}_x, \hat{d}_y) = I(x, y, t) - I(x + \hat{d}_x, y + \hat{d}_y, t + T) \qquad (2)$$

Disparity refers to the length in pixel units of the correspondence of an image intensity region from one image to another in a stereo camera setup.

Optical Axis is the perpendicular projection from the image plane through the center of the lens or the pinhole in a simple camera model, see Figure 1.2.

Optical Flow, mathematically introduced by Horn [27], is the apparent motion of intensity regions in an image due to motion of the object imaged, motion of the viewer, or a combination of these effects. This relative motion is described as vectors in a dense 2-D field. This field will be referred to as the displacement field.

Perspective Projection is a common method used to describe the object's image formation process. This projection relates a point, $\mathbf{X} = (X, Y, Z)$, in 3-D space back to a point, $\mathbf{x} = (x, y)$, in the 2-D image plane. This type of projection is also referred to as an ideal pinhole camera or back projection [53].

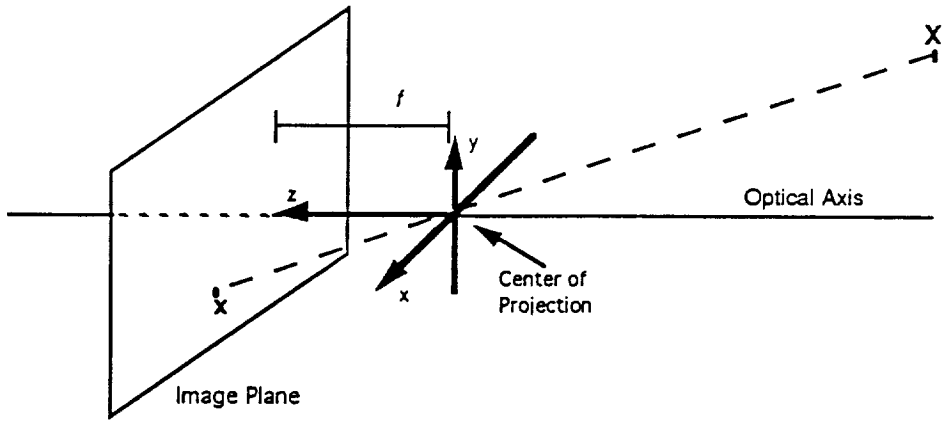

Figure 1.2 - Perspective Projection Camera Model

By using the simplifying assumption that light travels in straight lines, simple geometric relations, see Figure 1.2, can be developed. In this figure, the optical axis is defined as the perpendicular line from camera lens or center of projection to the image plane. By aligning a Cartesian

coordinate system with the origin coincident with the center of projection and the z-axis directed towards the image plane along the optical axis, a right handed system is defined. The optical focal length, f, is the linear distance of the image plane from the lens or center of projection. The following geometric relations are found [26]:

$$x = \frac{f\ X}{Z} \quad , \quad y = \frac{f\ Y}{Z} \tag{3}$$

Sparse Field is a field composed of a relatively small number of values computed for selected locations. When a sparse field is determined, the values are computed for a few identified feature points as opposed to every point in the image. (See Dense Field).

1.4. Applications and Relevance

Passive displacement and depth field determinations have become important components in robotic control and vision processing. The use of depth and displacement fields has found many applications ranging from autonomous navigation of planetary explorers, robotic assembly of structures with vision assisted path planning, to intensity image frame interpolation and medical imaging and diagnostics. Such robotic work is currently being carried out in the NASA CIRSSE Labs at Rensselaer [14].

Unfortunately, the depth and displacement fields are unknown in general and usually must be estimated from the input intensity sequence pairs. The above systems must work with intensity image data that may be corrupted by noise or other degradations. These degradations and various gradient operations on the corrupted intensity images greatly reduce the accuracy and reliability of information extracted from the estimated depth

and displacement fields, which in turn degrade the overall system's performance. To improve performance, the degraded depth and displacement fields must be restored. This thesis presents a successful modeling and filtering procedure that accomplishes this restoration. Thus, more accurate information is provided for the systems to process.

Displacement maps have become an important tool for recovering 3-D motion and depth from image sequences. Heeger and Jepson in [24] approach the nonlinear problem of recovering 3-D motion parameters and depth by separating the task into recovering first the translational components, then the rotational components, and finally the depth. To test their algorithms, the authors make use of an optical flow field calculated from known camera motion and a known depth map. This avoids the issues of degradations in computing displacement fields from the actual image sequences as would be required for an unknown depth field.

Another technique to estimate structure and 3-D motion [1] derives these properties in three dimensions from the estimated displacement field and its spatial first and second order derivatives. These techniques require a smooth variation in the displacement field and assume smooth surfaces on the objects imaged. This underlines the importance of restoration of the displacement field, since typical levels of distortions or noise sources can severely degrade the reliability of higher order derivatives.

Recovering the 3-D parameters and depth from images is a central issue to robotic applications. Autonomous land navigation and robotic assembly make use of displacement maps as a tool to establish the world environment [3, 10, 21, 25, 39, 50, 63]. The ability to navigate, avoid,

identify, and track objects are major goals for vision input to robotic applications. Dense depth and displacement field calculations from stereo pairs offer several advantages over other sparse depth estimation systems such as in structured lighting systems. Among the advantages are:

- A non-active operation which avoids alignment and detection of light patterns.
- Consumption of less power than active lighting methods, which may be of considerable importance for space operations.
- Capability for complete scene capture which allows for subsequent computations on a temporally consistent dense field.

Complete scene capture allows for calculations over most of the image features at a single time instant and thus is suited for determination of time varying object parameters. This allows modeling to be done with relatively fewer frames than is required with a sparse sampler such as laser spotter systems. A related topic to non-active sparse samplers is the feature matching procedure in which prominent features are extracted from the image and the 3-D parameters are determined. Several authors [16, 33] point out the difficulty in determining the features and the severe problems in sensitivity to ambiguity.

As with any method of solution, dense depth and displacement estimation calculations can present several problems. The major problem is the computational cost due to the large quantity of data required when dealing with images. Methods to overcome this cost will be addressed in Section 4.5 of this thesis. Areas in which there is little contrast pose problems for gradient based displacement estimators. One author [42]

suggests that the high contrast areas be identified and that these regions be used to estimate the displacement. Discontinuities in the depth field also present a challenge. This is of particular importance to robotic assembly tasks in which the identification of the boundaries of the object determined by the edges of the depth fields must be accurately known. Incorrect estimates may result in damage to the object or the robot. The model based restoration procedure detailed in this thesis overcomes these problems.

Bandwidth compression is fast becoming a required function to address the expanding requirements on information transfer. From its early beginnings [46, 47], image coding has made use of the temporal correlations between frames. Today, issues of High Definition Television, multimedia applications, video phones, and teleconferencing have extended the need for further research in motion compensated coding and compression.

One related area that has shown great promise is motion compensated image sequence restoration [29, 30, 34, 35]. A vast amount of material and algorithms exist for restoration and noise suppression in a single intensity image. With the availability of digital image sequences, the strong temporal correlation between successive images can be integrated into the modeling and restoration process, thus producing a superior image product. In the past, the high cost of computation and limitations on the systems resources available, such as processing speed, computer memory (amount of as well as access time), and limited disk based storage, have made motion compensated restoration prohibitive. With the introduction of massively parallel computers, dedicated array processors, and large memory banks, these techniques will become more prevalent. The results of this

thesis will be crucial in providing the displacement information necessary for these intensity image restoration algorithms.

1.5. Thesis Outline

This thesis is sectioned into six chapters that cover the background, research approach, results, and summary of the research.

Chapter 2 is concerned with the specific problem encountered with the determination of the depth and displacement fields. Pertinent literature is cited to provide a history of the techniques used to estimate depth and displacement fields. Several algorithms currently used to estimate the fields and the limitations of such systems are discussed. Since these algorithms work with corrupted input intensity images, various distortions are introduced by gradient or correlation operations applied to the intensity images. This historical treatment provides the background to the problem and presents the initial motivation for the study of the modeling and restoration of the depth and displacement fields.

Chapter 3 presents a detailed description of the main modeling and filtering contributions of this thesis. Models are developed to describe:

a) the underlying depth and displacement fields and

b) the observation process.

The models are incorporated into a system which employs a multi-dimensional recursive filter to restore the distorted fields. A computationally efficient algorithm based on the ROMKF is used, and the models developed for depth and displacement fields are incorporated into the framework of the reduced order model. Various issues in modeling including spatial, temporal, and spatio-temporal supports are discussed.

Chapter 4 is concerned with the application and implementation issues of this restoration procedure. The coefficients for the homogeneous support model are determined from the distorted fields. The issue of discontinuities in the depth fields and its importance to robotic vision is covered in detail. This motivates one to consider a multiple model approach that describes variations of the field while allowing for reduction of distortions and maintenance of sharp edge boundaries. Since we are dealing with images, vast quantities of pixels need be processed. To provide efficient processing, a parallel processing version is described.

Chapter 5 contains the results of the estimation and restoration of distorted or degraded depth and displacement fields. Several types of images are used to demonstrate the validity of this approach and the beneficial effects of the model based filter. Effects of prefiltering the intensity images, modifying the observation equation, and adaptively selecting model support are detailed.

A discussion and summary of the success of this approach to restoring depth and displacement fields from image pairs are presented in Chapter 6.

CHAPTER 2

2. Depth and Displacement Field Formulation

2.1. Introduction

In this chapter, procedures to estimate depth and displacement from pairs of images are reviewed from current literature to provide a detailed background study of this problem. The depth and displacement fields are calculated from pairs of intensity images either from stereo image pairs or sequences of images. Several authors [10, 26, 27, 40, 44, 54, 55] have proposed methods to smooth the resulting calculations by placing constraints on the neighboring values or by filtering the fields with methods currently applied to image processing, such as low pass filters or region smoothing operators. Many of the methods operate directly on the discrete intensity values to formulate the fields. Degradations in the image formulation process result in errors in the depth and displacement fields. Not only must the fields be smoothed to account for the local gradient operations, but noise and blurring must also be considered.

2.2. Geometry for Depth and Displacement Fields

Displacement maps represent the translations of intensity regions from one frame to another or, in the case of a stereo setup, from the image obtained from one camera to another. The displacement map is a 2-D field with vector values at each pixel to represent direction and distance of translation of intensity regions from one image to another. The displacement vector is assigned to a single pixel, but for implementation purposes most researchers have made a common selection for a region as a square patch centered on

each pixel (intensity patch region). The displacement value is assumed to be constant for the patch region, and the calculated displacement vector is assigned to the center pixel. Another option available is to make the selected region dependent on the image content. This requires some form of segmentation in an effort to group pixels with similar displacements.

Displacement fields, which indicate apparent movement of intensity regions, can be used to represent changes due to depth and relative camera translations as in stereo frames or velocity of objects captured in image sequences. Although tracking intensity patch regions is a popular procedure, other methods exist to determine displacement fields. One such method is the imaged object feature-based method [24, 25]. In feature-based methods, prominent intensity image features are extracted from each scene and tracked from image to image. This procedure usually produces a sparse field for the displacement in the scene.

When dealing with frame to frame changes taken over a period of time, assuming constant lighting conditions, the displacement map represents the temporal variation in the intensities that are due to motion of the observer, motion of individual objects in the scene being imaged, or a combination of these effects. A scene undergoing 3-D motion produces a projection of the motion as translations of intensity patch regions on the image plane. These sequences of time-variant images are represented as 2-D vector fields, referred to as the velocity map or optical flow. This velocity map represents the motion (both magnitude and direction) of the intensity patch regions in the image plane. Apparent motion in the image plane may also be due to rapid changes or movements in the lighting

conditions. Changes in light intensity, from one frame to the next, brought about by variation of the lighting conditions creates an additional modification of the intensity patch regions that is beyond the scope of this thesis.

Figure 2.1 shows a graphical example of a displacement field, **DF**(t), for a pair of sequential images and how a series of displacement fields can be generated for an extended set of image sequences.

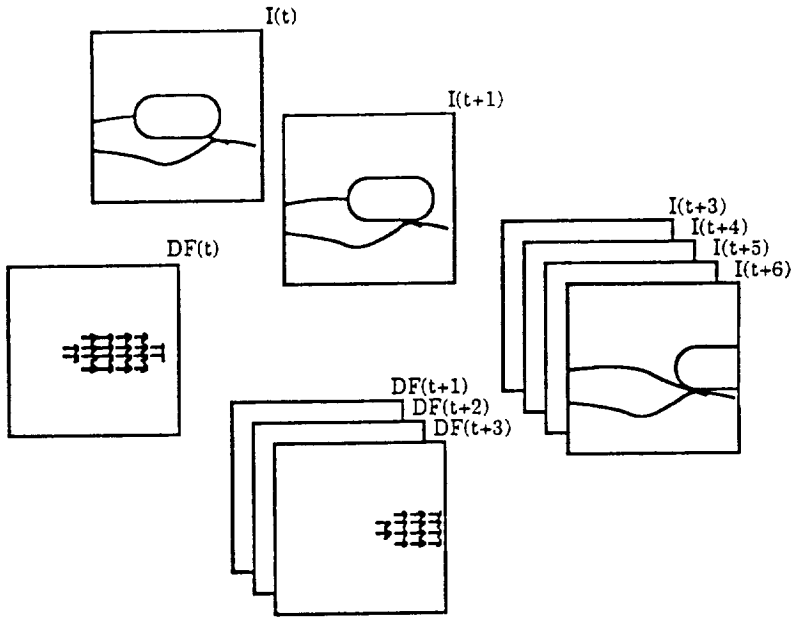

Figure 2.1 - Displacement Fields for Image Sequence

When dealing with a stereo camera setup, two images of a static scene are captured simultaneously. The displacement field is calculated from the stereo images on a per pixel basis and can be used to compute the depth of the imaged locations from the camera setup. The camera positions and optical properties of the imaging system are known *a priori* and are usually configured to facilitate the identification of the correspondence between objects in both images. A common method is to align the optical axes so that

they are mutually parallel and perpendicular to the connected baselines of the cameras [26] as represented in Figure 2.2.

To obtain an estimate of the depth, a pixel in the left image, $\mathbf{x}_l = (x_l, y_l)$, is matched to a pixel in the right image, $\mathbf{x}_r = (x_r, y_r)$. This matching procedure is referred to as the correspondence problem. Searches for matches need only be done along the x-axis since $y_l = y_r$ is fixed by the geometry of the cameras. The image points in correspondence must lie along the same line called the epipolar line which is parallel to the x-axis in this geometric configuration.

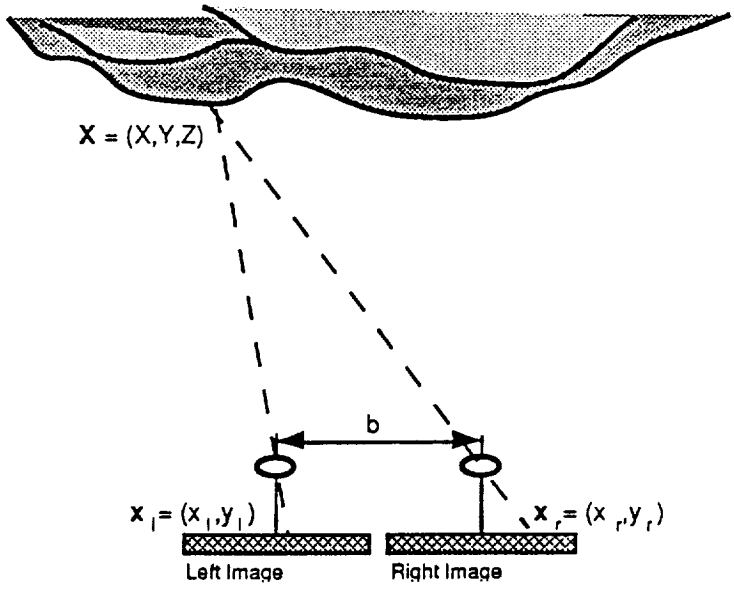

Figure 2.2 - Stereo Camera Configuration

Once the displacement, $x_l - x_r$, which is sometimes referred to as the disparity [26], is known, the 3-D coordinates of the point, $\mathbf{X} = (X, Y, Z)$, can be determined by:

$$X = b\frac{\frac{1}{2}(x_l + x_r)}{x_l - x_r}, \quad Y = b\frac{\frac{1}{2}(y_l + y_r)}{x_l - x_r}, \quad Z = b\frac{f}{x_l - x_r} \tag{4}$$

where f is the focal length of the lens and b the baseline length between the optical axes. Each component of the 3-D coordinate is inversely proportional to the estimated displacement between the frames. As the object becomes closer to the stereo camera system, the estimate of the coordinate becomes more accurate. Since the disparity is proportional to the baseline length between the cameras, making the baseline larger results in greater disparity and resolution, but the identification of the correspondence of matching regions in the stereo pair becomes increasingly more difficult to identify.

The above configuration for stereo cameras allows for rapid and efficient algorithms to be developed by restricting their "search" for image alignment to a 1-D space. The displacement field represents the distance for correspondence between the two images for this alignment, i.e., disparity, and thus represents the distance or depth of each projected pixel from the stereo camera setup.

The notation $I_l(x,y,t)$ and $I_r(x,y,t)$ denotes the intensity value at pixel location (x, y) at time t for the left and right images respectively. With the alignment established for the stereo cameras, objects imaged in the left camera appear on the same scanline as those present in the right image. In the ideal setup an imaged object appearing at $I_l(x,y,t)$ would also appear at $I_r(x-d(x,y,t),y,t)$, where $d(x,y,t)$ is the disparity value for location (x,y,t). Using $I(x,y,t)$ to denote the image signal of the scene, the left and the right stereo images may be expressed as:

$$I_l(x,y,t) = I(x,y,t) + n_l(x,y,t) \tag{5}$$

$$I_r(x,y,t) = I(x+d(x,y,t),y,t) + n_r(x,y,t) \tag{6}$$

where $n_l(x,y,t)$ and $n_r(x,y,t)$ account for the noise process in the left and right images of location (x,y) at time t. Once the disparity, $d(x,y,t)$, is estimated, the 3-D locations for the imaged points of the scene can be found relative to the stereo camera origin by (4). Notationally, the left and right camera images may also be viewed as images taken at t and $t+T$ by a single camera undergoing lateral translation.

2.3. Literature Review

Several reviewers [1, 40, 58] present a survey of the methods used in the determination of motion from a sequence of images. The most common methods are feature-based which produces sparse mapping and stereo matching and gradient based which produce dense mappings.

Aggarwal in [1] presents a comparative review of feature-based vs. gradient based (optical flow) algorithms for depth and motion estimation. Feature-based algorithms are separated into three major categories: direct methods on identified points, *a priori* knowledge on the rigidness of objects by multiple views, and extended sequence feature processing of monocular images. Feature-based algorithms have a common problem of identifying the feature points in the image, localizing the features, and then finding the correspondence between frames. These methods have an additional computational problem of how to automatically determine the number and type of features that are necessary to identify the object when attempting to characterize the displacement for complex scenes.

Whereas feature-based methods identify a few "key" components in the image to track and form a sparse field, the optical flow method produces

a dense field based on spatial and temporal gradients of the intensity images.

Some of the early work on the determination of displacement maps by optical flow is given by Horn and Schunck in [27]. They present a solution for determining the optical flow from a set of computer generated image patterns by deriving the image gradient constraint equation. This constraint equation relates the velocity of an imaged pixel to its change in intensity from sequences of images. An assumption is made that the observed intensity of a patch undergoes uniform translation over time. $I(x,y,t)$ denotes the intensity at image location (x,y) at time t. For a small increment in time δt, the assumption indicates that the same intensity would be observed at the point $(x+\delta x, y+\delta y)$ at time $t+\delta t$. This assumption is expressed as

$$I(x,y,t) = I(x+\delta x, y+\delta y, t+\delta t) \tag{7}$$

Equation (7) deals with intensity field shifts brought about through small incremental changes in x, y, and t. Although this equation appears to be similar to (1), it should be emphasized that (1) deals with interframe displacements for a frame rate T. The assumption, used in (7), of uniform translation allows for tracking the intensity field shifts over time. A Taylor series expansion can be formed about the pixel (x, y, t) to give:

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \varepsilon \tag{8}$$

for small values of δx, δy, and δt. The expansion is represented as a series of first order terms with ε accounting for the higher order terms of δx, δy,

and δt. By neglecting the higher order terms of (8) and substituting into (7), the following equation is formed [54]:

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t}+\frac{\partial I}{\partial y}\frac{\delta y}{\delta t}+\frac{\partial I}{\partial t}=0 \tag{9}$$

Taking the limit as $\delta t \rightarrow 0$, the final form for the gradient constraint equation becomes:

$$I_x v_x + I_y v_y + I_t = 0 \tag{10}$$

where

$$I_x = \frac{\partial I}{\partial x},\quad I_y = \frac{\partial I}{\partial y},\quad I_t = \frac{\partial I}{\partial t} \tag{11}$$

and (v_x, v_y) is the velocity component of the optical flow in the x and y directions respectively. The vector $\mathbf{d} = (d_x,\ d_y)$, where $d_x = v_x T$ and $d_y = v_y T$, is the desired displacement vector for the pixel located at (x, y) for time T between image frames. The collection of the displacement vectors for all pixels forms the velocity field, which is the same as the displacement field for the image pair. The gradient constraint equation (10) may be re-written in the form:

$$I_x u + I_y v = -I_t \tag{12}$$

where u and v are the x and y component of the velocity vector at location (x, y).

Equation (10) highlights the fact that the problem of velocity displacement estimation is ill posed. In this single constraint equation, since there are two unknowns, a unique velocity component can not be

determined from only a single measurement. Horn provides a graphical explanation of (10) as a constraint line in the velocity domain. Since the velocity field can not be locally determined uniquely, additional constraints are necessary.

To get around this difficulty, Horn proposes a modification of the problem where a smoothness penalty is imposed on the local velocity field. A set of iterative equations is developed [27], and results are provided only for a continuous pattern region with an optical flow that had no temporal change. The gradients are computed directly from differences in the intensity images with no consideration of image noise or other degradations. Although these results are adequate for the synthetic conditions imposed, they are not appropriate for more realistic scenes. Some problems with this technique include the lack of detection of motion boundaries and discontinuities in the intensity image, temporal changes in optical flow (i.e., translating objects against a background), noise sensitivity in the gradient estimation, and blurring of the motion boundaries. A detailed study of the errors, inherent in the local optimization of equation (10), due to the gradient measurements, non-uniformity in the flow field, and the condition on the linear equation is presented in Kearney [33].

Schunck in [54, 55] overcome some of these earlier deficiencies by making use of an algorithm that employs constraint line clustering to estimate image flow on discontinuous velocity fields. Additionally, this algorithm attempts to detect motion boundaries and turns off velocity field smoothing when in close proximity to a boundary. Boundary detection,

poorly conditioned solutions, and noise sensitivity still present problems for this algorithm.

Ballard and Kimball in [10] augment the traditional computation of general rigid body motion using displacement fields (optical flow) by also incorporating known depth information. They maintain that computation of the velocity vectors in three dimensions, which they term 3-D flow, needs to include the depth information, in addition to the optical flow, to be constrained and estimated. The Hough Transform is incorporated to relate the intrinsic image features to global parameter values and is used to obtain the solutions to the 3-D flow.

Heeger and Jepson [24, 25] compute 3-D motion parameters by decomposing the nonlinear problem of 3-D motion into three sets of equations. They propose a "direct" method based directly on the spatio-temporal gradients of the image intensity, but do not make use of it in the work presented [24]. Instead they compute the displacement field by known camera translations on a known scene. This avoids the issue of optical flow calculations on the input intensity images.

In the area of model-based displacement field estimation, Matthies, Szeliski, and Kanade in [43] estimate depth from image sequences taken with known camera motion. The optical flow equation as developed earlier in this chapter can be written in terms of known camera 3-D translational velocities, $\mathbf{T} = \left(T_x, T_y, T_y\right)^T$, and 3-D rotational velocities, $\mathbf{R} = \left(R_x, R_y, R_z\right)^T$. Using this notation, the optical flow with unit focal length, $(\delta x, \delta y)^T$, can be rewritten as in [43]:

$$\begin{bmatrix}\delta x\\ \delta y\end{bmatrix}=\frac{1}{Z}\begin{bmatrix}-1 & 0 & x\\ 0 & -1 & y\end{bmatrix}\begin{bmatrix}T_x\\ T_y\\ T_z\end{bmatrix}+\begin{bmatrix}xy & -(1+x^2) & y\\ (1+y^2) & -xy & -x\end{bmatrix}\begin{bmatrix}R_x\\ R_y\\ R_z\end{bmatrix} \quad (13)$$

where 1/Z is the inverse depth and (x, y) is the image coordinates. Equation (13) relates the known camera translations and estimated depth to the induced displacement field. A 1-D Kalman filter is used to update and filter the depth field. Their work investigates the use of small lateral camera translations in an effort to estimate the depth field. Using the known camera movements, the incremental depth, ΔZ, at a point (x,y) from time t to $t+T$ is predicted for the next frame as:

$$\Delta Z = -T_z - R_x yZ + R_y xZ \quad (14)$$

Due to the spatial quantization of images, the depth values have to be interpolated based on a neighborhood to re-orient them to the lattice. In a similarly manner, Bridwell makes use of lateral camera motion to estimate the depth of structures [12].

A method to find the displacement between image pairs that implements a simple correlation-based matching criteria is proposed by Anandan [4]. Anandan uses the Sum of the Squared Differences (SSD) based on a weighted difference between sliding intensity patches. One problem with this method is that these search methods are sensitive to interpixel interpolation methods, quantization of the motion vector, and computational complexity. This procedure will be detailed in section 2.4.2.

In contrast to the direct intensity gradient and the correlation-based methods to estimate the depth and displacement fields, Martinez in [41] proposes the use of parametric models from a set of basis functions to

describe intensity surfaces and computes the gradients directly from the model. When estimating the raw displacement map, a least squares fit to the intensity images is done on a 5x5 region centered about each pixel. To smooth the resultant displacement field, a weighted local average is used.

In contrast to the work presented in the existing literature, the work presented in this research makes use of a model-based approach to recursively filter the depth and displacement fields. The modeling is done with the inclusion of both the spatial and temporal information. By properly defining the system state vector, the displacement fields can be dynamically modeled and filtered using established multidimensional filtering algorithms. This new approach will be shown to provide improvements over existing algorithms because of the explicit modeling of the underlying depth and displacement field.

2.4. Existing Techniques for Depth and Displacement Estimation

This section reviews the current methods employed to estimate the depth or displacement using multiple images. This information is presented so that we may investigate areas that contribute to distortions or errors in the resultant depth or displacement fields.

2.4.1. Prominent Feature Matching

Prominent feature-based matching of images to estimate a displacement field results in a relatively sparse number of values in the field. Various "significant" image features are identified and then tracked to the corresponding feature points in the second image pair in a stereo setup or in the latter images of a sequence. Once these feature-based points have been obtained, a set of equations is solved to estimate the 3-D position of the

objects and possibly the 3-D velocity if the points are from a sequence of images.

Laing *et al.* [39] use a laser as an artificial light source to facilitate the identification of feature-based points by following the laser stripe in a stereo camera setup. The extended Kalman filter, along with the feature-based points extracted from the image, is used to develop a hand-eye calibration procedure for robotic assembly.

The issue of prominent features is not limited to only point locations, such as boundary intersections or object peaks, but can involve other features such as line correspondence and curve tracing. A combination of these approaches, where both point and line correspondences are determined, is described in [2].

Matthies in [43] makes use of small lateral translations to simplify the feature correspondence problem. The image features are restricted to a single scan line by the actual camera translations. Feature translations are on the order of a single pixel, thus a window width of two pixels is used to track the features. A 1-D Kalman filter in the temporal domain is used to track the image features and provides an on-line estimate of the variance of the depth estimate.

Potential problem areas with feature-based systems involve the correspondence problem between frames. This can be a difficult task depending on the number and kind of features extracted from each image. As with any matching scheme, loss of correspondence must be worked into the algorithm to avoid problems with the appearance of new features and the loss of currently tracked features. The issue of sensitivity to feature

correspondence is detailed in [16]. Finally, the feature-based systems are constrained due to the lack of a sufficient number of general models for non-rigid or curved objects (i.e., typical real objects).

2.4.2. Region Correspondence

The sum of the squared difference (SSD) is one means to establish the region correspondence between image pairs. The algorithm works by minimizing an error measure for various estimates on the displacement between the images. The displacement value that produces the minimal error is taken as the best estimate of the true displacement. The error measure is based on the intensity differences between two displaced images and can be described as:

$$e(x,y,t;d'_x,d'_y) = \left\{I_r\left(x-d'_x,y-d'_y,t\right)-I_l(x,y,t)\right\}^2 \qquad (15)$$

where d'_x and d'_y are estimates of the x and y components of the displacement vector that attempts to bring the two images into correspondence. In implementation, a square patch region is used around each location (x,y) in the image. The patch region should be large enough to capture the underlying intensity contours while reducing sensitivity to image noise, but at the same time the region must be kept small to achieve an acceptable resolution. These two requirements place an upper and lower limit on patch region size, and experimentally a 5x5 region is used for this research.

In general, the search region for correspondence between patches in the two images may extend in all directions in the image. By proper alignment of the camera's optical axes, as discussed in Chapter 1, the search

region of correspondence can be reduced to a single direction, usually confined to a single row in an image. Searches for matches need only be done along the x-axis since $y_l = y_r$ is fixed by the geometry of the cameras. The goal of region correspondence is to estimate the disparity between the pairs of stereo images described in equations (5) and (6).

A typical calculation is shown in Figure 2.3. The displacement estimate, $\hat{d}(x,y,t)$, is taken to be the "best" match defined by the minimal error measure. To estimate the displacement between two stereo images and taking into account the epipolar line, the following equations are used (utilizing a 5x5 patch):

$$e_t(x,y;d'_x) = \sum_{\gamma=-2}^{2}\sum_{\lambda=-2}^{2}\left\{I_r(x-d'_x+\gamma,y+\lambda)-I_l(x+\gamma,y+\lambda)\right\}^2 \qquad (16)$$

$$\hat{d}_x(x,y) = \min_{d'_x}[e_t(x,y;d'_x)] \qquad (17)$$

From previous knowledge of possible depth values in an imaged scene and (4), an upper and lower limit on displacement values can be established.

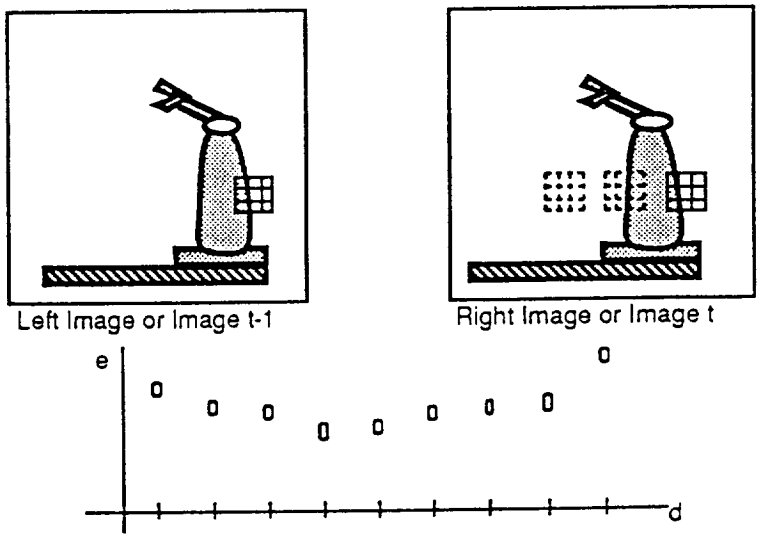

Figure 2.3 - Typical SSD Calculation

To speed the computation of (17) at each pixel location in the image, a course to fine technique is implemented. Candidate values of d'_x are first computed for integer pixel values at each location (x, y). After the minimal error measure is located for integer displacements, subpixel values are tested about that value to further refine the displacement value.

If the displacement change is small between sequences of image pairs, the previously estimated displacement field is taken as the initial estimate and a small localized area need only be searched. A gradient descent algorithm may also be implemented to provide an iterative type approach to obtain displacement estimates [46].

In addition to the pixel intensity based region correspondence algorithms, other image properties such as edges have been investigated [9, 49]. After edges are localized in an image, edge properties such as position, contrast, strength, neighboring values, and slope can be used to refine the matching procedure. Benefits of finding corresponding edges rather than intensity regions include: reduced computational requirements due to smaller data sets, greater resolution and localization, and greater invariance due to following geometric rather than photometric properties. Wohn *et al.* in [64] use contour tracing to establish the displacement vectors. A polygonal approximation is made to the contour, and an iterative scheme is developed to refine the estimated optical flow field.

2.4.3. Spatio-temporal Gradient Methods

Spatio-temporal gradient methods have been investigated by many authors. Most of the methods make use of equation (10) as the basis for computations. The central problem, termed the aperture problem, is that

(10) provides a single equation with two unknowns, i.e., the displacements in the x and y directions. Various constraints or smoothing techniques are proposed to estimate the displacement field while reducing noise due to gradient computations. These constraints usually follow the assumption that nearby pixels have similar flow characteristics. With this assumption, additional constraints [26, 27, 33, 46, 47, 54, 55] are placed on the flow equation to obtain a solution. Traditional formulations incorporate a deviation or departure from smoothness to constrain the estimated displacement field. A common method used to obtain the estimate of the displacement field is found by a combination of two error measures:

$$\varepsilon_b = \frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} \tag{18}$$

which measures the departure from the optical flow equation, and

$$\varepsilon_c^2 = \left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2 \tag{19}$$

which is a measure of the amount of departure from a smooth displacement field. In order to combine these two error measures a weighting factor, α, is introduced. The resulting equation is:

$$E^2 = \int\int \left(\alpha^2 \varepsilon_c^2 + \varepsilon_b^2\right) dx\, dy \tag{20}$$

where the total error, E, is minimized over the entire image. Using the calculus of variations and applying an iterative method based on Gauss-Seidel, a solution is possible [27]. Denoting $\bar{v}_x^n$ and $\bar{v}_y^n$ as the local average for the current estimates of the x and y component of the displacement vector at the n^{th} iteration, a set of iterative equations is developed:

$$v_x^{n+1} = \bar{v}_x^n - \frac{\frac{\partial I}{\partial x}\left[\frac{\partial I}{\partial x}\bar{v}_x^n + \frac{\partial I}{\partial y}\bar{v}_y^n + \frac{\partial I}{\partial t}\right]}{\left(\alpha^2 + \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2\right)}$$

$$v_y^{n+1} = \bar{v}_y^n - \frac{\frac{\partial I}{\partial y}\left[\frac{\partial I}{\partial x}\bar{v}_x^n + \frac{\partial I}{\partial y}\bar{v}_y^n + \frac{\partial I}{\partial t}\right]}{\left(\alpha^2 + \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2\right)} \qquad (21)$$

Finite differences on the input intensity images are used to approximate the spatial and temporal derivatives. Curiously, the current velocity components do not directly depend on the previous estimate at that point. Subsequently, various references [21, 29, 44, 45, 54, 55] have pointed out the sensitivity to image noise, boundary effects, and the over-smoothing effects of this iterative approach.

Other proposed solutions have looked at establishing a local optimization problem by utilizing flow values for neighboring pixels [33, 46, 47, 53]. The local optimization procedure is performed by solving (possibly in a least squares sense) a set of gradient constraint equations for a small neighborhood in the image. This can be written as:

$$\begin{bmatrix} \left.\frac{\partial I}{\partial x}\right|_{P_1} & \left.\frac{\partial I}{\partial x}\right|_{P_1} \\ \vdots & \vdots \\ \left.\frac{\partial I}{\partial x}\right|_{P_N} & \left.\frac{\partial I}{\partial x}\right|_{P_N} \end{bmatrix} \begin{bmatrix} v_x & v_y \end{bmatrix}^T = -\begin{bmatrix} \left.\frac{\partial I}{\partial t}\right|_{P_1} \\ \vdots \\ \left.\frac{\partial I}{\partial t}\right|_{P_N} \end{bmatrix} \qquad (22)$$

The assumption is that the neighborhood region will be large enough to contain sufficient variation to properly condition the solution, but small enough to provide good resolution. In Kearney's work [33], the conditioning

and sensitivity of these types of local optimization approaches are discussed. Additionally his results show that smoothing or blurring of the input intensity images to achieve better calculations on the gradients tends to make the linear equations more ill-conditioned.

A method that we use of in our research is based on a procedure presented by Martinez in [41]. This algorithm also starts with the optical flow equation (10) and minimizes the deviation over a parameterized intensity patch region to solve the under-constrained nature of the aperture problem. The local optimization procedure, described above, takes on the following form for N discrete points:

$$\min_{v_x,v_y} \frac{1}{N} \left\{ \sum_{i=1}^{N} \left(v_x \left.\frac{\partial I}{\partial x}\right|_{P_i} + v_y \left.\frac{\partial I}{\partial y}\right|_{P_i} + \left.\frac{\partial I}{\partial t}\right|_{P_i} \right)^2 \right\} \tag{23}$$

A parametric signal model is used to estimate the spatial and temporal gradients of the images to be used in solving equation (23). A linear parameter model with a set of basis functions ϕ, shown in (24), is assumed in order to make the computations simpler.

$$\begin{array}{lll} \phi_1(x,y,t) = 1 & \phi_2(x,y,t) = x & \phi_3(x,y,t) = y \\ \phi_4(x,y,t) = t & \phi_5(x,y,t) = x^2 & \phi_6(x,y,t) = y^2 \\ \phi_7(x,y,t) = xy & \phi_8(x,y,t) = xt & \phi_9(x,y,t) = yt \end{array} \tag{24}$$

The parametric signal model parameters, $\mathbf{S}$, were estimated by a least squares fit to the intensity image:

$$I(x,y,t) \approx \bar{I}(x,y,t) = \sum_{i=1}^{N} S_i \phi(x,y,t) \tag{25}$$

where $\bar{I}(x,y,t)$is the parametric surface approximation to the observed intensity $I(x,y,t)$. Once the parametric model parameters are known, the

gradients can be calculated directly. The issue of noise reduction is addressed by the nature of the least squares parameter estimation. The least squares motion estimator is computationally efficient for small local changes. A multi-resolution algorithm, where a set of scaled images is produced, or an extended search region may be required for displacements greater than the size of the modeling support region.

2.5. Differences in Depth and Displacement Fields

In this section we will review some important differences between depth and displacement fields. As described before, the displacement map is the 2-D vector field representing the matching of various intensity patches in one image compared with another image. In determining a displacement field from a sequence of images, a patch region centered on a pixel in an image at time t is matched to a region in another image at time t + T, where T is the sampling time between image frames. In this case the displacement field is sometimes referred to as the velocity field.

In determining a depth map from a stereo image pair, the correspondence between a patch in the left image to one in the right image is estimated. The left and right frames may also be captured by a single camera undergoing a known translation, thus falling into the sequential image notation. Since in both of these operations a 2-D vector map is produced, the term displacement map can be used to describe either of these depth fields.

There is a difference in the structure of the displacement field used to represent depth and velocity. In the stereo correspondence case, the camera setup or alignment is known *a priori*; that is the translation and rotation

between the two cameras are known. The displacement map representing depth can be reduced by the proper transformation into a 2-D field of **scalar** values. This is due to the patch translation constraints imposed by the orientations of the optical axes of the cameras. In contrast, since the motion of objects is usually not known *a priori*, the 2-D displacement map indicating pixel velocity is generally vector valued.

In addition to the differences in the dimensionality of the displacement field for depth and velocity, the relative magnitudes of typical estimates tend to be different. In stereo camera scene analysis, due to the baseline arrangement, displacements of 10 - 30 pixels are common. Correspondence may be difficult to determine. In velocity field estimation, the correspondence is not as much of a problem due to the temporal correlation between frames. If the sampling rate is sufficiently high, the displacement between frames will be small and contained within a local neighborhood. In cases of larger velocities, a greater sampling rate is selected or a multigrid algorithm is used to first estimate the larger components of the velocity field as in [41].

When the displacements are large relative to the support region used to solve the correlation problem, some type of multi-resolution or extended search region can be used in the estimation process. A subsampled image or a Gaussian image pyramid [13] may be used to aid in the correspondence problem and to estimate initial estimates of larger displacements. At each level in the pyramid structure, which represents different resolutions of the intensity image, the previous level's displacement estimate, magnified by the reduction factor, is used as the new estimate to establish the region of

correspondence. At each higher level, the estimate is refined, thus allowing the larger displacement to have improved accuracy.

In order for the Taylor series expansion used in (8) to be valid, the incremental displacements must be small. This is also important when using a window based estimator, such as in (23), due to the fact that the velocity vector usually must be constrained to the size of the analysis window. As the value of the displacement becomes larger, the accuracy of the approximation decreases. To maintain the validity of the expansion, equation (8) can be rewritten in the following form:

$$I(x+d_x, y+d_y, t+T) = I(x,y,t) + \frac{\partial I}{\partial x}d_x + \frac{\partial I}{\partial y}d_y + \frac{\partial I}{\partial t}T + \varepsilon \qquad (26)$$

and the linearization is then taken about the initial estimate of $\hat{\mathbf{d}}$. If the value is known to be large (such as if previous displacement frames are available or a pyramid algorithm is used), the displacement can be broken down into a known displacement and an incremental, subpixel, unknown component as described by:

$$\hat{\mathbf{d}}^{i}(m,n) = \begin{bmatrix} \hat{d}_x^{i-1} + \delta d_x \\ \hat{d}_y^{i-1} + \delta d_y \end{bmatrix} \qquad (27)$$

where $\hat{\mathbf{d}}^{i-1}$is the previous displacement estimate and $\delta\mathbf{d}$ is the update quantity to estimate. The unknown component, $\delta\mathbf{d} = \mathbf{d} - \hat{\mathbf{d}}^{i-1}$, can be estimated in a local neighborhood around the previous displacement estimate as was done by Netravali [46, 47]. The application of (7) and (27), to equation (26) yields:

$$I(x,y,t) - I(x+d_x-\hat{d}_x^{i-1}, y+d_y-\hat{d}_y^{i-1}, t) \cong -(d_x-\hat{d}_x^{i-1})\frac{\partial I}{\partial x} - (d_y-\hat{d}_y^{i-1})\frac{\partial I}{\partial y} \qquad (28)$$

This equation allows for large displacement values to be estimated, but the change between frames in the displacement field must still be small.

2.6. Summary

There has been considerable research applied to the determination of depth and displacement fields. This is due in part to the fact that there are many important applications which rely on the quality of the estimation of these fields. The quality of the results of these applications depends on the accuracy and reliability of the input depth and displacement fields. There exist several major methods to estimate these fields, each of which have its own benefits and limitations depending on the method of image acquisition and properties of the objects being imaged.

CHAPTER 3

3. Depth and Displacement Estimation and Restoration

3.1. Introduction

In the previous chapter, it was shown that the depth and displacement field estimation process could be determined from changes in intensity images. These input intensity images suffer from several types of degradations which in turn degrade the quality and reliability of the derived depth and displacement fields. Consequently, a process is needed which can be utilized to remove the degradations in these fields. Although much attention has been placed on the estimation procedure for depth and displacement fields, less emphasis has been placed on developing models that improve the quality (noise reduction, edge preservation, etc.) of the depth and displacement fields, thus improving the restoration process. This thesis shows that it is possible to reduce the degradations by incorporating the spatial and temporal correlations into a model of the field. The approach is to model the underlying fields and observation process and then apply this model to restore the estimated displacement fields from the corrupted intensity images.

In order to filter out degradation and noise in the fields, several authors [11, 20, 30, 43, 45, 48] have proposed modeling the displacement field in either the temporal or spatial domain and derived a filter for the field. The results of such attempts have been somewhat limited by the application of an empirical selection of modeling coefficients or the lack of adaptability to changes within the underlying field.

By employing a model-based approach that incorporates both spatial and temporal components and determining the modeling coefficients based on the underlying fields actually processed, this thesis presents a novel approach to the estimation and restoration of displacement maps which has escaped previous solutions. The restoration process derived is computationally efficient and allows for parameter modification for adaptive filtering.

3.2. Motivation for Model Based Restoration

In typical images, neighboring pixels that correspond to the same object have a strong correlation in the intensity domain. At each pixel in the depth or displacement field, an estimate of the depth or velocity is produced. These values are estimated from the corrupted intensity images, so that the estimates themselves are subject to degradations. In typical scenes, the depth or displacement along a body varies slowly for incremental spatial changes along its surface. Since the depth or displacement field of an imaged object forms a dense sampling grid, neighboring values in these fields are also assumed to have a strong correlation. It is this correlation between adjacent locations in the depth and displacement fields that motivates a model-based solution for the restoration of the corrupted field.

This restoration problem bears some similarity to the situation faced in intensity image restoration where the restored image is estimated from the corrupted intensity values given as an input. Model based restoration has been employed for many years in work with intensity image restoration [6, 7, 36, 60, 67]. Reduction in the size of the signal state has led to computationally efficient identification of model parameters and rapid

filtering techniques [6]. The type of modeling employed by this thesis is an extension of the successful models that have been developed for intensity image restoration.

In dealing with depth and displacement fields that contain discontinuous regions, an adaptive model process is used to restore these areas. In model-based image restoration, the filter parameters can be adjusted or adapted to account for changes in the image spatial content. The problems encountered with discontinuous regions and boundary smoothing motivate an adaptive parameter identification approach for depth and displacement estimation and restoration. Although random image capture noise can be handled by this process, blurring in the input intensity images, which results in a highly nonlinear effect on the displacement map, is beyond the scope of this thesis. Pre-filtering to restore the intensity images, however, may be used to reduce the blurring effects, as proposed in [60], before the displacement estimation stage. By properly defining the modeling processes for the depth and displacement maps, the problem of restoration can be addressed with well-established multi-dimensional filtering techniques.

3.3. Literature Review

Biemond *et al.* [11] propose a two-frame pel-recursive Wiener-based algorithm to estimate displacement for image sequences. They rewrite the constant displaced intensity assumption (1) and displaced frame difference equation (2) in terms of frame t and the previous frame t-T resulting in:

$$I(x,y,t) = I\left(x - d_x, y - d_y, t - T\right) \tag{29}$$

$$dfd(x,y;\hat{d}_x,\hat{d}_y) = I(x,y,t) - I(x-\hat{d}_x, y-\hat{d}_y, t-T) \tag{30}$$

where $\hat{\mathbf{d}} = (\hat{d}_x,\hat{d}_y)$ is an estimate of the true displacement vector, $\mathbf{d} = (d_x,d_y)$. Applying an initial displacement estimate, $\hat{\mathbf{d}}^{i-1} = (\hat{d}_x^{i-1},\hat{d}_y^{i-1})$, to equation (30) yields:

$$dfd(x,y,\hat{d}_x^{i-1},\hat{d}_y^{i-1}) = I(x,y,t) - I(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T) \tag{31}$$

Substituting (29) into (31) results in:

$$dfd(x,y,\hat{d}_x^{i-1},\hat{d}_y^{i-1}) = I(x-d_x, y-d_y, t-T) - I(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T) \tag{32}$$

Taking a Taylor series expansion of $I(x-d_x, y-d_y, t-T)$ about location $(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1})$ yields:

$$\begin{aligned} I(x-d_x, y-d_y, t-T) &= I(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T) \\ &\quad -(\mathbf{d}-\hat{\mathbf{d}}^{i-1})^T \nabla_x f(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T) + v(x,y,\hat{\mathbf{d}}^{i-1}) \end{aligned} \tag{33}$$

where $\nabla_x f(\cdot)$ represents the spatial gradient operator and v is considered a stochastic process that models the truncation error from linearization. Substituting (33) into (32) produces an observation equation:

$$\begin{aligned} dfd(x,y,\hat{\mathbf{d}}^{i-1}) &= -(\mathbf{d}-\hat{\mathbf{d}}^{i-1})^T \nabla_x f(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T) + v(x,y,\hat{\mathbf{d}}^{i-1}) \\ &= -\nabla_x^T f(x-\hat{d}_x^{i-1}, y-\hat{d}_y^{i-1}, t-T)\cdot(\mathbf{d}-\hat{\mathbf{d}}^{i-1}) + v(x,y,\hat{\mathbf{d}}^{i-1}) \end{aligned} \tag{34}$$

where $\mathbf{d}-\hat{\mathbf{d}}^{i-1}$ is assumed to be a sample of a stochastic process and the spatial gradient operator is viewed as a known transition vector. An update $\mathbf{u}$ is defined as $\mathbf{u} = \mathbf{d}-\hat{\mathbf{d}}^{i-1}$ and (34) is applied to a neighborhood of N points which produces a set of equations through which a Wiener-based estimator is derived for the displacement update estimate, $\hat{\mathbf{u}}$. Finally, the new displacement vector is estimated by:

$$\hat{\mathbf{d}} = \hat{\mathbf{d}}^{i-1} + \hat{\mathbf{u}} \tag{35}$$

Biemond *et al.* use the estimated displacement of the previous pixel as the initial estimate $\hat{\mathbf{d}}^{i-1}$. Their algorithm converged quicker than previously reported algorithms, but boundary effects and noise sensitivity were not reported. No other spatial support for this estimate was proposed. Further, discontinuities in the intensity image may cause trouble in the dfd(·) calculation and in the estimation of the update term.

An extension to incorporate multiple frames into the Wiener-based displacement estimator of (34) was presented by Efstratiadis and Katsaggelos [20]. Their algorithm is similar to that presented in [11], except that equation (34) is observed over a neighborhood of N pixels in the previous v frames. A 3-D autoregressive (AR) model with fixed model parameters was proposed to estimate the initial displacement. The estimate is sensitive to the initial displacement vector since v past frames are utilized. Sensitivity to image discontinuities and to motion boundaries presents even greater problems if motion compensation is used to develop an initial estimate based on v past frames.

Several researchers propose a variety of models for the depth and displacement fields. Among these Matthies in [42, 43] proposes the use of a 1-D Kalman filter to track depth estimates based on known camera rotations and translations. Unfortunately, no treatment of spatial discontinuities or spatial correlation appears with his work. Instead, a piece-wise continuous spline under tension, presented by Terzopoulos in [61], smoothes the results in the spatial domain and bilinear interpolation is used to predict the new depth value.

Stuller and Krishnamurthy in [57] propose a model of the displacement field that relates the spatial characteristics of the field along a single scan line as:

$$\mathbf{d}(i) = \Phi\mathbf{d}(i-1) + \mathbf{w}(i-1) \tag{36}$$

where d(i) is the displacement vector for column i of a single scan line, Φ is the transition matrix, and **w** is a random component used to describe the uncertainty of the modeling, with a covariance Q_w. A Kalman type filter is implemented to track the displacement values along a raster scan. This type of model has severe limitations in that the two dimensional spatial and the temporal characteristics of the displacement fields are not considered. In addition, results for parameter selection were limited to a few trivial applications.

Driessen at TU Delft, Netherlands, in [17-19] is presently investigating the 2-D AR model for the displacement field modeling, namely:

$$\mathbf{d}(x,y) = \sum_{(i,j)\in S} \mathbf{A}_{ij}\mathbf{d}(x-i, y-j) + \mathbf{v}(x,y) \tag{37}$$

where **A** is a set of coefficients for spatial support and **v** is a driving process. A nonlinear observation equation based on equation (1) is implemented in the filtering process. Additionally, a decoupled separable autocorrelation function is used for the displacement estimates and constant model parameters, **A**, are selected for the field. The nonlinear observation is very sensitive to discontinuities in the input intensity image.

3.4. Depth and Displacement Modeling

Section 2.5 showed that the depth field is composed of scalars and that the displacement field is vector valued at each location. Several papers [18, 19,

57] have proposed that the vector components can be decoupled and filtered separately. This operation is particularly valid when processing displacement fields computed from stereo pairs. The decoupling of the vector components allows for a series of modeling equations to be developed which are dependent only on scalar fields.

3.4.1. System Model

The system model describes the underlying characteristics of the depth and displacement field. In typical scenes, the depth or displacement of a body varies slowly for incremental spatial changes along its surface. Thus, it is reasonable to assume a model which relates a depth or displacement value at pixel location (x,y) to its neighboring values. To model the relationship between neighboring values, the field is assumed to be generated by the 2-D autoregressive model that is described by the following equation:

$$d(x,y) = \sum_{(k,l)\in S} c_{kl}(x,y)d(x-k,y-l) + w(x,y) \tag{38}$$

where $d(x,y)$ is the depth or displacement component at location (x,y), $c_{kl}(x,y)$ are the system model coefficients at location (x,y) for support region S, indexed by (k,l), and $w(\cdot,\cdot)$ is a noise term to account for inaccuracies in the modeling procedure. $w(\cdot,\cdot)$ is assumed to be an independent zero mean, white Gaussian process with statistical properties used to describe the field. This modeling is consistent with that which has been successfully applied by Jain [28] to the description of intensity images.

An example of a possible support region S, which may be extended to various neighboring values, is the $M_1 \times M_2 \times M_3{}^{th}$ order Non-Symmetric Half Plane (NSHP) shown in Figure 3.1.

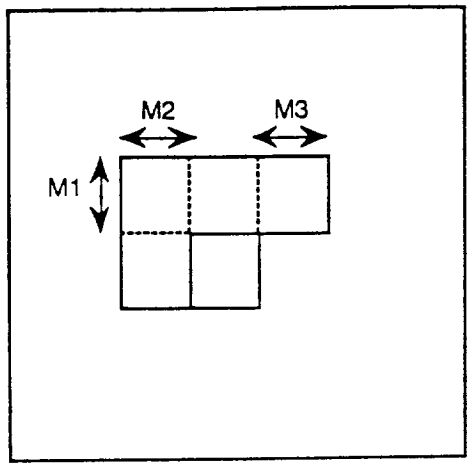

Figure 3.1 - Non-Symmetric Half Plane

The $M_1 \times M_2 \times M_3{}^{th}$ order Non-Symmetric Half Plane can be described by:

$$R_{\oplus} = \left\{ k,l \middle| \begin{matrix} [1 \le k \le M_2, 0 \le l \le M_1] \cup \\ [-M_3 \le k \le 0, 1 \le l \le M_3] \end{matrix} \right\} \tag{39}$$

where M_1 is the vertical extent of the support, M_2 is the left side extent and M_3 is the right side support extent.

The model coefficients, $c(\cdot)$, are generally space variant. In this thesis, a $1 \times 1 \times 1$ NSHP support region is utilized following a raster scan ordering. Specific reasons for the selection of the NSHP are due in part to its causality properties in filtering which is discussed in Section 3.5. When estimating homogeneous parameters for the entire field, the model coefficients are assumed to be wide-sense stationary

$$c_{kl}(x,y) = c_{kl} \quad \forall \ (x,y) \tag{40}$$

(38) may be re-written as:

$$d(x,y) = \sum_{(k,l) \in R_{\oplus}} c_{kl} d(x-k, y-l) + w(x,y) \tag{41}$$

To be sure of stability, the following constraint is assumed [6]:

$$\left| \sum_{(k,l)\in S} c_{kl} \right| < 1 \tag{42}$$

3.4.2. Observation Equation

An observation equation must be generated to describe the observation process for the field in order to complete the model needed for the estimation and restoration of depth and displacement fields. Several options are available from which to choose the most appropriate observation equation.

One option is presented if we obtain a depth or displacement map from an external source, such as in a block matching SSD algorithm or a gradient-based algorithm. This is referred to this as the **indirect** observation method for depth and displacement field estimation. It is indirect in that the observation is based on an estimation of the field at pixel location not directly involving the input intensity images. Mathematically this may be stated as:

$$d_o(x,y) = \hat{d}(x,y) \tag{43}$$

where d_o is the observed depth or displacement and $\hat{d}$ is the externally estimated value. In the calculations for the estimate of the depth and displacement by the external source, various errors are introduced due to intensity image noises, quantization, and image artifacts. We model these unknown errors as additive noise to the displacement estimate:

$$\hat{d}(x,y) = d(x,y) + v(x,y) \tag{44}$$

and $v(x,y)$ is assumed to be from an independent zero-mean Gaussian stochastic process with variance σ_v^2. In conventional notation, equations (43) and (44) can be combined as:

$$r(x,y) = d(x,y) + v(x,y) \tag{45}$$

where $r(x,y)$ is the observation at location (x,y), and v is an observation noise process.

The second type of observation equation will be referred to as the **direct** observation method. Here the displacement is recognized as an intrinsic component of equation (1). A non-linear observation model based on stereo frames can be described by:

$$I_r(x,y,t) = I_l\big(x - d_x(x,y,t), y - d_y(x,y,t), t\big) + n(x,y,t) \tag{46}$$

where $I_l(x,y,t)$ and $I_r(x,y,t)$ denote the intensity value at pixel location (x, y) at time t for the left and the right images respectively, and n accounts for the intensity image observation noise which is assumed to be independent, zero mean Gaussian noise. Care must be taken when applying equation (46) on noncontinuous intensity surfaces. One example of such a case occurs with the problem created with discontinuities between object and background. An intensity value sampled must be authenticated to verify that it belongs to the translated intensity region or an erroneous result will be produced.

Another possibility for direct measurement assumes that the displacement is estimated along a continuous intensity surface, as would be visualized in a shaded object or a landscape. Here we can rewrite equation (10) as:

$$\begin{aligned} \Delta I(x,y,t) &= I(x,y,t) - I(x,y,t+T) \\ &\equiv I_t \\ &\equiv -\left(\frac{\partial I(x,y,t)}{\partial x} \quad \frac{\partial I(x,y,t)}{\partial y}\right)\big(d_x(x,y,t) \quad d_y(x,y,t)\big)^T + v(x,y,t) \end{aligned} \tag{47}$$

where $d_x(x,y,t)$ and $d_y(x,y,t)$ are the x and y component of the displacement for location (x, y) at time t, and v accounts for uncertainty in the gradient and modeling approximations.

Equations (43), (46), and (47) deal with the observation of estimated displacement between image pairs.

If estimated depth is to be observed, as in a stereo setup, equation (4) must be incorporated into the observation equation. In this case, the observation equation contains only a single unknown, i.e., estimated depth.

With indirect observations, an external source provides the estimated disparity or displacement, $d_x(x,y)$, between the two stereo images, thus the depth observation can be written as:

$$d_p(x,y) = b\frac{f}{d_x(x,y)} + v(x,y) \tag{48}$$

where $d_p(x,y)$ is the depth estimate for pixel location (x,y), b is the baseline length, f is the focal length, and $v(x,y)$ is the observation noise process at location (x,y).

For direct observation of depth from the input intensity images, (3), (13), and (47) are combined to yield:

$$\Delta I(x,y) = \frac{1}{z(x,y)}\left[\frac{\partial I_1(x,y)}{\partial x} f\,T_x + \frac{\partial I_1(x,y)}{\partial y} f\,T_y\right] + v_i(x,y) \tag{49}$$

where $I_1(x,y)$ denotes the intensity value at pixel location (x,y) for the left image, $z(x,y)$ is the depth at location (x,y), T_x and T_y are the camera translations in the x and y directions, respectively, f is the camera focal length, and $v_i(x,y)$ accounts for uncertainty in the gradient and modeling

approximations. The image gradients in (49) are taken from the left image data.

By the geometry of the stereo cameras, $d_y(x,y)$ is identically equal to zero for all locations (x,y) in the displacement field. With these conditions, (49) can be simplified to:

$$\Delta I(x,y) = \frac{1}{z(x,y)} \frac{\partial I_l(x,y)}{\partial x} bf + v_i(x,y) \tag{50}$$

where $I_l(x,y)$ denotes the intensity value at pixel location (x,y) for the left image, $z(x,y)$ is the depth at location (x,y), b is the baseline width or translation in the x direction, f is the camera focal length, and $v_i(x,y)$ accounts for uncertainty in the gradient and modeling approximations. The image gradients in (50) are taken from the left image data.

3.5. ROMKF-based Field Estimation

Now that a modeling formulation for the underlying depth and displacement fields is developed, the filtering process can be examined. The Reduced Order Model Kalman Filter, ROMKF, is used to filter the depth and displacement maps based on the models presented in the previous sections. The ROMKF is an optimal filter applied to a sub-optimal state. This procedure allows for easier estimation of parameters, lower computational complexity, and greater facility for parameter adaptation, while producing results that are of comparable merit to those of more complex and time consuming recursive procedures [6]. The ROMKF type of filtering procedure is recognized as having been successfully applied to the restoration of intensity images and is shown to work effectively in its application with the models presented in this thesis.

The Kalman filter is based on modeling the dynamics of a system with a state-space model. In a single dimension, the state-space model is described by:

$$\mathbf{s}(m) = \mathbf{F}\mathbf{s}(m-1) + \mathbf{E}\mathbf{w}(m) \tag{51}$$

$$r(m) = \mathbf{H}\mathbf{s}(m) + v(m) \tag{52}$$

where $\mathbf{s}(m)$ is the signal space at time m, r is the observation or measurement of the system, $\mathbf{w}$ is a random process to account for uncertainties of the state model, v is the observation noise, and $\mathbf{F}$, $\mathbf{E}$, and $\mathbf{H}$ are system matrices. $\mathbf{F}$ and $\mathbf{H}$ are usually referred to as the state transition and observation matrices respectively. $\mathbf{w}$ and v are assumed to be uncorrelated zero-mean white Gaussian noise processes with covariances of σ_v^2 and Q_w respectively.

The Kalman filter extension into processing images of two dimensions as proposed by Woods and Radewan in [67] can be described by:

$$\mathbf{s}(m,n) = \mathbf{F}\mathbf{s}(m-1,n) + \mathbf{E}\mathbf{w}(m,n) \tag{53}$$

$$r(m,n) = \mathbf{H}\mathbf{s}(m,n) + v(m,n) \tag{54}$$

A raster scan format is assumed by ordering the pixels from left to right and top to bottom. This provides a horizontal direction of recursion where the scanline transition is from a pixel on the rightmost side of the image to the leftmost element on the next row is assumed. This ordering made on the 2-D image data provides the context of past, present, and future states. The past, present, and future pixels are defined with respect to the current location (x,y) as shown in Figure 3.2.

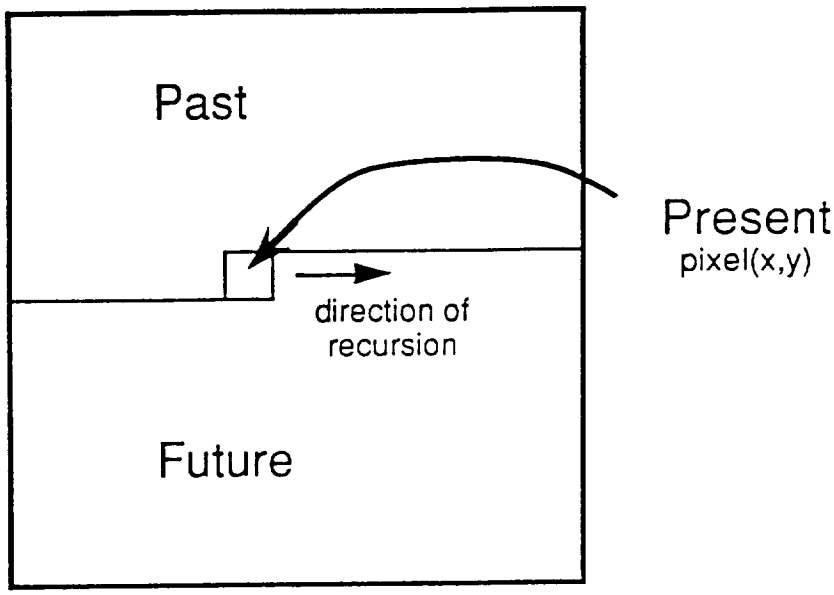

Figure 3.2 - Pixel Ordering by Raster Scan Format

We need to formulate our model describing the depth and displacement fields developed in Section 3.4 into a state space model notation. In equation (41), a NSHP is used for the support region of the underlying depth and displacement field. This yields:

$$d(x,y) = \sum_{(k,l)\in R_{\oplus}} c_{kl} d(x-k, y-l) + w(x,y) \tag{55}$$

where the spatial support region $R_{\oplus}$ is given in (39). The recursive properties of the Kalman filter and the ordering assumed by the raster scan format dictate the use of a causal field model such as provided by the NSHP support region.

Using the raster scan format, a field model described by (55), and a depth or displacement field of $N_h \times N_v$ pixels, in the horizontal and vertical dimensions respectively, the state $\mathbf{s}(m,n)$ at location (m,n) is described in [67] as:

$$
\begin{aligned}
\mathbf{s}(m,n) = \; & [d(m,n), d(m-1,n), \cdots, d(1,n); \\
& d(N_h, n-1), \cdots, d(1, n-1); \\
& \vdots \\
& d(N_h, n-M_1), \cdots, d(m-M_2, n-M_1); \\
& b(1-M_2, n-M_1+1), \cdots, b(0, n-M_1+1); \\
& \vdots \\
& b(1-M_2, N_v), \cdots, b(0, N_v); \\
& b(N_h+1, n-M_1), \cdots, b(N_h+M_3, n-M_1); \\
& \vdots \\
& b(N_h+1, N_v-1), \cdots, b(0, N_v-1)]^T
\end{aligned}
\tag{56}
$$

where $b(\cdot,\cdot)$ are boundary pixels and a $M_1 \times M_2 \times M_3{}^{th}$ order NSHP is used for the field model support. The dimension of this state vector (56) is on the order of $O[M_1 N_h]$. This state model support is shown in Figure 3.3.

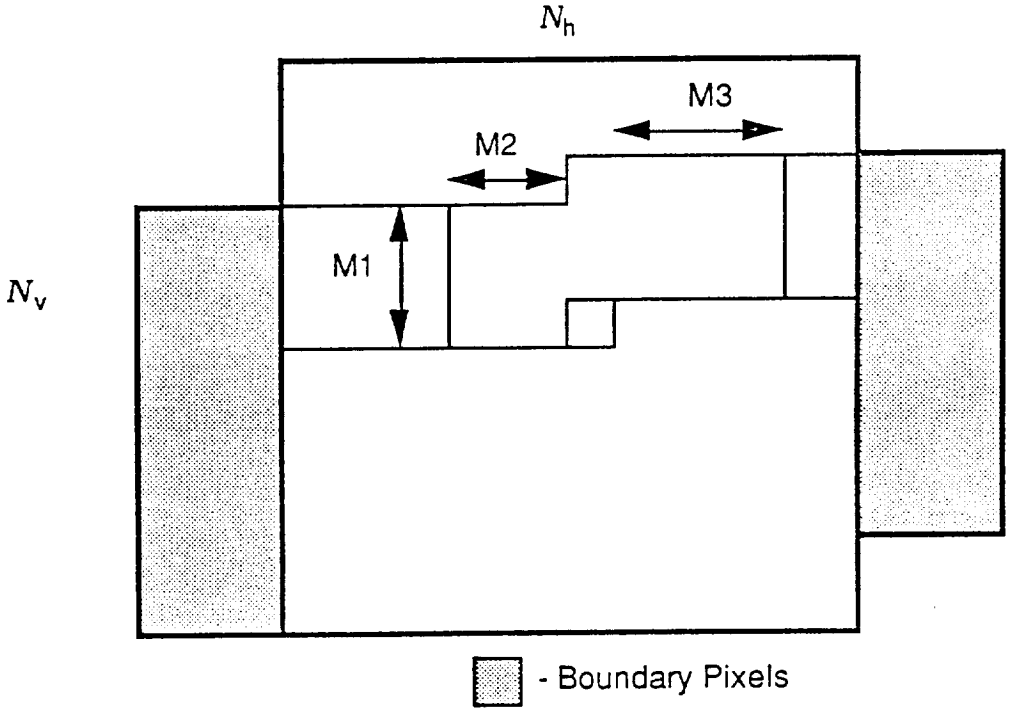

Figure 3.3 - NSHP State Model Support

With the above state vector, equation (53) has a state transition matrix **F** which is sparse, consisting mainly of shift operators which contain 1's and 0's.

The equations to implement a Kalman filter are described in Appendix A. The filter consists of a prediction and an update equation. The update equation requires the computationally intensive solution of a nonlinear equation to compute the filter gain coefficients. The number of computations in the filtering procedure that are required per pixel is on the order of $O\left[M_l^2 N_h^2\right]$. The amount of storage required to maintain the error covariance prediction and update for a state size similar to (56) is considerable. To reduce the computational complexity and the amount of storage required to implement the filter, several approximations are required.

In implementing the Reduced Update Kalman Filter, RUKF, Woods and Radewan [67] make the observation that the image pixels are not significantly correlated with the pixels outside of a certain region or neighborhood of a particular location. The error covariance update is reduced to contain only those pixels contained within this local neighborhood. While very successful results have been reported [59, 65, 67], the reduced error covariance update region is still larger than the model support region.

In an effort to substantially reduce the amount of computational complexity, Angwin and Kaufman in [7] describe an alternate state description that has a much lower dimension. The state size is reduced to the order of the support region utilized in equations (38) and (45). They

propose a Reduced Order Model state vector, based on the pixels actually required for the computation of (41) and (45), which can be described by:

$$\begin{aligned}\mathbf{s}(m,n) = [&d(m,n), d(m-1,n), \cdots, d(m-M_2,n);\\ &d(m-M_2,n-1), d(m-M_2+1,n-1), \cdots, d(m+M_3+1,n-1);\\ &\vdots\\ &d(m-M_2,n-M_1), d(m-M_2+1,n-M_1), \cdots, d(m+M_3+1,n-M_1);]^T\end{aligned} \tag{57}$$

where a $M_1 \times M_2 \times M_3{}^{th}$ order NSHP is used for the field model support. This state model support is shown in Figure 3.4.

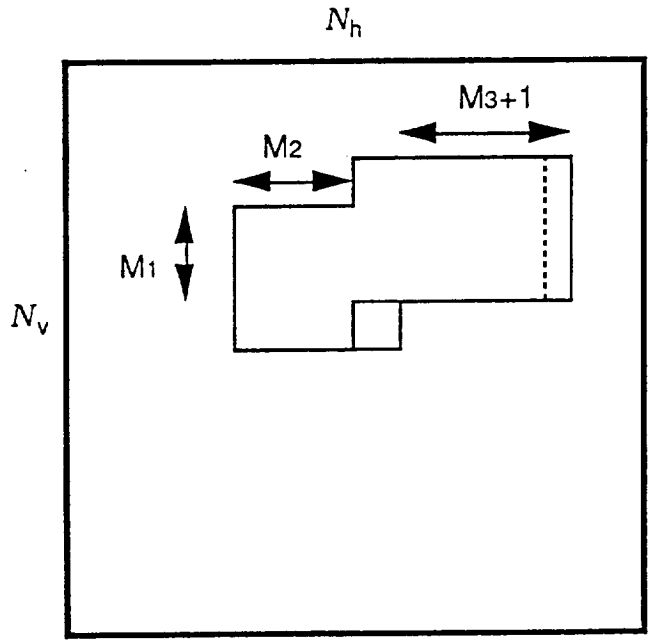

Figure 3.4 - ROM State Model Support

As an illustration, the reduced order model for the depth or displacement field state for the $1 \times 1 \times 1$ NSHP support may be written as:

$$\mathbf{D}(x,y,t) = \begin{bmatrix} d(x,y,t) \\ d(x,y-1,t) \\ d(x+1,y-1,t) \\ d(x+2,y-1,t) \\ d(x-1,y,t) \end{bmatrix} \tag{58}$$

where $d(x,y,t)$ is the depth or a single component of the displacement field.

The dimension of this state vector (57) is on the order of $O[M_1(M_2+M_3)]$. The state vector, $\mathbf{s}(m,n)$, as written in (57), contains elements that can not be written in terms of the previous state, $\mathbf{s}(m-1,n)$. This is the basis for the ROMKF procedure where these elements are approximated by their most recent estimate [7]. This approximation is made as:

$$s(m+M_3+1,n-j) = \hat{s}(m+M_3+1,n-j) + w_j(m,n) \tag{59}$$

where $\hat{s}(\cdot,\cdot)$ is the best available estimate of the field and w_j is a noise term to account for approximation uncertainties. The most recent update of the elements at the time that pixel (m,n) is filtered is taken as the best available estimate. For example: the $1\times1\times1$ NSHP case requires approximation for only a single term.

$$d(x+2,y-1,t) = \hat{d}(x+2,y-1,t) + w_1(x,y,t) \tag{60}$$

The previous estimate used in the approximation is included in the state model equations, (53) and (54), as a deterministic input term $\mathbf{u}$. Thus the ROMKF state equations become:

$$\mathbf{s}(m,n) = \mathbf{F}\mathbf{s}(m-1,n) + \mathbf{G}\mathbf{u}(m,n) + \mathbf{E}\mathbf{w}(m,n) \tag{61}$$

$$r(m,n) = \mathbf{H}\mathbf{s}(m,n) + v(m,n) \tag{62}$$

where $\mathbf{F}$, $\mathbf{E}$, and $\mathbf{H}$ are system matrices and $\mathbf{G}$ incorporates the ROM approximations into the state vector.

The ROMKF implements an optimal Kalman filter on the reduced signal state space. The lower dimensionality of the state allows for a more rapid computation of model coefficient parameters and associated gains calculated from an iterative solution of the discrete-time Riccati equation (A.9). This reduced state allows for parameter adaptation in a computationally efficient manner. Usually, the new Kalman gains can be found with relatively few additional iterations.

Simpson in [56] reviews the effects of varying the extent of the NSHP support. She reports that filter performance was improved by moderate extensions of the left side of the support region, while extensions to the right side resulted in negligible performance improvements over the reduced order model described by (57). Taking into account the results which Simpson presents and the computational complexity of larger models, a $1\times1\times1$ NSHP will be used as the basis for the ROM state. As mentioned previously, with the use of the ROM state vector, approximations are needed for the components in the state that can not be modeled by terms in the previous state. For the research conduced in this thesis, the most recent estimate is used as the approximation for those terms not represented in the previous state.

Since blurring is not to be considered, the $\mathbf{H}$ matrix reduces to a selector matrix picking out the appropriate component of $\mathbf{D}(x,y)$. The state models for the $1\times1\times1$ NSHP for spatial displacement filtering can be written as:

$$\mathbf{D}(x,y,t) = \mathbf{F}\mathbf{D}(x-1,y,t) + \mathbf{G}\mathbf{u}(x,y,t) + \mathbf{E}\mathbf{w}(x,y,t)$$
$$r(x,y,t) = \mathbf{H}\mathbf{D}(x,y,t) + v(x,y,t) \tag{63}$$

$$F = \begin{bmatrix} c_{-10} & c_{1-1} & c_{0-1} & c_{-1-1} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$
$$E = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad H = [1 \;\; 0 \;\; 0 \;\; 0 \;\; 0] \tag{64}$$

$$\mathrm{var}(w) = Q_w \qquad \mathrm{var}(v) = \sigma_v^2$$

where $\mathbf{F}$, $\mathbf{E}$, and $\mathbf{H}$ are system matrices, $\mathbf{G}$ incorporates the ROM approximations into the state vector, and the deterministic input $\mathbf{u}$ contains the most recently updated previous estimate of depth or displacement values for the terms in the current state that can not be written in terms of the previous state. For example: the $1 \times 1 \times 1$ NSHP case requires only a single term:

$$d(x+2,y-1,t) = \hat{d}(x+2,y-1,t) + w_1(x,y,t) \tag{65}$$

Hence the deterministic input is:

$$\mathbf{u}(x,y,t) = \left[\hat{d}(x+2,y-1,t)\right] \tag{66}$$

3.6. 2-D Spatial and 3-D Spatio-Temporal

If a sequence of displacement maps is available, it is possible to extend the spatial ROMKF filter into a spatial-temporal ROMKF by the inclusion of

previously filtered information into the state models. The temporal contribution, included in the deterministic input from the previous frame's data, has been filtered and is available. Care must be taken to make sure that the previous data are taken from the appropriate location and are tested for erroneous results. There are two special cases to consider for the temporal component term. The first case is that the displacement vector is correlated with the previous frame at the same location. For example, this occurs in a sequence of stereo images in which there is little or no motion in the scene, and leads to a steady-state displacement map for that scene. The deterministic input as shown in (63) for stationary temporal displacements can be described as:

$$\mathbf{D}(x,y,t) = \mathbf{F}\mathbf{D}(x-1,y,t) + \mathbf{G}\mathbf{u}(x,y,t) + \mathbf{E}\mathbf{w}(x,y,t)$$
$$\mathbf{r}(x,y,t) = \mathbf{H}\mathbf{D}(x,y,t) + v(x,y,t)$$

$$\mathbf{u}(x,y,t) = \begin{bmatrix} \hat{d}(x+2,y-1,t) \\ \hat{d}(x+1,y,t-1) \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 0 & c_t \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{67}$$

where c_t is the temporal coefficient.

The other case occurs if the scene contains dynamic or moving objects. In this case, the current state's displacement value is better represented by compensating for the motion of the fields. Here we are assuming a relatively constant velocity model to describe the change in the field from frame to frame, i.e., little or no acceleration. Denoting $\left(\hat{d}_x^-, \hat{d}_y^-\right)$ as the previous displacement estimate for pixel (x,y), the deterministic input can now be written as:

$$\mathbf{D}(x,y,t) = \mathbf{F}\mathbf{D}(x-1,y,t) + \mathbf{G}\mathbf{u}(x,y,t) + \mathbf{E}\mathbf{w}(x,y,t)$$
$$r(x,y,t) = \mathbf{H}\mathbf{D}(x,y,t) + v(x,y,t)$$

$$\mathbf{u}(x,y,t) = \begin{bmatrix} \hat{d}(x+2,y-1,t) \\ \hat{d}\left(x+1-\hat{d}_x^-, y-\hat{d}_x^-, t-1\right) \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 0 & c_t \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{68}$$

In equation (68) interpolation of interpixel displacement values may be required if $\left(\hat{d}_x^-, \hat{d}_y^-\right)$ does not fall on a lattice location. Alternatively, the motion compensation can be rounded to the nearest lattice location, thus avoiding the interpolation of interpixel displacement values.

3.7. Summary

A model for describing the depth and displacement field is presented. An optimal Kalman type filter, which is based on a sub-optimal state space approximation to the models, is used. The filter, ROMKF, has the benefits of lower dimensionality thus allowing for simpler parameter estimation, adaptive filtering, and less computational complexity. Extensions to the 2-D Kalman filter with the inclusion of a temporal component are discussed.

CHAPTER 4

4. Extensions and Implementation Issues

4.1. Introduction

In this chapter various extensions and implementation issues related to the basic modeling and filtering procedure are addressed. Due to the properties of the ROMKF, a lower order system state vector, compared to the full state vector, is required and thus fewer parameters need to be calculated and stored. The lower order system state vector reduces the computational complexity of homogeneous field parameter identification and requires smaller storage requirements for a set of parameters tuned to filter the discontinuities in the underlying depth or displacement field.

Section 4.2 presents methods that may be used for parameter identification of the model coefficients. Section 4.3 investigates several sources of degradations and errors in the estimation of depth and displacement fields that may be obtained from the corrupted intensity images. This provides a better understanding of the error terms included in the modeling process.

Issues related to processing images representative of typical robotic scenarios are addressed in Section 4.4. In such scenes, many objects have well defined boundaries which should be retained both in the intensity images and in the depth and displacement fields.

Finally, since the fields contain typically large data sets, such as images of 256x256 or 512x512 pixels, the filter procedure using methods of parallel computation is discussed in detail.

4.2. Parameter Identification and Filter Selection

In Chapter 3, a set of models is provided to describe the underlying depth or displacement field and the observation process. In most situations in which the depth or displacement field is estimated from the noisy intensity images, the system model parameters, $c_{kl}(x,y)$ in (38), and statistics on the model and observation noise processes are unknown and must be estimated. Several methods have been proposed to estimate these values.

When estimating field dependent parameters, the model coefficients for use in the ROMKF must be determined from the corrupted depth or displacement fields. A ROMKF/maximum likelihood parameter identification method incorporating nonstationary models for adaptive restoration is proposed by Angwin and Kaufman in [8]. Lagendijk in [37, 38] makes use of an iterative procedure based on expectation-maximization (EM) to simultaneously identify spatially variant coefficients and restore noisy blurred images. A least mean squares approach presented by Gelb in [22] can be used over the field to estimate a set of non-adaptive homogeneous system model parameters.

In the work presented in this thesis, the homogeneous system model parameters are determined by a least mean squares fit to the entire field as done by Kaufman *et al.* [32]. The computation is carried out over the field that contains valid depth or displacement data. Recalling (41), the system equation can be rewritten as:

$$d(n) = \mathbf{C}^T\mathbf{d}(n-1) + w(n) \tag{69}$$

where $d(n)$ is the depth or displacement at the current location (x,y), $\mathbf{C}$ is a 4x1 vector with the system coefficients for the 1x1x1 NSHP support region,

$\mathbf{d}(n-1)$ is a 4x1 vector with the depth or displacement values for the support region, and $w(n)$ is a modeling error term. The estimate of $\mathbf{C}$, denoted $\hat{\mathbf{C}}$, is found by:

$$\hat{\mathbf{C}} = \min_{\mathbf{c}} \sum \left(d(n) - \mathbf{C}^T \mathbf{d}(n-1)\right)^2 \tag{70}$$

where the summation is taken over the entire depth or displacement field values. The least squares estimate of $\mathbf{C}$ can be determined by:

$$\nabla_{\hat{\mathbf{c}}} \left[\sum \left(d(n) - \hat{\mathbf{C}}^T \mathbf{d}(n-1)\right)^2 \right] = 0 \tag{71}$$

The solution of which yields:

$$\hat{\mathbf{C}} = \left(\sum \mathbf{d}(n-1)\mathbf{d}(n-1)^T \right)^{-1} \sum \mathbf{d}(n-1)d(n) \tag{72}$$

Recalling the observation equation (45), only noisy observations of $d(n)$ and $\mathbf{d}(n-1)$ are available for the calculation of (72). These noisy observations bias the parameter estimate, $\hat{\mathbf{C}}$. By incorporating the noisy observations, taking expectations, and following the above least squares approach, the estimate of $\mathbf{C}$ with bias removal is [32]:

$$\hat{\mathbf{C}} = \left(\sum \mathbf{d}_o(n-1)\mathbf{d}_o(n-1)^T - N\sigma_v^2 \right)^{-1} \sum \mathbf{d}_o(n-1)d_o(n) \tag{73}$$

where N is the summation count, σ_v^2 is the variance of the observation noise process in (45), and $d_o(n)$ and $\mathbf{d}_o(n-1)$ are the noisy observations of $d(n)$ and $\mathbf{d}(n-1)$ respectively.

Tekalp in [60] proposes a method for adaptive filtering, based on the use of multiple parameter sets, called the multiple-model approach. While the least squares method makes use of the corrupted fields to estimate the system model parameters, the multiple-model method is made independent of the field by establishing a set of fixed parameters based on the underlying

content of representative field variations. Due to the high computational cost involved with the parameter identification and space-variant filtering procedures, Tekalp proposes the use of L *a priori* designed filters to adapt the restoration to the variations of the underlying fields. The identification of adaptive parameters is reduced to the less computationally intensive detection problem of filter bank selection. A finite bank of model parameters is pre-calculated and a selection is made as to which model "best applies" to the given observation window at each pixel location. At the filtering stage, Tekalp implements a maximum *a posteriori* probability (MAP) logic procedure to identify which model to use at each pixel location. Figure 4.1 gives a block diagram of the processing done at each pixel, where $\mathbf{r}(x,y)$is the observation window at location (x,y) and $\hat{\mathbf{s}}(x,y)$is the updated state.

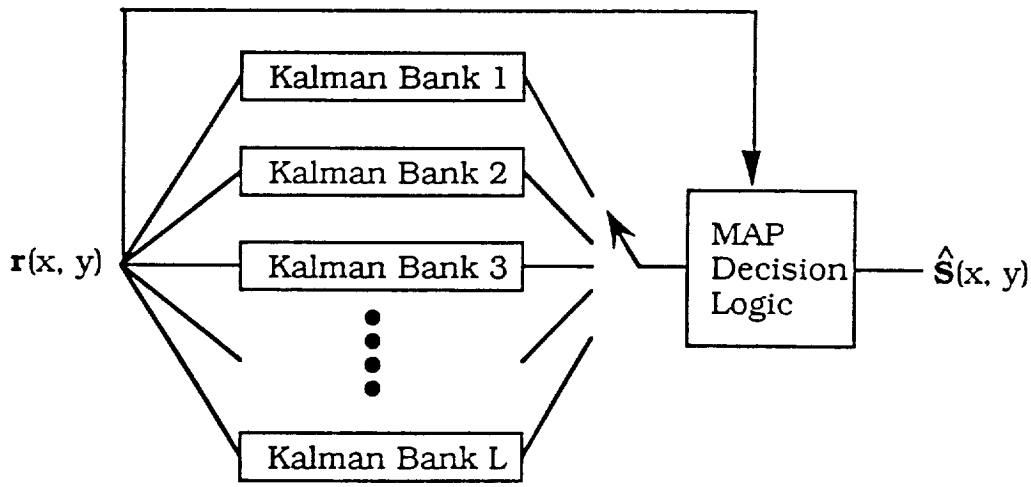

Figure 4.1 - Multiple-model Kalman Filter

Tekalp has reported very successful restoration results for intensity images using the multiple-model approach over space-invariant models at a fraction of the computational cost of a continually adaptive identification/restoration procedure. A significant extension of the multiple-model approach tailored

for depth and displacement restoration is a portion of the work associated with this thesis and is presented in Section 4.4.

A useful property of the models presented in Section 3.6 for the 3-D ROMKF is that the same program can be used to model the depth or displacement field in the spatial or temporal domain, either independently or collectively. If the spatial support, $\{c_{11}, c_{01}, c_{-11}, c_{10}\}$, where c_{kl} as described in (38), is set identically to zero, the filtering is reduced to temporal only. If the temporal coefficient c_t, described in (67), is forced to zero, the filter becomes a spatial filter only. This allows for an efficient performance comparison between the spatial and temporal filters currently presented in the literature.

Once the system model coefficients have been estimated with any of the procedures presented above, the Kalman gains for the update equations, provided in the Appendix as (A.6) and (A.7), can be calculated. In this thesis, the Kalman gains are found by an iterative solution [31] to the Riccati equation (A.9). This solution can be written as:

$$\begin{aligned} \mathbf{P}_-(y,x) &= \mathbf{F}\mathbf{P}_+(y-1,x)\mathbf{F}^T + \mathbf{E}\mathbf{Q}_w\mathbf{E}^T \\ \mathbf{K}(y,x) &= \mathbf{P}_-(y,x)\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_-\mathbf{H}^T + \mathbf{Q}_v\right)^{-1} \\ \mathbf{P}_+(y,x) &= \left(\mathbf{I} - \mathbf{K}(y,x)\mathbf{H}\right)\mathbf{P}_-(y,x) \end{aligned} \tag{74}$$

where $\mathbf{P}_+$ and $\mathbf{P}_-$ are the updated and predicted covariance matrices for the system state vector, $\mathbf{F}$, $\mathbf{E}$, and $\mathbf{H}$ are system matrices described in (61) and (62), $\mathbf{I}$ is the identity matrix of the appropriate dimension, and $\mathbf{Q}_w$ and $\mathbf{Q}_v$ are the covariances for the system model and observation noise, respectively. The iteration terminates when the updated covariance matrix converges to a

constant matrix. The Kalman gains, **K**, found in (74) are kept constant during the filtering process of the depth and displacement fields.

4.3. Degradations in Depth and Displacement Fields

This section covers the statistical methods used to determine the noise processes included in the system and observation models. Errors in the gradient calculation can cause inaccuracies in the displacement estimates and, in the case of indirect observations, the externally supplied estimates are sensitive to the noise of the intensity image. The distribution of the depth field is investigated using images obtained from a stereo camera setup and the estimated displacement field.

4.3.1. Modeling Noise Variance Determination

In Section 4.2, methods to determine the system model coefficients were presented. Once a set of coefficients has been estimated, the statistics of the system modeling error, **w**, can be estimated. An estimate of the variance of the modeling error, $\hat{\sigma}_w^2$, is determined by using the system model (41) and the observation equation (45) as shown by Angwin and Kaufman in [8].

$$\hat{\sigma}_w^2 = \frac{1}{N}\sum\left[\left[d(x,y) - \sum_{(k,l)\in R_{\bullet\cdot}} c(k,l)\,d(x-k,y-l)\right]^2 - N\left(\sum_{(k,l)\in R_{\bullet\cdot}} c^2(k,l) + 1\right)\sigma_v^2\right] \quad (75)$$

where $d(\cdot)$is the distorted depth or displacement value, σ_v^2 is the variance of the observation noise, $c(\cdot)$ is the system model coefficients for the NSHP support region, and N is the number of locations (x,y) used in the summation. The variance of the ROM approximation [6] described in (60) was set to $\sigma_{w_1}^2 = \frac{1}{4}\sigma_w^2$

For a field with negligible observation noise, (75) can be simplified to:

$$\hat{\sigma}_w^2 = \frac{1}{N}\sum\left[d(x,y) - \sum_{(k,l)\in R_\bullet} c(k,l)\,d(x-k,y-l)\right]^2 \qquad (76)$$

4.3.2. Observation Noise Variance Determination

An estimate of the variance of the observation noise process can be made on a region of the depth or displacement field known *a priori* to originate from a constant field. Such regions would occur in the displacement field for an image section undergoing uniform translation or in the depth field for a flat portion of an object viewed normal to the stereo cameras. An estimate of the variance of the observation noise, $\hat{\sigma}_v^2$, can be determined by:

$$\hat{\sigma}_v^2 = \frac{1}{N}\sum_{(i,j)\in W}\left[d(i,j) - \bar{d}\right]^2 \qquad (77)$$

where N locations are taken from the assumed uniform area W, and $\bar{d}$ is the local mean value for the area.

A lower bound on the variance of the observation noise process may be developed from a statistical treatment of the noise process in the intensity images. The bound is derived following the procedure outlined by Van Trees in [62].

In Chapter 2 the correspondence for left and right frames of a stereo camera setup was established. In the ideal setup an imaged object appearing at $I_l(x,y,t)$ would also appear at $I_r(x-d(x,y,t),y,t)$, where $I_l(x,y,t)$ and $I_r(x,y,t)$ are the intensity values at location (x,y) at time t for the left and right images respectively, and $d(x,y,t)$ is the displacement value that location. To find an estimate of the displacement, $\hat{d}(x,y,t)$, (the

measure of the correspondence between the images), an intensity error function, e, is given:

$$e(x,y,t;d') = I_r(x-d',y,t) - I_l(x,y,t) \tag{78}$$

where d' is a possible estimate for the displacement at location (x,y). The estimate for the displacement is taken as the value of d' which results in an intensity error closest to zero. Unfortunately, the left and right images are corrupted by noise. Using $I(x,y,t)$to denote the imaged scene signal, the left and the right stereo images may be expressed as:

$$I_l(x,y,t) = I(x,y,t) + n_l(x,y,t) \tag{79}$$

$$I_r(x,y,t) = I(x+d(x,y,t),y,t) + n_r(x,y,t) \tag{80}$$

where $n_l(x,y,t)$ and $n_r(x,y,t)$ are uncorrelated zero-mean Gaussian terms that account for the noise process in the left and right images with variances σ_l^2 and σ_r^2, respectively. Substituting (79) and (80) into the intensity error function (78) yields:

$$\begin{aligned} e(x,y,t;d') &= I_r(x-d',y,t) - I_l(x,y,t) \\ &= I(x-d',y,t) - I(x,y,t) + n_r(x-d',y,t) - n_l(x,y,t) \end{aligned} \tag{81}$$

Since $n_l(\cdot)$ and $n_r(\cdot)$ are uncorrelated zero-mean Gaussian processes,

$$e(x,y,t;d') = I(x-d',y,t) - I(x,y,t) + n(x,y,t) \tag{82}$$

where $n(\cdot)$ is a zero-mean Gaussian noise term with variance $\sigma_n^2 = \sigma_l^2 + \sigma_r^2$. To reduce sensitivity to image noise in calculating (82), a patch region is centered about each possible estimate d', and the error intensity function is evaluated for each location in the region. A 3x3 observation patch is shown in Figure 4.2.

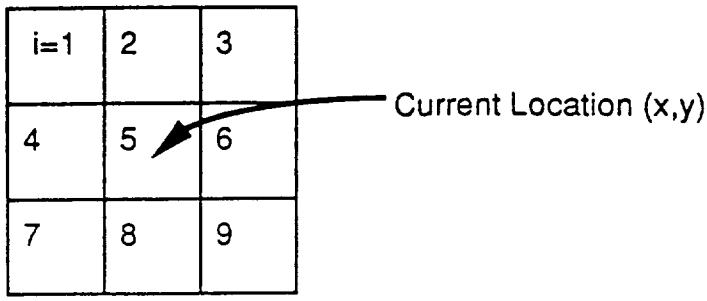

Figure 4.2 - Pixel Indices for 3x3 Observation Patch Region

Assuming that the displacement is constant within this region, a set of observations may be made. This may be written in the form:

$$\mathbf{R}_i = s_i(d) + \mathbf{N}_i \qquad i = 1,2,\ldots,W \tag{83}$$

where i is the pixel index within the observation patch region, random variable (RV) $\mathbf{R}_i$ is the measurement of the intensity error, $\mathbf{N}_i$ is the random noise process at index location i, and $s_i(d)$ is the intensity difference which is non-linear in the displacement term, d. With the given distribution of **N**, the conditional density function of **R** given d is:

$$f_{R|d}(r|d) = \left(\frac{1}{\sqrt{2\pi}\sigma_n}\right)^N \exp\left(-\frac{1}{2\sigma_n^2}\sum\left(r_i - s_i(d)\right)^2\right) \tag{84}$$

A lower bound based on the Cramér-Rao inequality for any unbiased estimate, $\hat{d}$, of the actual displacement d is given by Van Trees [62] as:

$$\mathrm{Var}\left[\hat{d}(\mathbf{r}) - d\right] \geq \frac{\sigma_n^2}{\sum_{i=1}^{W}\left[\frac{\partial s_i(d)}{\partial d}\right]} \tag{85}$$

where $\sigma_n^2 = \sigma_i^2 + \sigma_r^2$, and $s_i(d)$ is the intensity difference at pixel index i within the observation window centered at displacement d.

Using (82) and (83), the lower bound of the variance of the estimated displacement from a location in the left image to the right image is found to be:

$$\sigma_v^2 = \mathrm{Var}\left[\hat{d} - d\right] \geq \frac{\sigma_n^2}{\sum_{i=1}^{W}\left[\frac{\partial I_1}{\partial x}\right]} \tag{86}$$

For computational purposes, the gradient of the imaged scene, I, is estimated from the left image, I_1.

4.3.3. Density Distribution of Depth and Displacement

Applying the stereo camera setup described in Chapter 2, equation (4) relates the displacement found between the left and the right images to the 3-D location of the point on the object imaged. In particular, there is an interesting relationship between depth and displacement (disparity for a stereo camera) given by [26]:

$$Z = b\frac{f}{x_l - x_r}; \qquad f > 0, b > 0 \tag{87}$$

where b is the baseline between the cameras, f is the focal length and $(x_l - x_r)$ is the displacement between the images at location (x, y). This relationship allows for a direct mapping from a given displacement field to a depth field.

The question of two possible computation procedures arises. One is the issue of filtering then calculating, i.e., filtering the displacement first and then calculating the depth field, and the other is calculating then filtering, i.e., calculating the depth from displacement then filtering the depth field.

To answer this question, the following assumption is made: the displacement estimates follow a normal distribution with a mean m_s and variance σ_s^2. This assumption follows from the observation equation presented in (44) and (84). The RV S is used for the displacement random process with the notation $S \sim N(m_s, \sigma_s)$ to indicate that S is normal with mean m_s and variance σ_s^2. The density function of a RV X is indicated by $f_X(x)$. S has a density function described by:

$$f_S(s) = \frac{1}{\sigma_s \sqrt{2\pi}} e^{-\frac{(s-m_s)^2}{2\sigma_s^2}} \tag{88}$$

Since the displacement is a random process, one can clearly see that the depth will also be a random process. The RV Z is used to represent the depth random process. The relation (87) can be written in terms of RV S and Z as:

$$Z = g(S) = \frac{bf}{S} \tag{89}$$

where the baseline, b, and the focal length, f, are assumed to be known constants.

The Fundamental Theorem presented by Papoulis in [51] can be used to find the distribution of Z. The Fundamental Theorem determines the density of Z = g(S) in terms of the density of S. The theorem states that:

Denoting the real solutions of $z = g(s)$ by s_n:

$$z = g(s_1) = \cdots = g(s_n) \tag{90}$$

the density function of Z, $f_Z(z)$, in terms of the density function of S, $f_S(s)$, is:

$$f_z(z) = \frac{f_s(s_1)}{|g'(s_1)|} + \cdots + \frac{f_s(s_n)}{|g'(s_n)|} \tag{91}$$

where $g'(s)$ is the derivative of $g(s)$.

Following the above notation, a single solution of (89) is found:

$$s_1 = \frac{fb}{z} \tag{92}$$

and the gradient of $g(s)$ at that point equals:

$$\begin{aligned} g'(s) &= -\frac{fb}{s^2} \\ g'(s_1)\Big|_{s_1=\frac{fb}{z}} &= -\frac{fb}{(fb/z)^2} \\ &= -\frac{z^2}{fb} \end{aligned} \tag{93}$$

Combining (92) and (93) into (91), the density function of the depth estimates, $\mathbf{Z}$, becomes:

$$f_z(z) = \frac{fb}{z^2} f_s\left(\frac{fb}{z}\right) \tag{94}$$

The implications of this distribution, with several representative examples, and the answers to the computation procedure question are presented, with the appropriate experimental results, in Chapter 5.

4.4. Robotics and Rigid Body Considerations

In the area of robotic vision and assembly, many of the objects in the scene that contribute to the depth and displacement fields are known *a priori*. Most of the objects being constructed are rigid. In a particular application, a series of struts and nodes are assembled into a space based structure [14]. The struts provide a well defined structure that can be exploited in

processing the displacement fields. An image representative of a typical robotic scenario is shown in Figure 4.3.

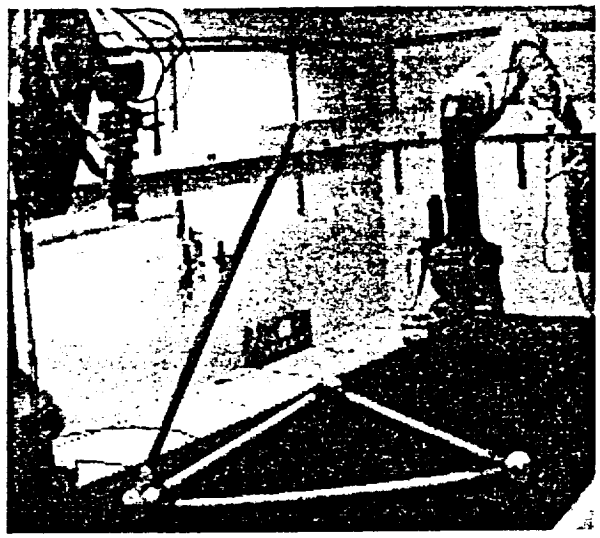

Figure 4.3 - Image Representative of a Typical Robotic Scenario

In processing rigid bodies, additional information can be extracted which provides for further processing and filtering of the depth and displacement fields. In robotic assembly applications, the construction typically deals with rigid objects of known shape and dimension. Some examples of such objects include sections on a robot arm, assembly struts, and various assembly-tool components. With these rigid bodies, information such as length, width, and geometric shape can be measured beforehand. This information is then available for processing fields that contain these items.

A simple example will demonstrate this more clearly. When the objects in the image are of known shape and length and their orientation can be determined, e.g., by feature points, this information can be added into

the depth computation. For example: if the object is a rigid beam of known length, the depth at one end of the section can be functionally related to the depth at the other end. Using this additional constraint, the number of unknowns can be reduced (e.g., depth at both end points can be reduced to only the depth at one end point).

As was previously mentioned, a problem exists with current methods used to estimate depth and displacement fields when confronted with discontinuities. A restoration algorithm that treats the field as a homogeneous continuous signal tends to greatly smooth out the edges in the fields. The exclusive use of homogeneous system model parameters used to filter fields that contain discontinuities would also be a problem. This is due to the fact that edge regions are treated in the same manner as continuous regions. In our robotic scenario, the smoothing of edges in the depth and displacement fields is undesirable in that the boundary areas of the object may be misidentified possibly causing the robot to strike the object.

A better solution is to provide a space variant system model and match the coefficients to the underlying field or local statistics. This brings up the issues of adaptive parameter identification. In the most general case, system model coefficients would be identified at each location in the depth or displacement field. Implementation of continuously adaptive image restoration is computationally intensive, and would not be appropriate for high-speed operations. This motivates the search for a solution based on detecting suitable model parameter sets rather than identifying the coefficients.

A series of models is constructed with each one based on a different direction for the discontinuity. Although a wide range of discontinuity directions is possible, the $1 \times 1 \times 1$ NSHP model support region is amenable to 4 major directions [59, 60]: 0°, 45°, 90°, and 135°. A non-edge model, with equal coefficient weights or a least squares fit to the field, is used for regions that do not contain detected discontinuities. Since the models are chosen before filtering, detection of the edges is computationally less intensive than identification of the model parameters. The detection of edges permits adaptive behavior in restoring the fields. Noise and distortions can be suppressed while minimizing the loss of edge information in the depth and displacement fields.

Now that a set of directional model parameters is established, the decision procedure may be described. In intensity restoration, the multiple-model method implemented by Tekalp uses a decision algorithm based on a local decision window. This decision window is shown in Figure 4.4.

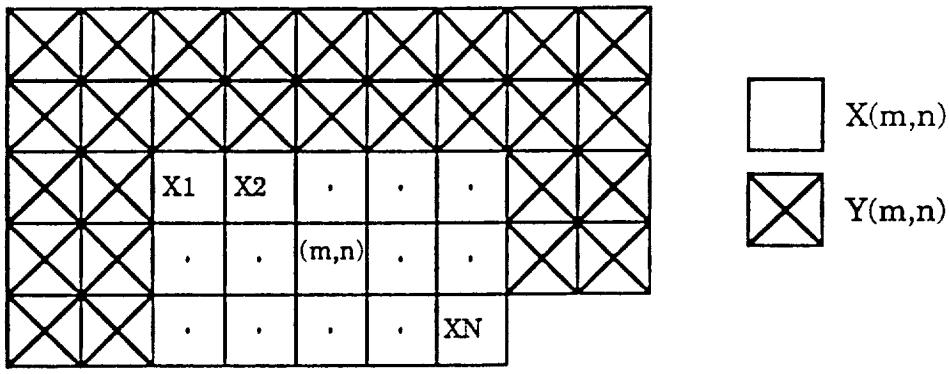

Figure 4.4 - Decision Window for Intensity Image Restoration

The decision is made based on the noisy observations $X(m,n)$ about the current location (m,n). $Y(m,n)$ is the boundary data consisting of the best past estimates and the future noisy pixels. An assumption is made that

there are no model transitions inside the decision window. A maximum *a posteriori* probability (MAP) decision logic is used to adaptively estimate the edge orientation based on the noisy observation X(m,n) in the intensity image.

For this thesis, the determination of which model applies to a given location in the field is based not only on the field that is being restored, but also on the additional information obtained from the input intensity images. This method differs from that proposed by Tekalp in that the direction of the discontinuity is not determined by the distorted field being restored by the use of a MAP decision, but that the direction is estimated by the edge operators applied to the input intensity images. The input intensity images are processed to provide information about discontinuity directions. The procedure begins by preprocessing the input intensity images to determine edge strength and orientation. The existence of edges in the intensity image indicates the possibility of edges or discontinuities in the depth or displacement field. Edges in the intensity image may appear from actual object boundaries or from changes in the surface characteristics such as abrupt changes in paint, tint, color, or reflectivity.

There are many proposed methods to identify edge strengths and directions of intensity images [26, 53]. One method that accomplishes this identification is the selection by maximal response from several convolution masks which are passed over the image [53]. Convolution masks are generated by rotating a base kernel in a clockwise direction. Several types of kernels can be used to determine edge strength and orientation. An

example of one, the Compass kernel and its associated masks, is shown in Figure 4.5.

$$\text{Compass Kernel} \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix}$$

Convolution Masks:

$$\mathbf{D}_1 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix}, \mathbf{D}_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix}, \mathbf{D}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \mathbf{D}_4 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix},$$

$$\mathbf{D}_5 = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}, \mathbf{D}_6 = \begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{D}_7 = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{D}_8 = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

Figure 4.5 - Compass Kernel and Its Convolution Masks

At each pixel, each mask is convolved with the image and the mask that has the largest response is taken as the edge direction and strength.

$$S(x,y) = \max_{j:1\to 8}\left[\text{convolve}\left(\mathbf{D}_j, I(x,y)\right)\right] \tag{95}$$

In looking at the masks given in Figure 4.5, $\mathbf{D}_1$ and $\mathbf{D}_5$ have large responses for vertical intensity image edges; thus a selection by (95) of either of these masks, indicates the presence of a 90° edge. Masks $\mathbf{D}_2$ and $\mathbf{D}_6$ respond to 45° edges, $\mathbf{D}_3$ and $\mathbf{D}_7$ respond to 0° edges, and $\mathbf{D}_4$ and $\mathbf{D}_8$ respond to 135° edges. A threshold is used to determine low contrast or non-edge regions.

The direction and strength are calculated by (95), and the results are saved for the filtering processing stage. The maximum kernel response is used in place of the MAP decision. In the actual filtering process, the edge map only gives locations for possible discontinuities. To determine if there is a discontinuity in the depth or displacement field, a threshold test is

conducted using the homogeneous model. The residue portion of the Kalman update equation (A.6) is used as a measure for discontinuities. The test is shown in (93):

$$\left| \left(d(x,y,t) - \sum_{(k,l)\in R_s} c(k,l)\, d(x-k, y-l, t) \right) \right| \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{96}$$

where H_1 is the hypothesis that a discontinuity is present, H_0 is the null hypothesis, $c(\cdot)$ is the homogeneous model coefficients for the $1 \times 1 \times 1$ NSHP model support region, and T is a threshold. This multiple-model filter for depth and displacement restoration is shown in Figure 4.6. The threshold is empirically determined based on the variance of the model and observation noise.

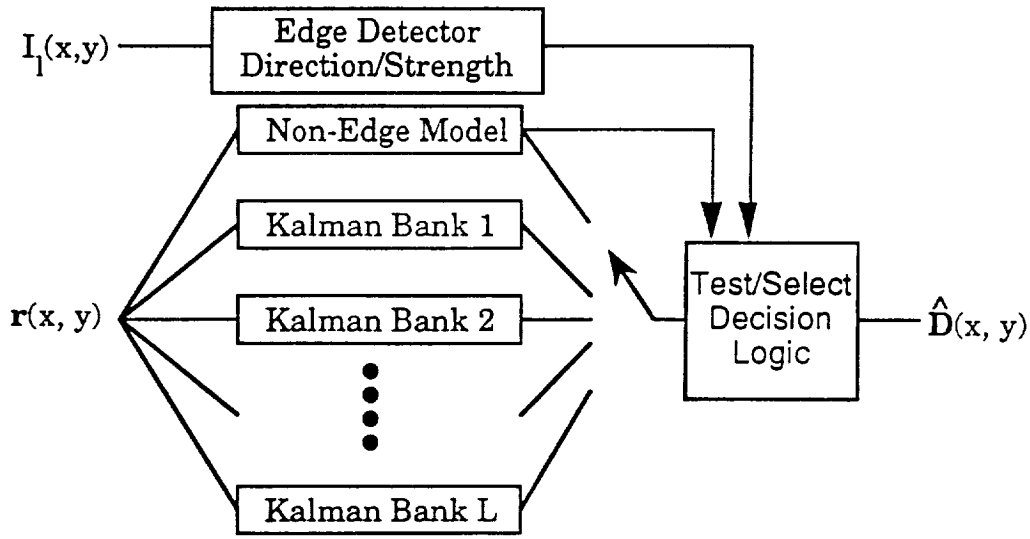

Figure 4.6 - Multiple-model Depth and Displacement Filtering

If the residue is found to be large and H_1 is accepted, the edge map is consulted to see if an edge model applies at that location in the field. If there is an edge detected at that location, the edge direction is used to select

appropriate parameters from the multiple-model set, and a new discontinuity compensated prediction is calculated. If a better fit is achieved, the edge directed model is used. Results from this procedure are effective on synthetic as well as the real test images containing discontinuities. Supportive experimental information is presented in Chapter 5.

4.5. Parallel Processing Implementation

Digital image processing involves a significant amount of data manipulation and calculations because of the nature of the data involved. Common image sizes range from 96 x 96 pixels to 1024 x 1024 pixels. In order to process that amount of data in a timely fashion, it is necessary to incorporate faster sequential computers, i.e., faster clock rate or more bandwidth, or break down the tasks into smaller sections that can be pipelined or computed in parallel.

The material presented in this section is described without restrictions to a particular parallel processing system, although it is particularly suited for implementation on multi-CPU shared bus systems. The method was originally proposed by Damour in [15]. The approach used to obtain a parallel version of the filtering process is to isolate the data independence in calculations. The general algorithm is described and results are provided on its implementation on a specific MIMD machine in Chapter 5.

The parallel algorithm begins by looking at the support regions for the model. If the prediction or update region incorporates all of the past pixels then the filtering must be done sequentially since each calculation

must use the most recently updated pixel values. This is due to the fact that no pixel may be processed until all of its past region, as defined in Figure 3.2, has been completely processed. If the support regions are reduced to a localized neighborhood, as used in the ROMKF, then parallelism is possible. Since digital images involve a great number of image points, it is advantageous to seek a method that processes the field in parallel.

To investigate the data independence of the ROMKF filtering procedure, recall the $1 \times 1 \times 1$ NSHP model support region. Assuming that the filtering begins in the top left hand corner and proceeds left to right and top to bottom, a diagonal of independent pixel calculations is formed as shown in Figure 4.7.

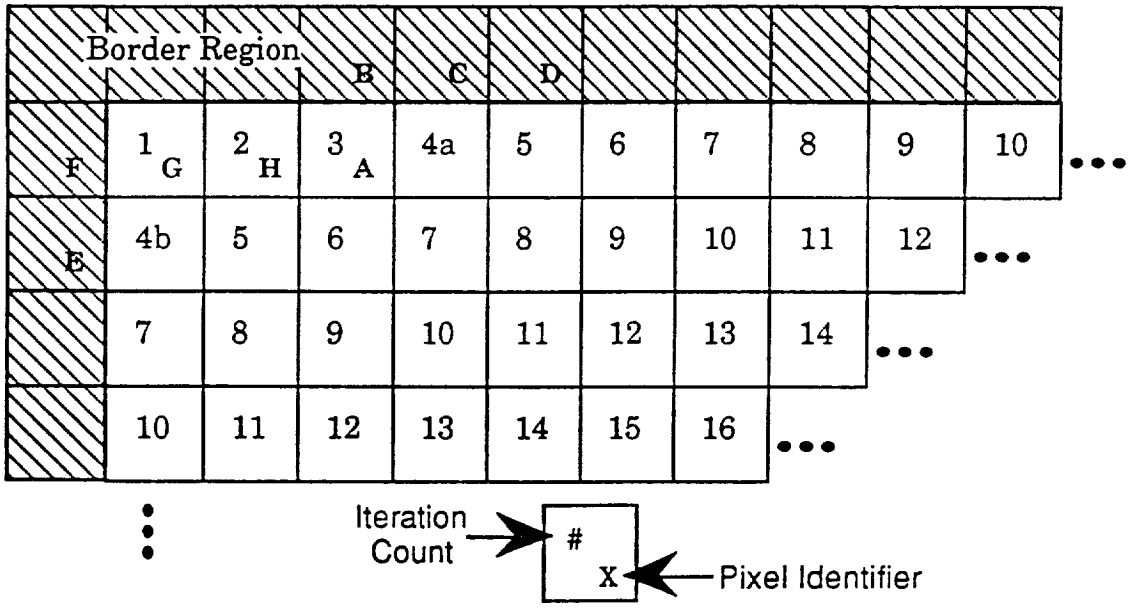

Figure 4.7 - Possible Parallelism in ROMKF

Each number represents the iteration number in which the pixel can be processed. Each iteration number process must wait until the smaller iteration values have completed. Those pixels with the same iteration

values may be processed concurrently. This can be seen by looking at the 4^{th} iteration. Pixel 4a's support region covers the pixel labeled **A** and the border pixels **B**, **C**, and **D**. Pixel 4b's support region covers the pixels labeled **G** and **H** and border pixels **E** and **F**. There is no overlapping in support regions thus 4a and 4b can be processed in parallel. At the 7^{th} iteration number, data independence implies that three locations may be processed in parallel.

The pattern can be generated for larger dimensional models by insuring that the top-most right pixel in the support region does not overlap with the bottom-most left pixel in the support region of the calculated pixel to the right. Clearly with larger support regions, the space between the pixels that can be processed independently must increase. This increase between processed pixels decreases the parallelism. The number of pixels that can be processed in parallel begins at a single location in the top left hand corner, increases to a maximum close to the diagonal of the image, and then decreases to a single location at the bottom right pixel of the image. Given that a precise ordering must be maintained due to the data dependencies between individual processed pixels, a parallel algorithm must totally control the sequence of operations.

The Row Method ROMKF is now described. The basic algorithm treats each row as a single processing element's (PE) job allocation. First, all of the PE's are placed on a queue awaiting for a message to be sent. This message will enable a PE to process a row in the field. Once a PE is assigned a row, the processor will filter every pixel from the left margin to the right margin. When the PE processes a "sufficient number" of pixels so

that no overlapping of model regions will occur, a message is sent and a flag is set for the next PE in the queue to begin processing the next row. Once a PE has processed the last pixel with respect to the right margin of the field, it places itself on the waiting list. When the last row with respect to the bottom margin has been processed, a flag is changed to EXIT, and all processors exit from the parallel ROMKF routine.

A "sufficient number" refers to the number of pixels that must separate two concurrent processing PE's so that the support regions do not overlap. Any overlapping would cause undetermined results since the order of pixel references and calculations would be random. Since the algorithm allocates a PE to process an entire row, additional care must be taken to prevent a PE that is processing in a future row from overtaking the support region of a PE operating in a past row. This is accomplished by requiring that all PE's complete their computations before the next parallel iteration begins.

Since the exact order of PE execution or effective execution is not known, due to bus and memory contention and PE time sharing, it is necessary to place a control on execution. To prevent any differences in PE execution order, all PE's that are allocated to process rows are synchronized after processing each pixel. Before a PE can move on to the next pixel, a synchronization occurs which assures that proper ordering is maintained and that the past iteration's computations are completed. The PE is guaranteed that the region it is using has been completely updated by all past PE's.

As an example, the following steps occur in processing a field with a $1\times1\times1$ NSHP model support region. $\mathbf{PE}_1$ begins processing ROW = 1. Once the processor completes the third pixel, a message is passed, enabling $\mathbf{PE}_2$ to begin ROW = 2. On the next iteration $\mathbf{PE}_1$ and $\mathbf{PE}_2$ are computing the filter in parallel. $\mathbf{PE}_2$ signals after processing the third pixel in the row, enabling $\mathbf{PE}_3$ to begin processing. The next PE on the queue will be allocated to process the next row, one at a time. If no PE's are available, as in the case of greater number of concurrent pixels present to be processed than the number of PE's allocated to execute the filter, that row will wait until the top most processing PE completes its row. Once the PE on the last row reaches the third pixel, it sets a flag that instructs each of the other PE's to exit as soon as it completes processing its row.

This algorithm is completely general for the model size as well as the field size. It can easily be extended to include various support regions by simply calling the algorithm routine with the maximum extent for each direction of the region. Supportive data which displays the reduction in execution time associated with the parallel row method compared to a single processor is given in Chapter 5.

A method to perform the multiple-model ROMKF method follows in a fashion similar to that described above. A finer parallelism technique, which makes the processing time independent of the number of filter models, is detailed in [15].

4.6. Summary

Identification of the parameters for the homogeneous model support is done on the corrupted field data. The statistics of the various noise processes in

the model and observation equations are evaluated and a lower bound based on a Cramér-Rao treatment of the estimation process is provided. To address the problems faced by oversmoothing of boundaries in the depth and displacement fields, a modified multiple-model procedure, based on edge content in the intensity images, is presented. Since the restoration process generally works with large amounts of data due to the size of input images, a parallel version based on data independence is detailed.

CHAPTER 5

5. Experimental Results

Presented in this chapter are the results obtained from applying the model-based restoration procedure of depth and displacement fields which was previously described in Chapters 3 and 4. A portion of the reported results involves the use of synthetic images, which were generated with known displacements, so that a controlled environment could be available to demonstrate the benefits of the applied method of restoration of the distorted fields. At the other end of the experimental spectrum, a set of real images, taken with a Panasonic M2.6 wv-BD400 camera with varied baseline widths, is presented in order to demonstrate the effects of restoration of fields generated from actual stereo images.

5.1. Synthetic Images and Known Displacement

In this set of experiments, a field is generated with known motion and the results obtained from an indirect measurement and restoration are given. These experiments can be considered as a basis of comparison since the exact motion is known throughout the field. The objective of the experiments in this section is to demonstrate the effectiveness of the model-based restoration process in restoring displacement fields estimated from noisy images in regions of low contrast. The SSD algorithm, described in Section 2.4.2, is used with varied sizes for the patch region utilized in the error measure. The effect of the patch region size on the improvements due to restoration is investigated. A ROMKF with a single spatial model is used in the restoration of the distorted fields. The model coefficients for these

experiments were found by a least squares fit of the $1 \times 1 \times 1$ NSHP support region to the estimated field. To quantify the restoration improvement, the dB improvement for the restoration of the distorted displacement field is defined as:

$$R_{dB} = 10 \log \frac{\sum_{(x,y) \in F} \left(d_o(x,y) - d(x,y)\right)^2}{\sum_{(x,y) \in F} \left(\hat{d}(x,y) - d(x,y)\right)^2} \tag{97}$$

where F is a region selected within the field, d_o is the corrupted displacement field estimated from the noisy images, d is the true field known *a priori* or calculated from the noiseless images, and $\hat{d}$ is the restored field.

To generate the test cases, images were produced with a linearly increasing intensity value along the horizontal direction. The equations that generate these images are:

$$I_l(x,y) = Gx + B + n_l(x,y) \tag{98}$$

$$I_r(x,y) = G(x + D_c) + B + n_r(x,y) \tag{99}$$

where $I_l(x,y)$ and $I_r(x,y)$ are the intensity values at pixel location (x,y) for the left and right images, respectively, G is the gradient value for the test image, B is the starting intensity, D_c is the constant displacement, and $n_l(x,y)$ and $n_r(x,y)$ are zero-mean uncorrelated Gaussian image observation noise processes with variances σ_l^2 and σ_r^2, respectively. Table 5.1 defines a set of experiments for each gradient value run with various levels of additive noise in the intensity images.

Test Name	Gradient	Constant Displacement
Test1	1.0	2.0
Test2	2.0	2.0
Test3	4.0	2.0

Table 5.1 - Synthetic Image Parameters

The experiments were performed with a constant displacement, D_c, of 2 along the horizontal direction. Two images, 60 by 60 pixels, were generated for the image pairs with a constant displacement. The starting intensity value was 6. The SSD algorithm, described in Section 2.4.2, with a 5x5 patch region is used to estimate the field. In the second part of the experiment, the window size was decreased to a 3x3 to see the effects of restoration on varied window sizes for the SSD estimator. For the results presented in this section, a 36 by 30 pixel region centered in the field is used for F in (97). The restoration improvement is then calculated from (97).

Figure 5.1 shows the dB improvements of the restored displacement field for Test1. Table 5.2 contains the model coefficients for Test1 found by a least squares fit to the displacement field. The image gradient was increased to 2.0 for Test2. Figure 5.2 and Table 5.3 show the dB improvement in restoration and model coefficients, respectively. Figure 5.3 and Table 5.4 contain the results for Test3, which has an image gradient of 4.0. As the gradient of the synthetic images was increased, the ability of the SSD algorithm to estimate a more accurate field increased, and thus the amount of restoration, primarily due to smoothing, had greater effects as seen in the increase in the dB improvement at smaller intensity gradient values, i.e., smaller G values. The model coefficients were found to be consistent over the range of noise levels.

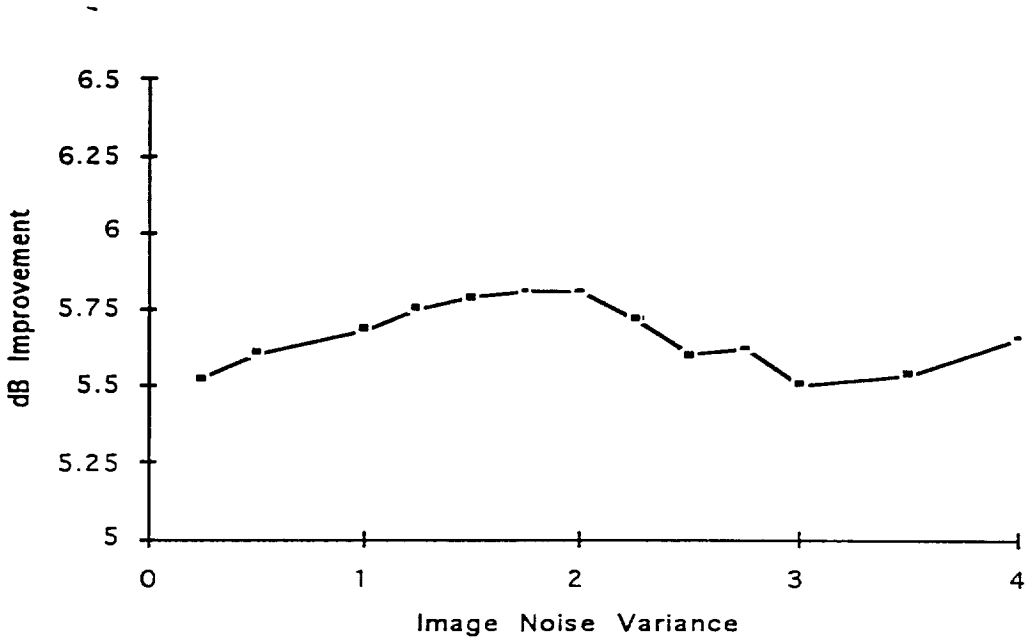

Figure 5.1 - Synthetic Image Test1 Displacement dB Improvement

Image Noise Variance	c_{11}	c_{01}	c_{-11}	c_{10}	σ_w^2
0.25	-0.253838	0.480539	0.089381	0.590033	0.021
0.50	-0.193104	0.426388	0.103746	0.561594	0.043
1.00	-0.198325	0.400132	0.110987	0.575931	0.068
1.25	-0.196602	0.395352	0.109460	0.573996	0.078
1.50	-0.149406	0.355295	0.115935	0.553915	0.096
1.75	-0.158263	0.367996	0.111321	0.556315	0.107
2.00	-0.155577	0.383068	0.109718	0.530657	0.200
2.25	-0.159606	0.368803	0.134305	0.524753	0.133
2.50	-0.161023	0.389643	0.118011	0.508085	0.151
2.75	-0.160399	0.397533	0.120095	0.500720	0.163
3.00	-0.144955	0.357980	0.146730	0.491243	0.185
3.50	-0.112115	0.361107	0.130766	0.463093	0.223
4.00	-0.118534	0.386517	0.113569	0.451274	0.264

Table 5.2 - C Model Coefficients for Test1

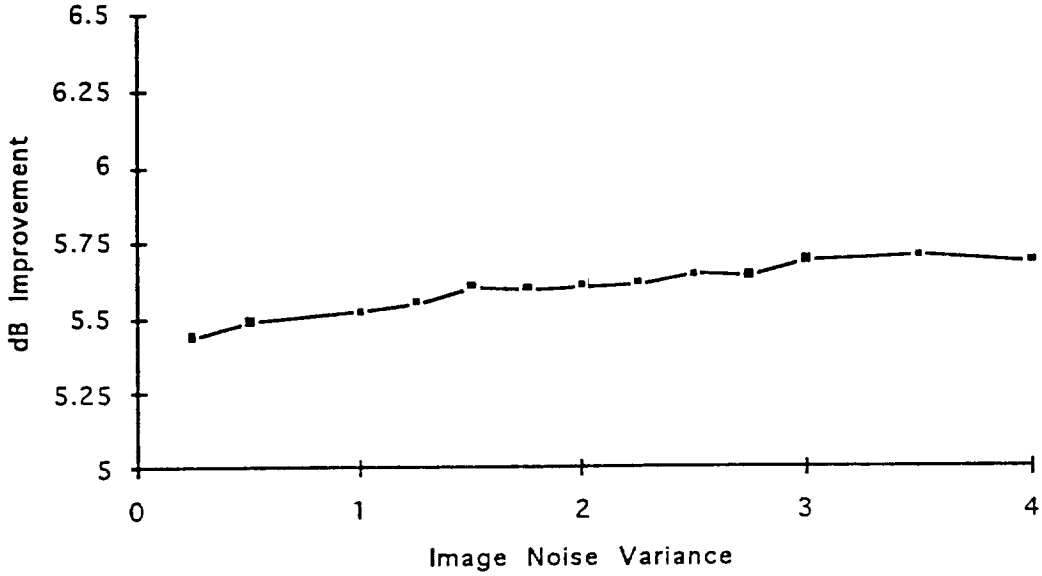

Figure 5.2 - Synthetic Images Test2 Displacement dB Improvement

Image Noise Variance	c_{11}	c_{01}	c_{-11}	c_{10}	σ_w^2
0.25	-0.332085	0.552803	0.066914	0.634953	0.004
0.50	-0.276414	0.501239	0.085115	0.600687	0.010
1.00	-0.253838	0.480539	0.089381	0.590033	0.021
1.25	-0.237760	0.459220	0.097369	0.584822	0.027
1.50	-0.223530	0.458222	0.090899	0.575862	0.032
1.75	-0.202800	0.430005	0.105329	0.565599	0.039
2.00	-0.193104	0.426388	0.103746	0.561594	0.043
2.25	-0.178989	0.407997	0.111205	0.553254	0.048
2.50	-0.183220	0.405837	0.111297	0.559946	0.052
2.75	-0.202473	0.409384	0.112932	0.573637	0.054
3.00	-0.209133	0.422398	0.102219	0.578342	0.057
3.50	-0.212750	0.415691	0.104634	0.585184	0.062
4.00	-0.198325	0.400132	0.110987	0.575931	0.068

Table 5.3 - C Model Coefficients for Test2

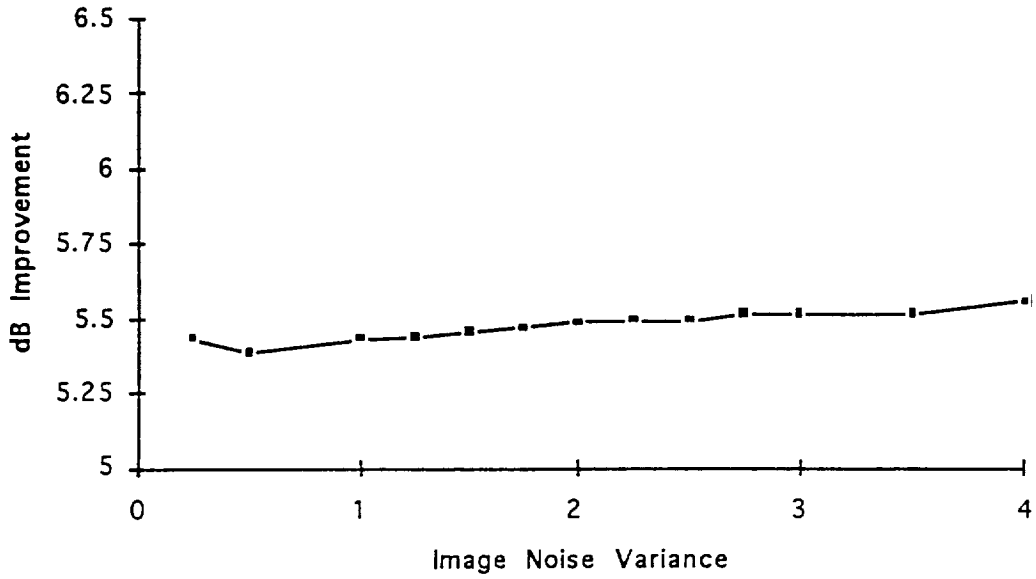

Figure 5.3 - Synthetic Images Test3 Displacement dB Improvement

Image Noise Variance	c_{11}	c_{01}	c_{-11}	c_{10}	σ_w^2
0.25	-0.478257	0.683371	0.025255	0.702028	0.0005
0.50	-0.405868	0.617935	0.051354	0.663716	0.001
1.00	-0.332085	0.552803	0.066914	0.634953	0.004
1.25	-0.317141	0.545553	0.067819	0.622703	0.005
1.50	-0.308074	0.537613	0.070110	0.615781	0.006
1.75	-0.289710	0.516094	0.078659	0.606907	0.008
2.00	-0.276414	0.501239	0.085115	0.600687	0.010
2.25	-0.268939	0.494040	0.087354	0.596997	0.011
2.50	-0.266136	0.485874	0.094072	0.593846	0.013
2.75	-0.262641	0.474641	0.094751	0.599060	0.014
3.00	-0.254343	0.475108	0.092907	0.593036	0.016
3.50	-0.253031	0.478133	0.089159	0.592872	0.019
4.00	-0.253838	0.480539	0.089381	0.590033	0.021

Table 5.4 - C Model Coefficients for Test3

The effects of changing the size of the support region used in the indirect measurement of the displacement field were investigated. In the previous tests in this section, a 5x5 window was used to evaluate the error measure described in (15) for the SSD algorithm. The comparison is made between a 5x5 and a 3x3 window. Figures 5.4, 5.5, and 5.6 show the dB improvements in the restoration of the displacement field when the observation window of the SSD operation is reduced to a 3x3 patch region. The estimates in the unrestored field for the 5x5 window used above are less noisy than the unrestored field for the 3x3 window due to the smoothing effects of the larger observation patch region utilized in the SSD error measure function, (15). The effect of this smoothing is carried over into the restoration process and a greater dB improvement in restoration is observed for the 3x3 patch region.

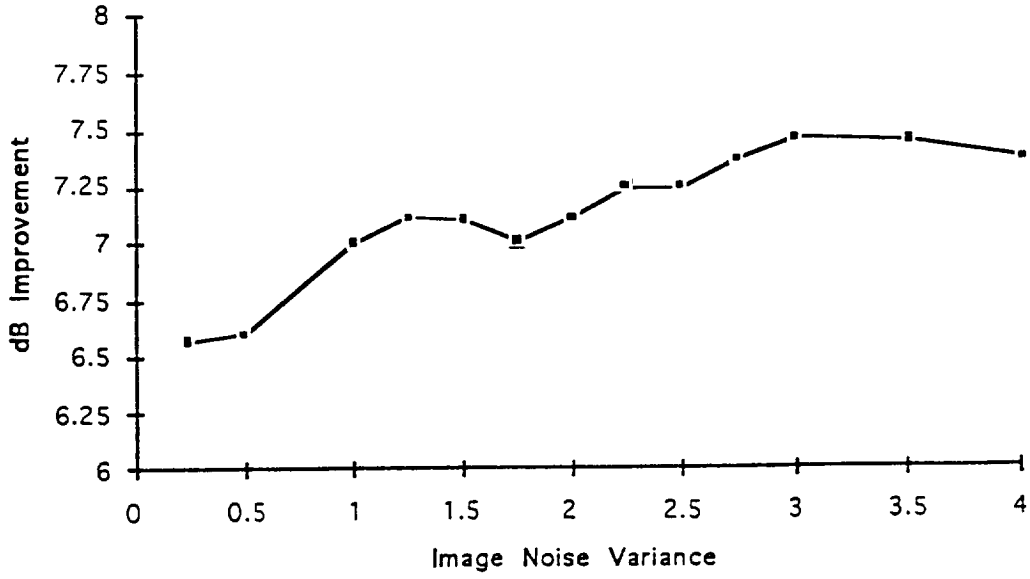

Figure 5.4 - Synthetic Image Test1 Displacement dB Improvement

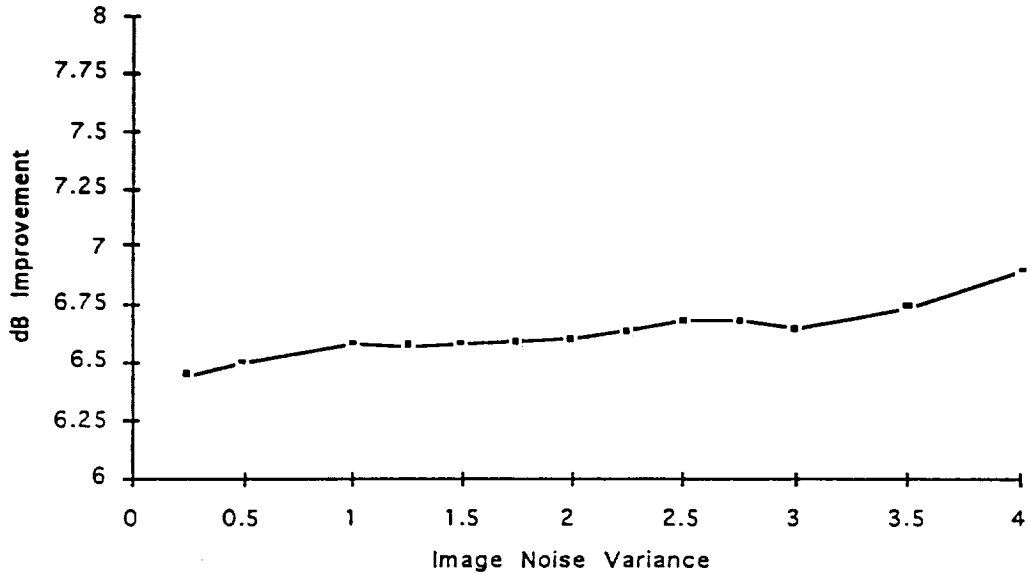

Figure 5.5 - Synthetic Image Test2 Displacement dB Improvement

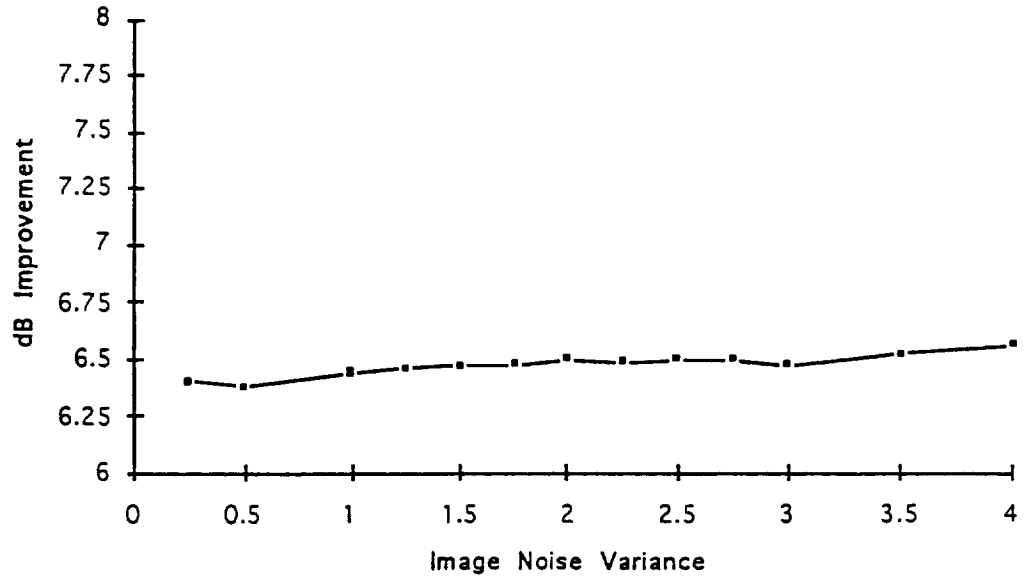

Figure 5.6 - Synthetic Image Test3 Displacement dB Improvement

The figures show that the restoration process is effective and consistent in its dB improvement over a wide range in input intensity noise levels and small gradients for the image pairs. The distorted displacement fields derived from the images created with the smaller gradient, Test1, are noisier than the images with the larger gradient, Test2 and Test3. The dB improvement for Test1 is greater than in the less noisy case, Test3. The estimates found by the SSD procedure with a 3x3 patch region are more sensitive to the additive image intensity noise than the larger 5x5 patch region's fields. The restoration of the fields found by the SSD with a 3x3 patch region has a greater dB improvement due to the smoothing property of the restoration procedure. The restoration of the noisy displacement field found by the SSD with 1x1 patch region produced even greater dB improvements, but the estimated model coefficients were very sensitive to the intensity noise level. Generally, the amount of improvement due to the restoration will decrease as the patch region increases due to the smoothing effects of the larger regions involved in the error measure of the SSD algorithm.

5.2. Real Images and Known Displacement

This experiment involved the capture of an image of an actual scene, a single block taken from Figure 5.13, and the generation of a second image of the scene obtained after shifting the image horizontally by two pixels. A series of various noise levels of uncorrelated white Gaussian noise was added to each image. The SSD algorithm, described in Section 2.4.2, with a 3x3 patch region is used to estimate the field. A ROMKF with a single spatial model is used in the restoration of the distorted displacement field.

The model coefficients for this experiment were estimated by a least squares fit of the $1\times1\times1$ NSHP support region to the estimated field. Figure 5.7 shows the dB improvements of the restored displacement field for the real images corrupted by additive noise. The results of the experiment show an average of a 7.1dB improvement in the displacement field estimates due to the restoration of the distorted displacement field.

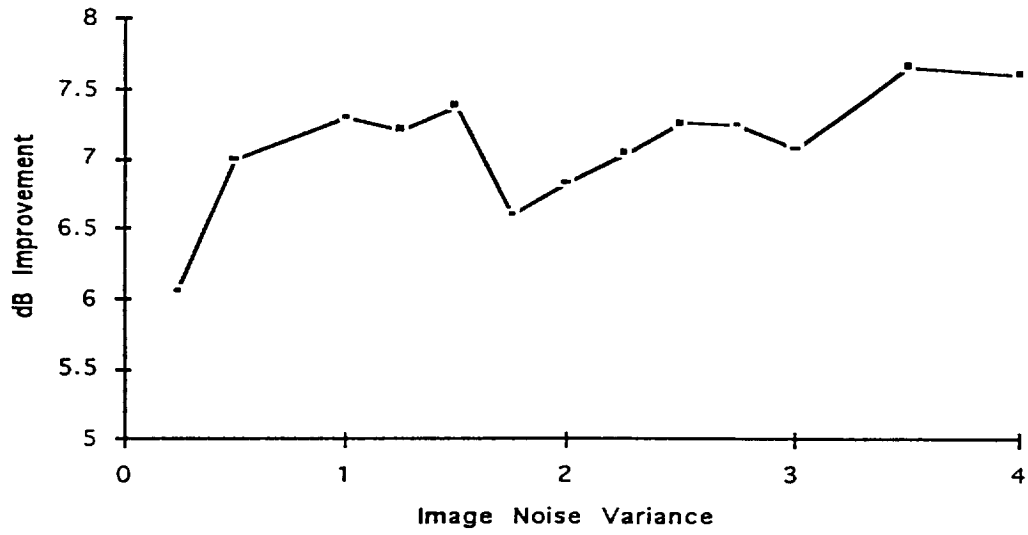

Figure 5.7 - Real Image with Known Displacement

5.3. Effects on Prefiltering of Images

The issue of the effectiveness of prefiltering the intensity images to reduce the noise contribution prior to the estimation of depth and displacement is now addressed. Since the depth and displacement fields are estimated by the apparent motion in the images generated by changes in the intensity values, the presence of significant noise in the image will distort these intensity values. The noise can present a significant problem in those image

regions that have sub-pixel displacement vectors. It is difficult to estimate the displacement vector if the image intensity noise has the same magnitude as the changes due to motion. In the extreme case of regions that have no motion, such as in a still background scene, noise in the images can produce random displacement values. In algorithms that require gradient operations on the input intensity images, additive image noise will degrade the accuracy of these estimates. Filtering the input intensity images to reduce the noise can be used to provide smoother estimates for the displacement field.

To construct this experiment, a procedure similar to that described in Section 5.1 is used. Two synthetic images are generated with a known displacement of 2 pixels in the horizontal direction, various noise levels are applied to the intensity images, and displacement field estimates are found for the prefiltered and unfiltered image cases. The displacement field is found by the use of the SSD estimator with a 5x5 patch region. A ROMKF with a single spatial $1\times1\times1$ NSHP support region is used for the prefiltering stage. The displacement field is filtered using a single model ROMKF. Figure 5.8 shows the sum squared error between the estimated and the true displacement fields for a 36 by 30 pixel region centered in the field obtained for Test1.

The markers in Figure 5.8 have two terms. These indicate the filtering condition of both the input intensity image and the estimated displacement field. For example: Unfiltered-Filtered indicates that no prefiltering of the input intensity images was done but that the estimated displacement field was filtered (restored).

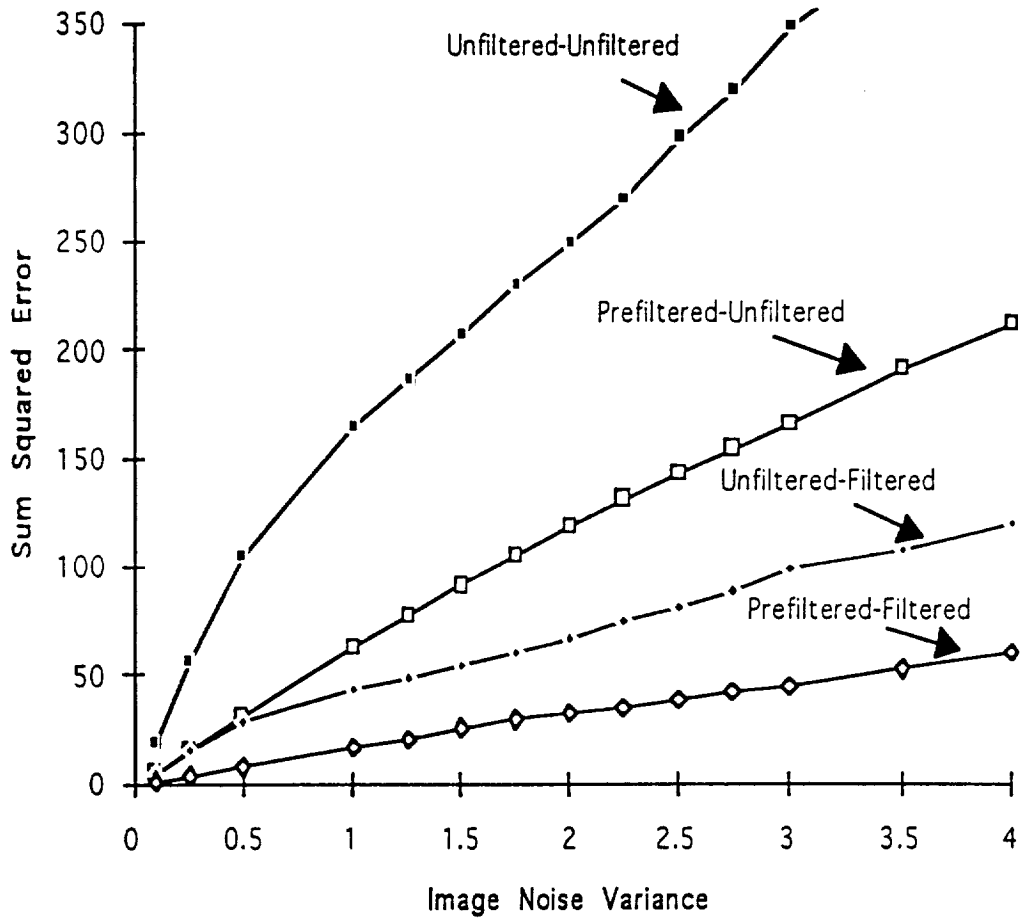

Figure 5.8 - Effects of Prefiltering Images

Several observations can be made on the results of this experiment. First, filtering of the displacement field estimates produces significant improvement in the results regardless of whether the input intensity images are prefiltered or unfiltered. This shows that for realistic noise levels the restoration of the displacement field estimates provides superior results to those obtained by just filtering the input intensity images. Secondly, prefiltering the input intensity images produces a more accurate field and

has its greatest effect when higher noise levels in the images are involved. As the image noise level increases, prefiltering the input intensity images becomes of greater importance; however, at all noise levels filtering of the displacement field estimates is of greatest importance.

Figure 5.9 shows the dB improvement, defined by (97), of the restored displacement fields with prefiltered and unfiltered intensity images. A region of 36 by 30 pixels centered in the field is used to calculate the restoration dB improvement.

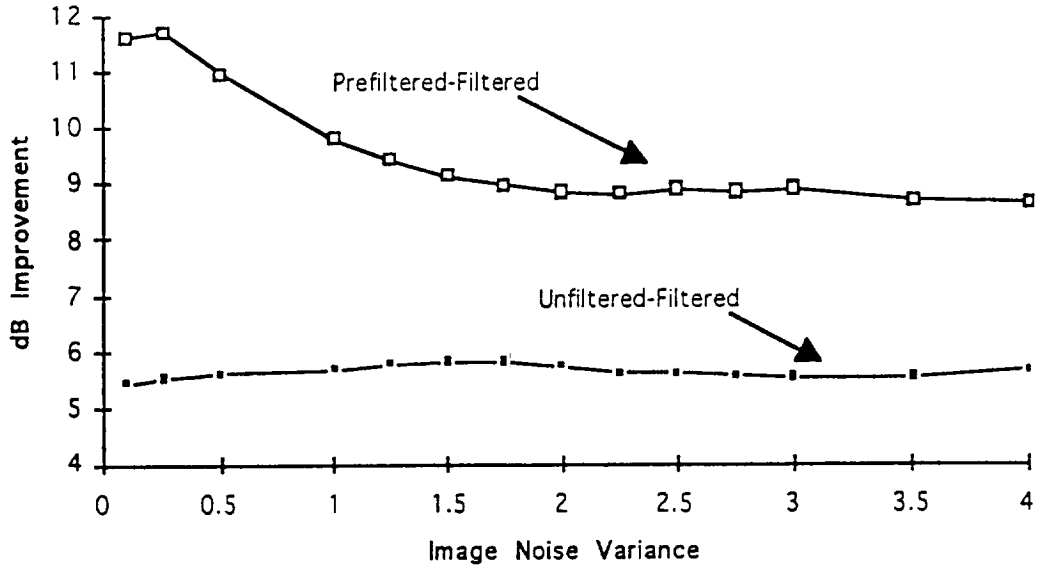

Figure 5.9 - dB Improvement in Restoration using Prefiltering

For these synthetic scenes, which contain no discontinuities in either the image or displacement fields, prefiltering the input intensity images produces better results. When dealing with complex scenes, care must be taken not to blur the edges when prefiltering the intensity images as this can remove sharp boundaries present in the actual depth field.

5.4. Direct Observation Equation Experiments

In this experiment, an observation model that is based on a direct measurement of the corrupted field will be used. Recalling (46) for the stereo camera setup, the observation of the displacement, $d(x,y,t)$, at location (x,y) at time t is:

$$I_r(x,y,t) = I_l(x - d(x,y,t),y,t) + n(x,y,t) \tag{100}$$

where $I_r(\cdot)$ and $I_l(\cdot)$ are the intensities for the right and left images, respectively, and $n(\cdot)$ accounts for the observation modeling error. For this test the observation is made at a specific location, a single pixel, in each image according to (100). A ROMKF with a single spatial model is used in the restoration of the distorted fields. The model coefficients for these experiments are taken from Table 5.4. The left and right images are generated by (98) and (99) given in Section 5.1 with $G = 4.0$ and $D_c = 1.0$. A 3x3 Sobel operator [23] is used to estimate the gradients in the left image. Since the stereo images have a constant displacement in the horizontal direction, the variances of the estimated field can be used as evaluation criteria for the direct observation. Figure 5.10 shows the effect of varied window sizes on the sum squared error between the estimated and the true displacement field for a 55 by 54 pixel region centered in the field. The variance of the estimated displacement field increases consistently with the increase in intensity of the noise since the update equation is based on single noisy intensity values.

The indirect methods previously described, in Section 3.4.2, made use of a larger patch region for each displacement estimate. The larger patch region lessens the effect of the additive noise by an averaging procedure.

The patch region must be large enough to reduce the sensitivity to noise while small enough to maintain the local characteristics of the intensity image.

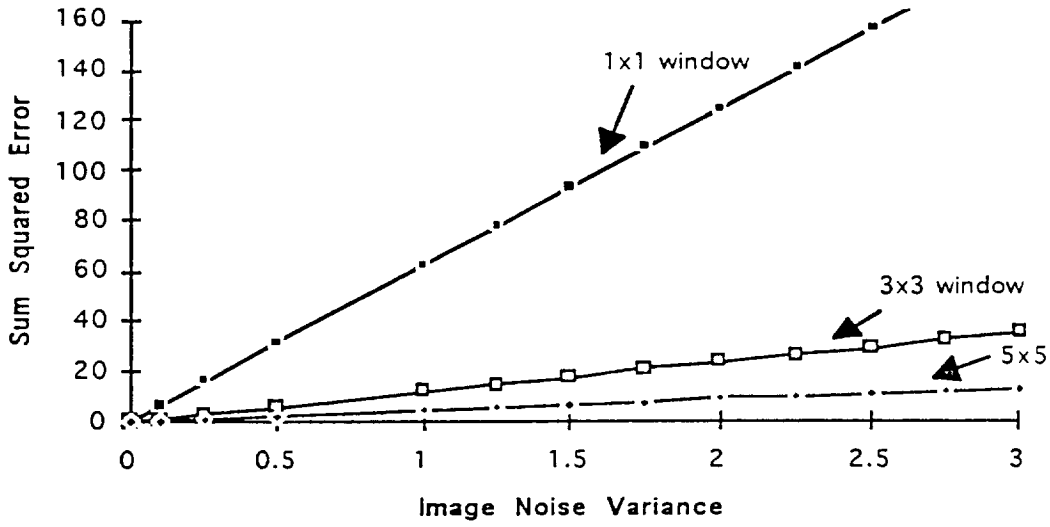

Figure 5.10 - Sum Squared Error of Direct Observation

In the next direct measurement experiments, the ROMKF with a $1\times1\times1$ NSHP single model support region and an observation equation based directly on the input intensity images is used to estimate the depth field. Recalling (50) in Section 3.4.2,

$$\Delta I(x,y) = \frac{1}{z(x,y)} \frac{\partial I_l(x,y)}{\partial x} bf + v_i(x,y) \tag{101}$$

where $I_l(x,y)$ denotes the intensity value at pixel location (x,y) for the left image, $z(x,y)$ is the depth at location (x,y), b is the baseline width or translation in the x direction, f is the camera focal length, and $v_i(x,y)$ accounts for uncertainty in the gradient and modeling approximations. An observation is made on "inverse depth", i.e., $1/z$. This permits a linear

observation of the depth at each pixel location. To the system model equations developed in Chapter 3 is applied a state vector composed of inverse depth terms. A 3x3 Sobel operator [23] is used to estimate the gradients needed in the left image. The change in intensity, ΔI, is estimated over a square patch region for a single point difference provided very noisy estimates on the intensity difference. To obtain a smoother estimate the average intensity difference over a 3x3 or a 5x5 pixel area is utilized. The results using each patch region size, shown in Figure 5.11, are then compared.

In the direct observation of the depth field experiment, the left and right images are generated by (98) and (99) given in Section 5.1 with $G = 4.0$ and $D_c = 2.0$. Equation (101), with a focal length of 14mm and a 1mm baseline length, is applied on the noiseless images to estimate the true inverse depth field. The distorted inverse depth field is estimated by evaluating (101) over the noisy stereo image pairs. A ROMKF with a single model is used to estimate a restored inverse depth field. The model coefficients were found by a least squares identification on the corrupted field. To evaluate the improvement due to the restoration process, a dB improvement similar to (97) is used on the estimated inverse depth fields. Figure 5.11 shows the dB improvement of the restoration process for the 3x3 and the 5x5 observation window sizes used in the evaluation of the intensity difference for (101). The dB improvement is calculated over a 26x26 patch region centered in the field.

The distorted inverse depth field estimated with a 5x5 observation window is less noisy than with the smaller 3x3 window estimator, and thus

there is a smaller amount of dB improvement due to restoration. The estimation procedure breaks down at larger levels of noise added into the input stereo images. This is mainly due to the gradient operation involved in (101), which is sensitive to the noise level. At the larger noise values, the signal gradient is poorly estimated and thus the inverse depth is unreliable.

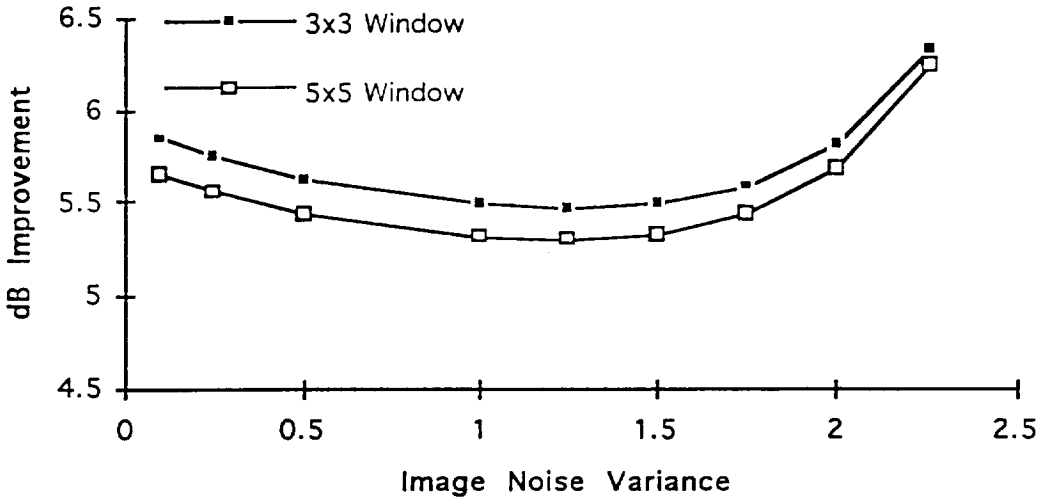

Figure 5.11 - dB Improvement of Direct Observation of Depth Field

In the previously described direct depth experiments, the images were generated with a constant depth field. A spatially variant depth field can be created by modifying equations (98) and (99) in the following way:

$$I_l(x,y) = Gx + B + n_l(x,y) \tag{102}$$

$$I_r(x,y) = G\big(x + D(x,y)\big) + B + n_r(x,y) \tag{103}$$

where $I_l(x,y)$ and $I_r(x,y)$ are the intensity values at pixel location (x,y) for the left and right images, respectively, G is the gradient value for the test image, B is the starting intensity, $D(x,y)$ is the spatially variant displacement, and $n_l(x,y)$ and $n_r(x,y)$ are zero-mean uncorrelated

Gaussian image observation noise processes with variances σ_l^2 and σ_r^2, respectively. The spatially variant displacement would be observed for a sloped surface imaged in the left and right stereo frames. For this experiment, the function for the spatially variant displacement is

$$D(x,y) = \frac{x}{120} + \frac{y}{240} \tag{104}$$

As before, equation (101), is applied on the noiseless images to estimate the true inverse depth field. The distorted inverse depth field is estimated by evaluating (101) over the noisy stereo image pairs. A 5x5 observation window is utilized for the estimated intensity difference at each location. A ROMKF with a single model is used to estimate a restored inverse depth field. The dB improvement is calculated over a 26x26 patch region centered in the field. Figure 5.12 shows the dB improvement of the restoration process and demonstrates that the filtering algorithm can be successfully applied to the restoration of depth fields found by direct observation of sloped surfaces.

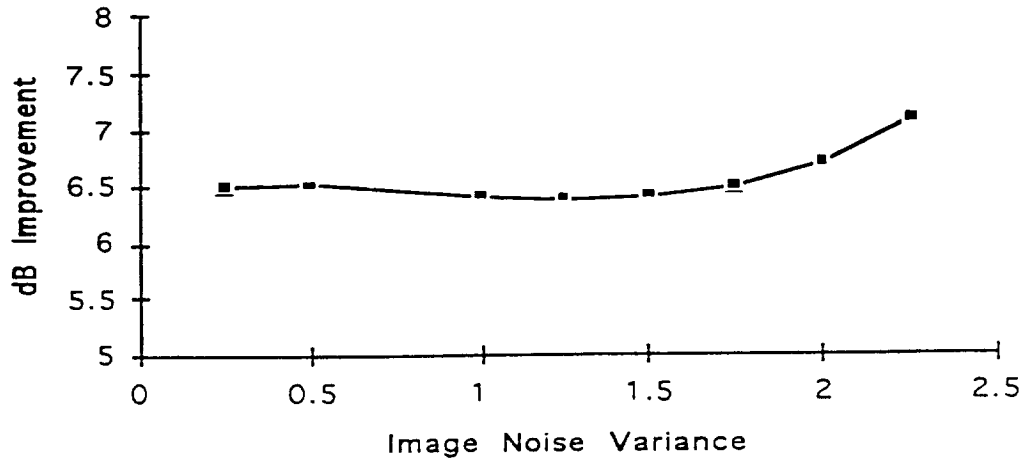

Figure 5.12 - dB Improvement of Spatially Varying Depth Field

5.5. Spatial-Temporal Experiments

For this experiment, the images created for Test3, described in Table 5.1, are used for the image pairs. The model coefficients are found by a least squares fit to the field with fixed temporal model coefficients in (67). The additive image noise has a variance of 1.0. The SSD algorithm with a 5x5 patch region is used to estimate the displacement field. The results in this section are reported in terms of the sum squared error between the estimated and the true displacement field. The results for the unfiltered, spatially restored, and spatio-temporally restored fields, for a 36 by 30 pixel region centered in the field, are listed in Table 5.5.

Temporal Coefficient	Unfiltered	Spatial Model	Spatio-Temporal Model
0.2	9.835247	2.612037	2.573474
0.3	9.835247	2.612037	2.511555
0.4	9.835247	2.612037	2.588500

Table 5.5 - Sum Squared Error for Spatial vs. Spatio-Temporal Restoration

The field is very smooth in the spatial domain, and there is no temporal change, therefore the improvement is small. Although the addition of the temporal component moderately increases the overall dB improvement for the field, the effect is not dramatic since the spatial field has rather good estimates. The contribution of the temporal component will be more significant for sequences that contain spatial discontinuities.

5.6. Adaptive Filtering - Multiple Model Results

In previous experiments, the model parameters were kept constant throughout the filtering of the displacement or depth fields. These parameters were estimated by a least squares fit to the entire field. While

this technique works well in homogeneous regions where the model represents the underlying field, the homogeneous parameters can have unacceptable effects when discontinuities appear in the field. Such discontinuities occur in regions in which the depth or displacement field is obtained on different objects, a single object moving across a stationary background, or several objects moving at different velocities or at different depths.

In images representative of robotic scenarios, depth and displacement boundary discontinuities are coincident with image intensity boundaries. The converse of this statement is not true since there can be intensity edges on a surface that undergo similar displacement between image pairs. To permit the model parameters to adapt to the underlying field, the edge information in the intensity image will be referenced to select the most appropriate model from a bank of model parameters, i.e., the multiple model process, as described in (95) and (96) of Section 4.4. This selection allows for adaptive processing by modifying the parameters to follow the estimated underlying depth or displacement field.

To demonstrate the effectiveness of the multiple model ROMKF over the single model ROMKF for fields which contain discontinuities, a pair of stereo images was taken of a scene constructed with a set of blocks. A Panasonic M2.6 wv-BD400 camera attached to a translation table is used to capture the images. Table 5.6 lists the parameters for the camera used for this experiment.

Property	Value
Focal Length	14.37mm
Horizontal Scale	0.985313
Horizontal CCD Spacing	13 microns
Image Center	(230, 244)
Baseline Width Increment	0.5mm

Table 5.6 - Stereo Camera Parameters

The single camera acquires a sequence of images by undergoing small lateral translations. A depth field can then be acquired with a small baseline width between frames. These images were selected since there are clearly defined boundaries between the various objects (the blocks) and the background. An arbitrary selection of using Frame 0 from the image sequence for the left frame and Frame 30 from the sequence for the right frame was made. Figure 5.13 shows the left and the right images for the stereo pair.

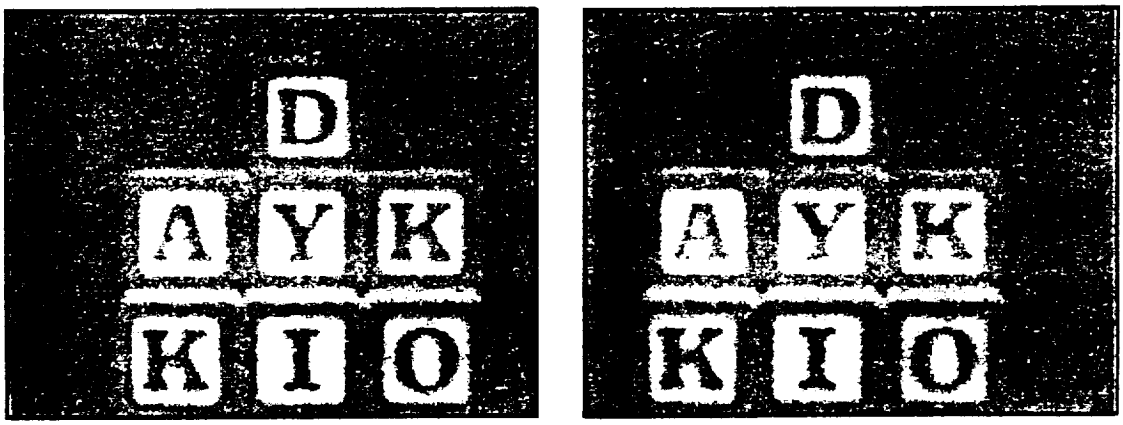

Figure 5.13 - Left and Right Image for Multiple Block Scene

In the estimation of the displacement fields for the image pair, a simple threshold test was used to segment objects from the background. A threshold intensity of 90 produced acceptable quality for the segmentation. A compass operator, shown in Figure 4.5, was applied to the left image to

produce the edge directions and strengths needed for the multiple model ROMKF. Figures 5.14 and 5.15 show the unfiltered depth field for the back and side view. These stereo images were selected since they contained clearly defined boundaries between the objects and the background. A displacement of about 20 pixels was observed between the left and right images.

As a means to show the differences between restoration involving a single model and restoration using multiple models, the displacement fields from both the single and multiple model ROMKF will be compared. The first set of results involves the use of the ROMKF with a single spatial model for the restoration of the distorted fields. The displacement field was then filtered with a single model employing the parameters, found by a least squares identification over the entire field, listed in Table 5.7.

c_{11}	c_{01}	c_{-11}	c_{10}
-0.528959	0.728204	0.018224	0.722108

Table 5.7 - Model Coefficients for Single Model for Block Image

The results from the restoration procedure are shown for the back and side views in Figure 5.16 and 5.17. There is a considerable amount of smoothing of the edge boundaries, in the order of 10 pixels, due to the mismatch between the model and the underlying field. This could provide difficulty in obtaining accurate object edge boundaries for calculations which work on the depth field.

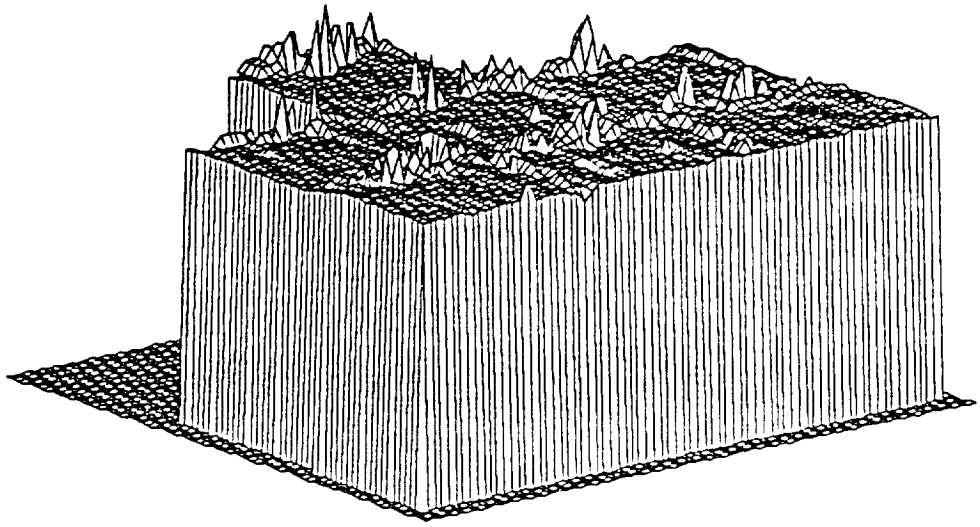

Figure 5.14 - Unfiltered Displacement Field of Block Set (Back View)

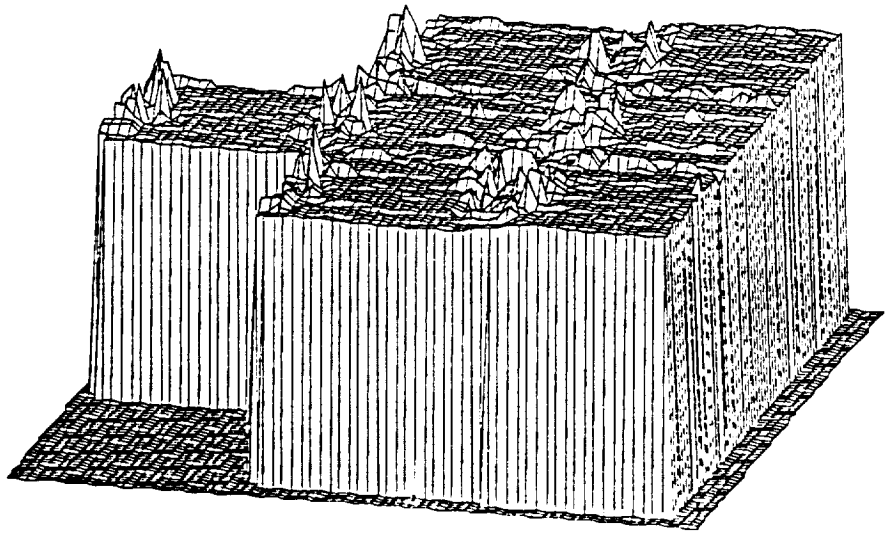

Figure 5.15 - Unfiltered Displacement Field of Block Set (Side View)

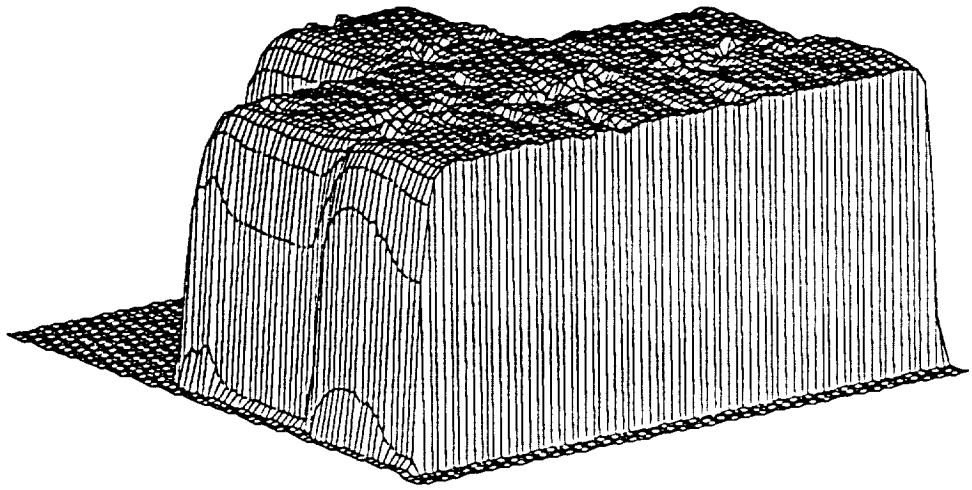

Figure 5.16 - Single Model Filtered Displacement Field of Block Set (Back View)

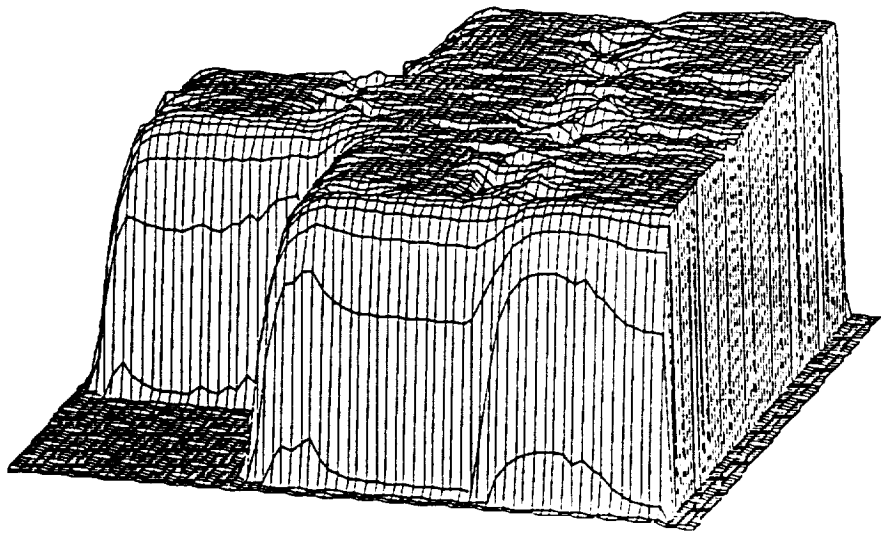

Figure 5.17 - Single Model Filtered Displacement Field of Block Set (Side View)

The theory behind using the intensity edge directions to select model parameters is based on the fact that the pixel locations in the intensity image and those in the displacement field which are from the same object should be modeled as such. Clearly there is a benefit to selectively weighting the coefficients in the state model to include only those state elements that correspond to the same object. The homogeneous model tends to smooth out the edges in the depth and displacement field which is undesirable if the field will be used in subsequent processing, such as boundary detection for robotic object manipulation.

The next restored fields are obtained by the application of the ROMKF with a multiple model approach to better follow the underlying field by using information extracted from the intensity images. For the ROMKF restoration, five models were designed for the bank of multiple models. The model coefficients are shown in Figure 5.18. These models were used with the multiple model procedure described in Section 4.4.

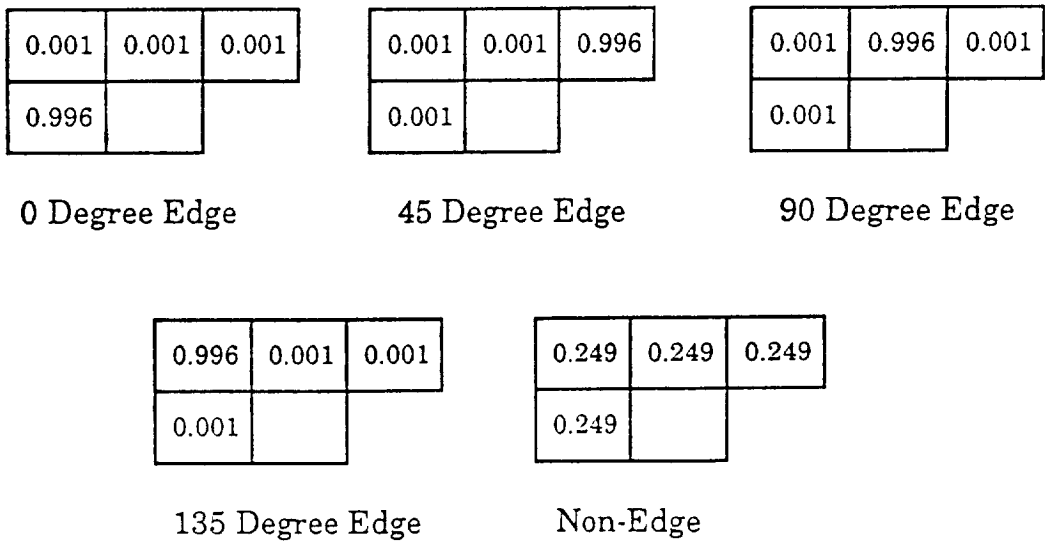

Figure 5.18 - Coefficients for Multiple Model Bank

In the compass edge detection procedure on the input intensity images, edges are found along the edges of the blocks as well as along the edges of the painted surfaces. The edges that correspond to the painted surfaces of the blocks do not represent discontinuities in the depth or displacement field. Since there is no discontinuity in the displacement field as detected by (96), the non-edge model is used over this area. This test set has a large difference between the various values of the displacement field. A threshold of 40 percent of the value was used to detect possible discontinuities in the displacement field.

Figures 5.19 and 5.20 unequivocally show the edge preservation provided by the use of multiple models to adjust the restoration procedure to changes in the underlying displacement field. A couple of artifacts appear due to misidentification of edge directions. The models, tuned to various directions for possible discontinuities, maintain the sharp features of the field. At locations where there are no discontinuities in the field, the non-edge model is applied to reduce the noise content in the field.

To show the effect of the multiple model more clearly, the profiles from a row are shown in Figure 5.21. The figure shows the edge of the blocks along line 35 of the image. The dashed line represents the results from the multiple model restoration, dotted line is the single model restoration, and the solid line is the unfiltered displacement field. The non-edge model for filtering is selected at all columns in Figure 5.20 except for column 96 where the 90° edge model was selected. The smoothing of the edges in the displacement field by the use of a spatially invariant model is clearly seen.

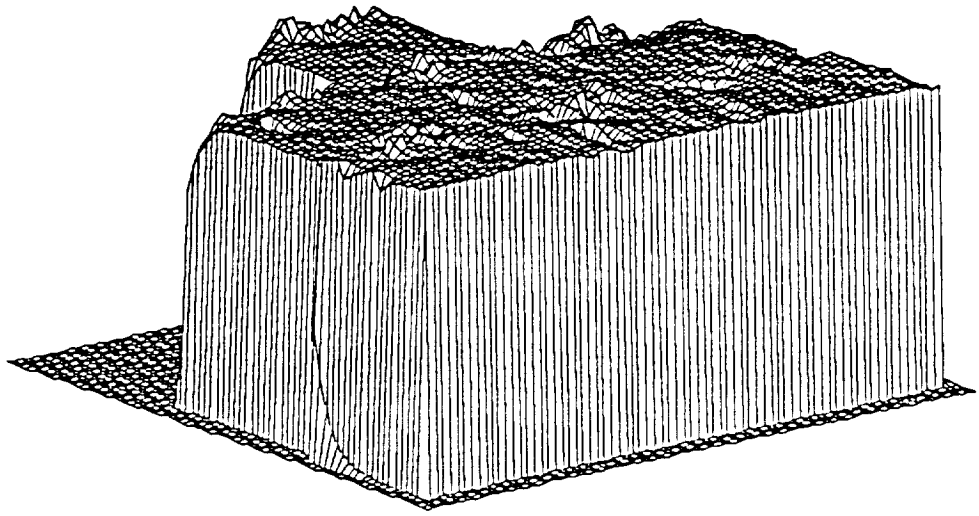

Figure 5.19 - Multiple Model Filtered Displacement Field of Block Set (Back View)

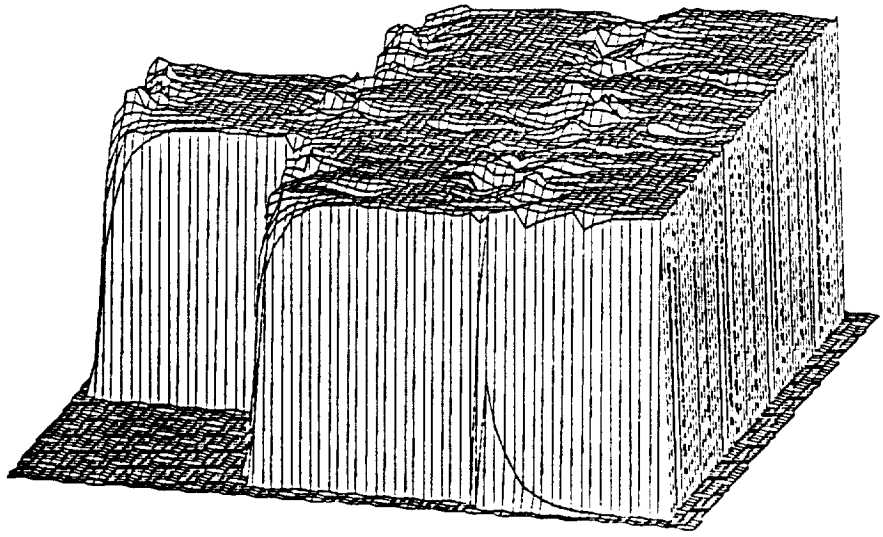

Figure 5.20 - Multiple Model Filtered Displacement Field of Block Set (Side View)

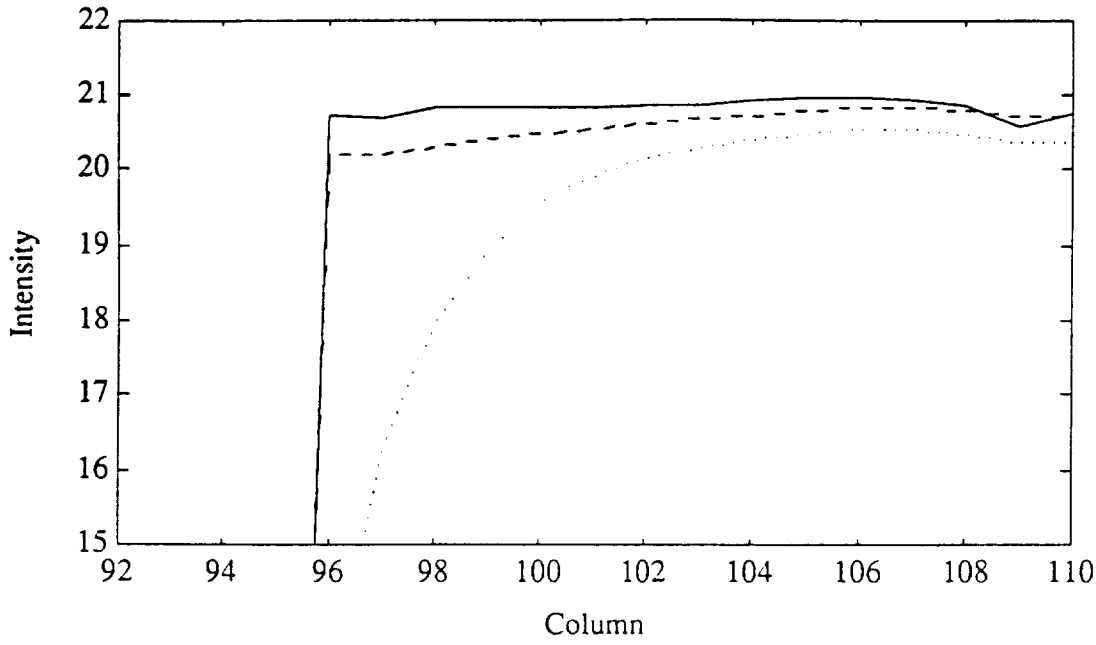

Figure 5.21 - Intensity Profile for Displacement Field Restoration

Figure 5.22 displays the depth field from the results of the multiple model ROMKF restoration of the displacement field. The image is extruded to simulate the depth at each location.

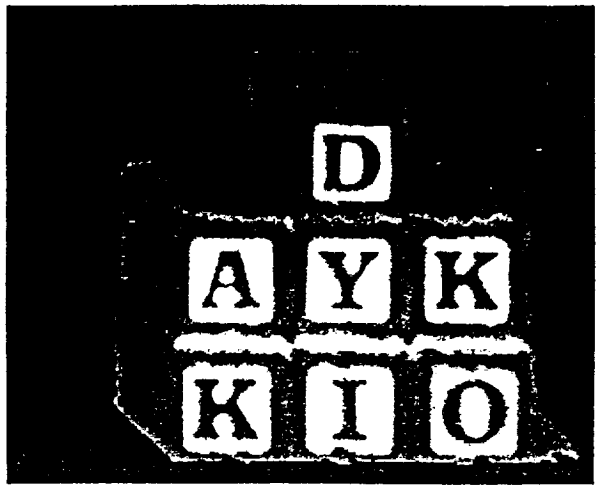

Figure 5.22 - Extruded Intensity Image to Indicate Depth Field

5.7. Depth Density Distributions

The stereo camera setup and equations to relate displacement between the left and right frames to the location of the point in 3-D space is detailed in Section 2.2. Recalling (4), the depth, Z, is found to be inversely proportional to the displacement, $(x_l - x_r)$, and can be written as:

$$Z = b\frac{f}{x_l - x_r}; \qquad f > 0, b > 0 \tag{105}$$

where b is the baseline between the cameras, and f is the focal length. In Section 4.3.3, the relation between depth and displacement was detailed in terms of density distributions. The density distribution of the depth field was written in terms of the density distribution of the displacement. The answer to the computational procedure question posed in Section 4.3.3 (i.e., to filter the displacement and then calculate the depth field, or to calculate then filter) can now be achieved through the application of the following representative cases.

For the cases in this experiment, the camera parameters shown in Table 5.6 are substituted into (105) resulting in:

$$Z = \frac{1.218614}{x_{Pl} - x_{Pr}} \tag{106}$$

where $x_{Pl} - x_{Pr}$ is the displacement in pixel units between the left and right images. A set of images was synthesized of objects located 0.5 to 2 meters away from the stereo camera setup with a baseline width of 1mm. The left and right images from the stereo pair are corrupted with additive white Gaussian noise with a variance of 1.0. The surface of the object had a local gradient of 2.0 pixels in the horizontal direction. The displacement field between the left and right image is estimated by the SSD algorithm with a

5x5 observation patch region. The resulting variance of a 36 by 30 pixel region in the estimated distorted displacement field was 0.029 (the Cramér-Rao lower bound for this case is 0.020). Figures 5.23, 5.25, and 5.27 show the distribution for the estimated displacements for objects located at 0.61, 1.22, and 2.44 meters, respectively. The displacement values were generated by 5000 samples from a Gaussian number generator with the appropriate mean. The distributions for the depth calculations, found by (105), are shown in Figures 5.24, 5.26, and 5.28 for the three distances. Figure 5.28 has an upper threshold of 10 meters for plotting purposes.

The following figures clearly show the nonlinear properties of (105). Although the input displacement estimates have a Gaussian distribution, the calculated depth values are clearly non-Gaussian. The actual distribution is described by (94).

$$f_z(z) = \frac{fb}{z^2} f_s\left(\frac{fb}{z}\right) \tag{107}$$

where $f_s(\cdot)$ and $f_z(\cdot)$ are the distributions for the displacement and depth, respectively.

As the displacements approach zero, the calculated depth approaches infinity. Thus the depth calculation is very sensitive when there are errors present in small displacement values. Due to these nonlinear effects, filtering the displacement before calculating depth is preferred.

The effects of filtering the displacement before calculating the depth field can be shown by having the displacement field filtered with the models described in Chapter 3 (ROMKF with a $1 \times 1 \times 1$ NSHP model support). The estimates of the model coefficient parameters are listed in Table 5.8.

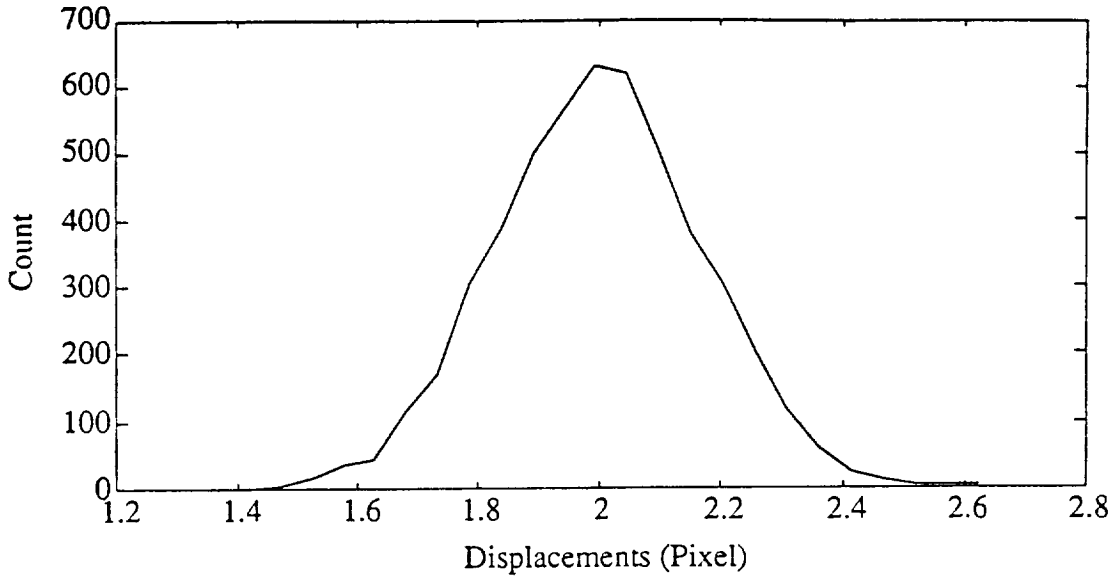

Figure 5.23 - Distribution of Estimated Displacement for 0.61 meters

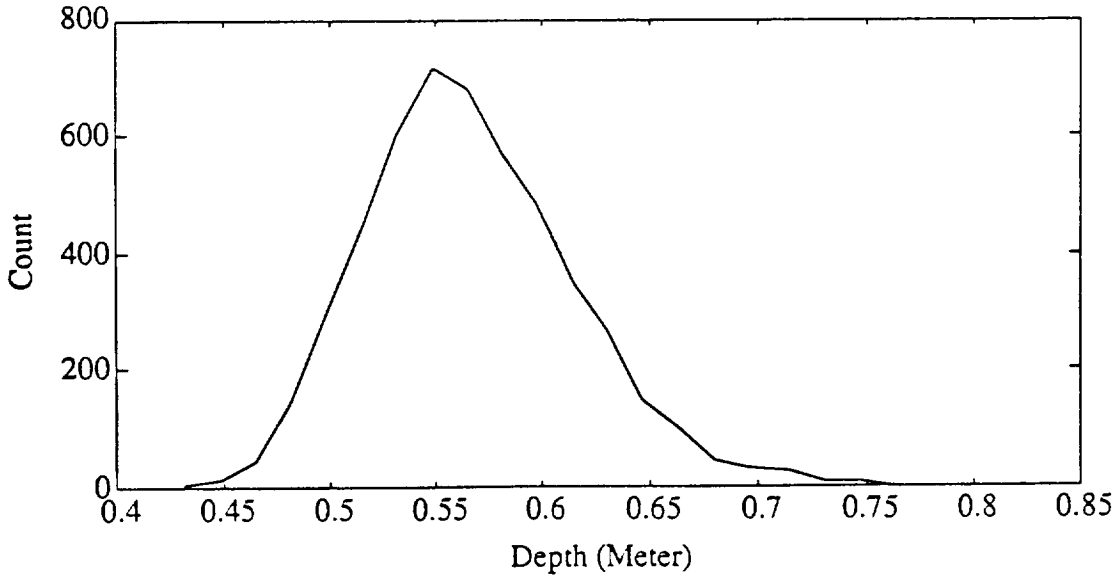

Figure 5.24 - Distribution of Estimated Depth for 0.61 meters

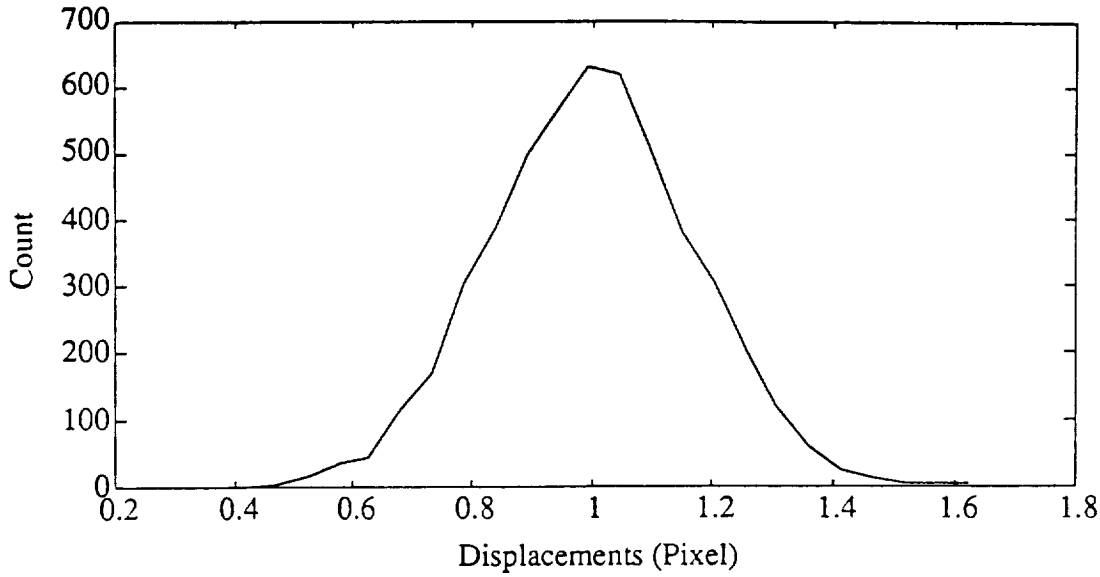

Figure 5.25 - Distribution of Estimated Displacement for 1.22 meters

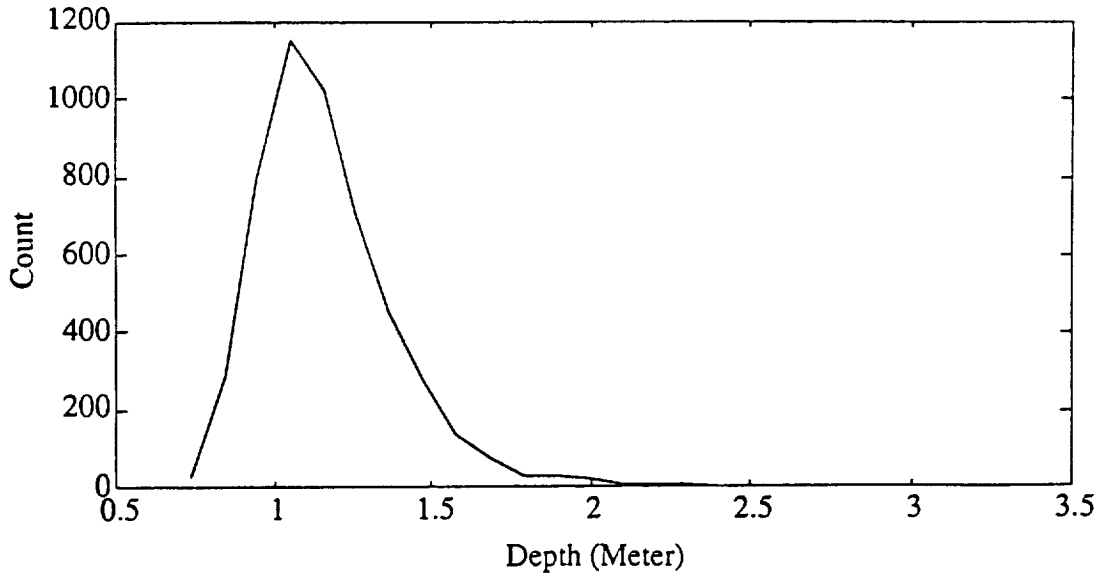

Figure 5.26 - Distribution of Estimated Depth for 1.22 meters

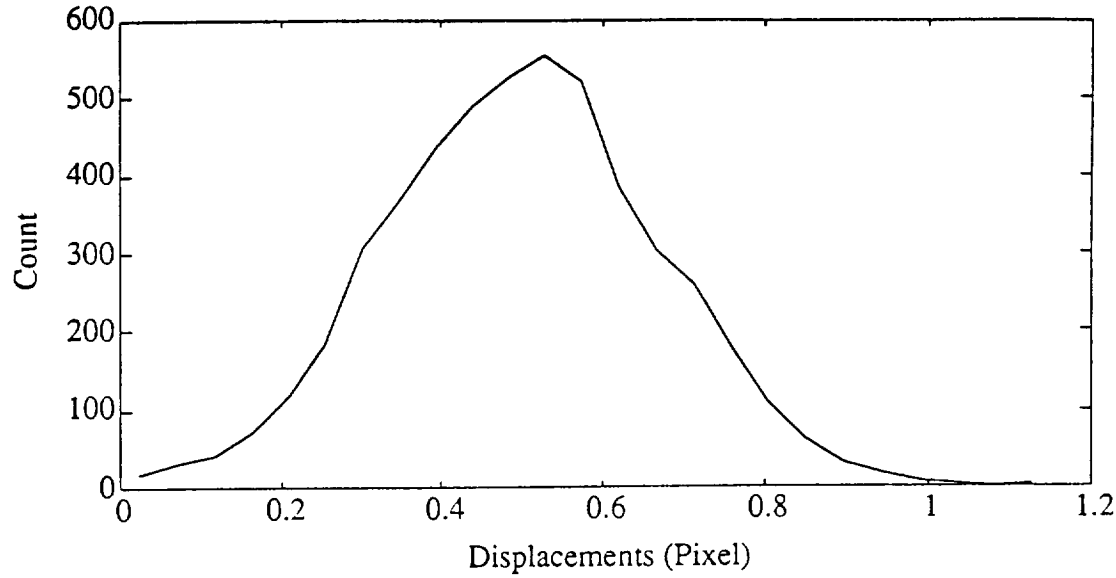

Figure 5.27 - Distribution of Estimated Displacement for 2.44 meters

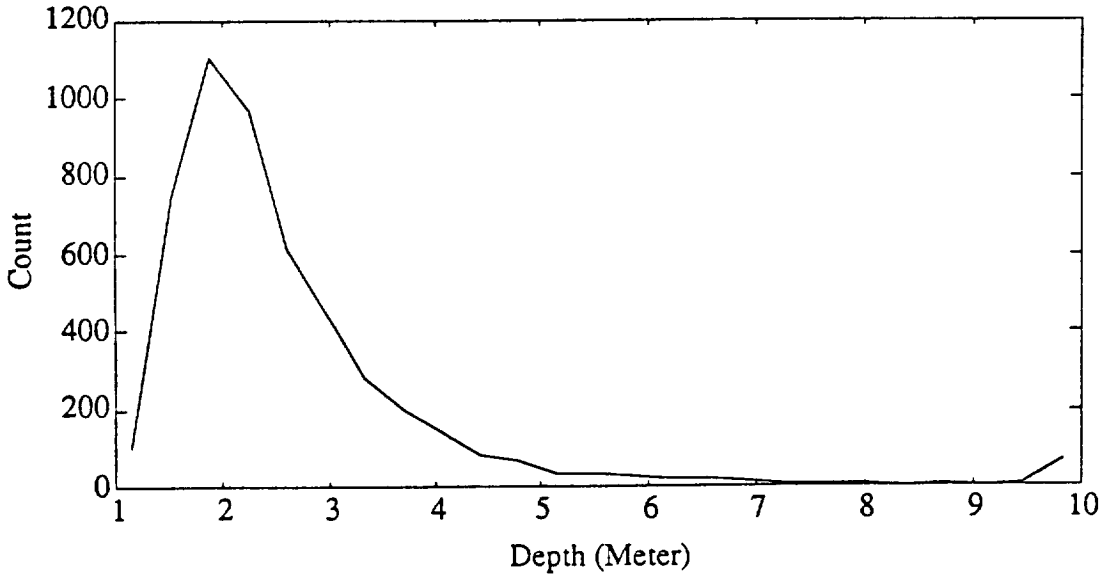

Figure 5.28 - Distribution of Estimated Depth for 2.44 meters

c_{11}	c_{10}	c_{-11}	c_{01}
-0.528959	0.728204	0.018224	0.722108

Table 5.8 - Model Coefficients for Distribution Experiment

After the restoration procedure on the displacement field, the variance of the filtered displacement estimates was 0.009. As before, 5000 samples were used to produce the distribution plots shown below. The results of the depth calculations from the filtered displacements and (106) are shown in Figures 5.29, 5.30, and 5.31.

Here the benefits of filtering the displacement are evident. The overall spread in the depth distributions is much smaller for the restored displacements than in the case of the unfiltered displacements. Two observations can be made on the distribution of depth calculated from displacement estimates, (84): first, the distribution is non-Gaussian, and second, the calculations are ill-behaved and unbounded for small displacement values. Application of these observations shows the benefit of filtering the displacement fields before performing the depth calculation.

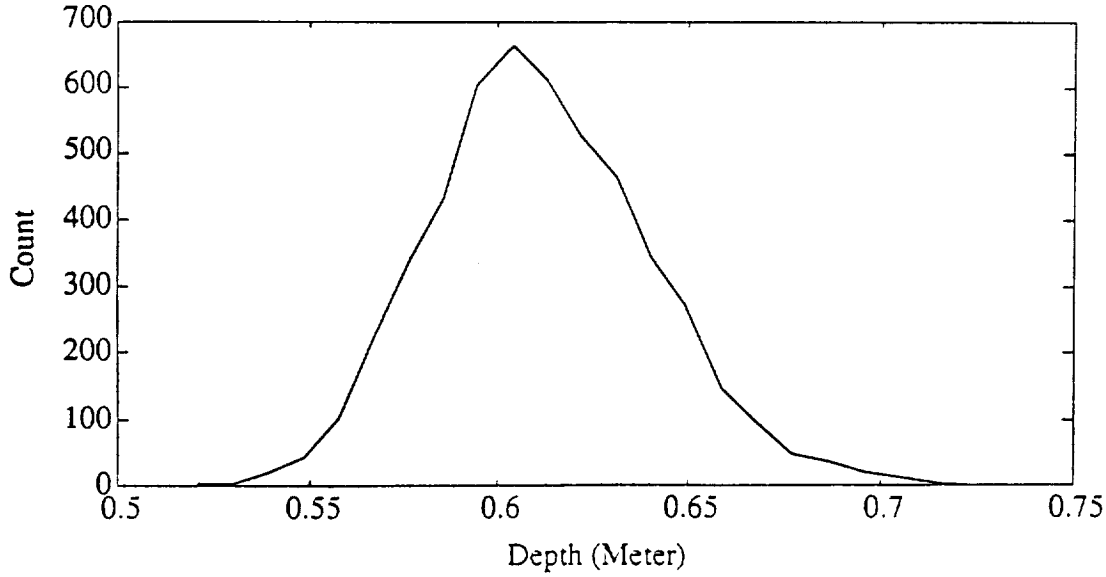

Figure 5.29 - Distribution of Estimated Depth from Filtered Displacement for 0.61 meters

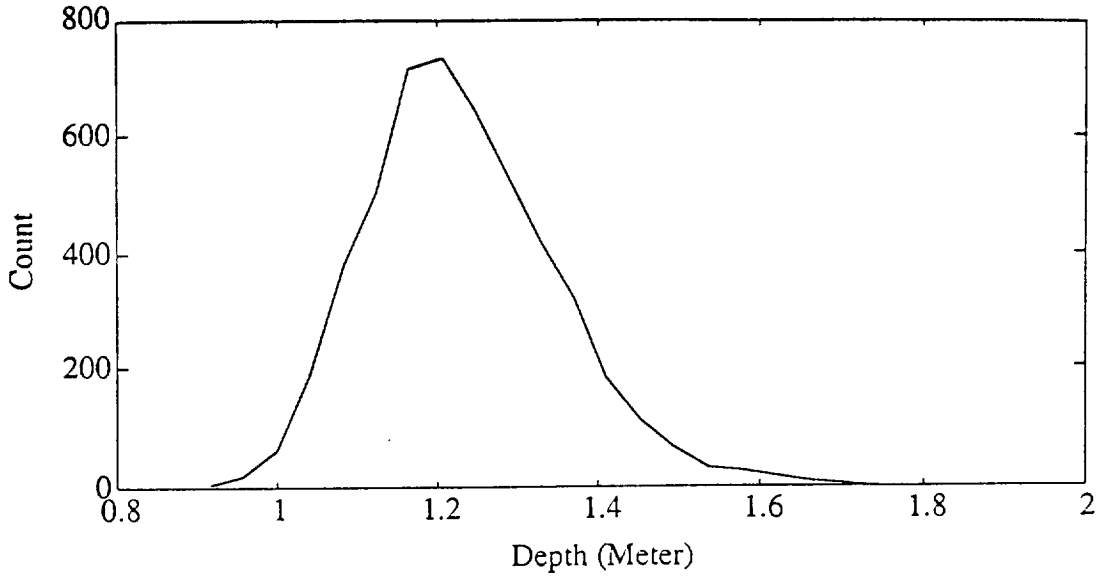

Figure 5.30 - Distribution of Estimated Depth from Filtered Displacement for 1.22 meters

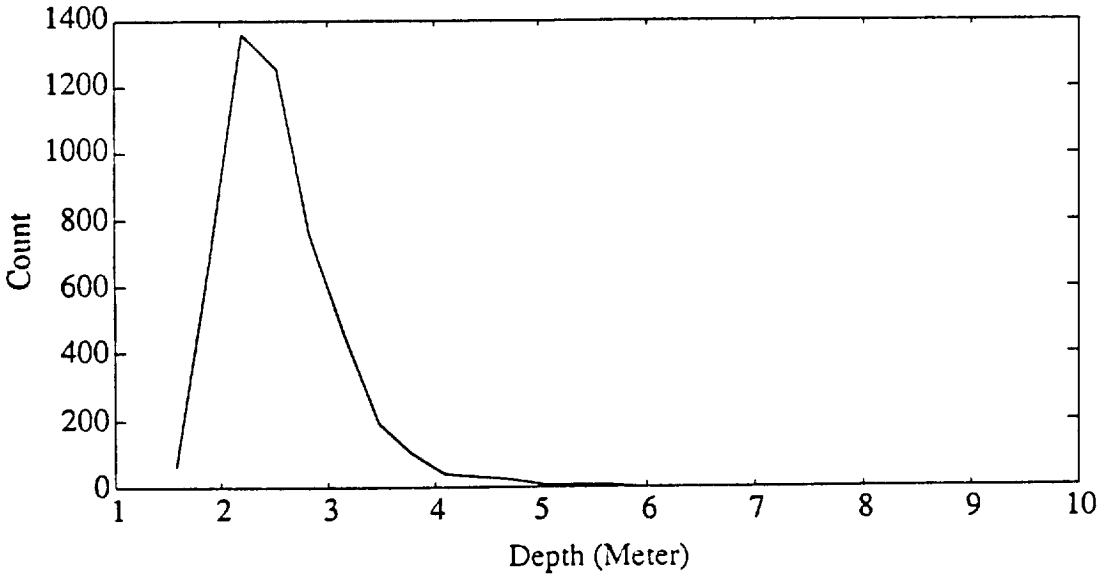

Figure 5.31 - Distribution of Estimated Depth from Filtered Displacement for 2.44 meters

5.8. Parallel Processing and ROMKF

Experiments were run to determine the effectiveness of the Row Method ROMKF. The algorithm was tested on the Sequent Balance 2100. This multi-CPU shared bus system is a MIMD computer with 16 processing elements (PE). Each PE is built around the National Semiconductor 32032 processor with a total 16 Mbytes of shared memory and 1 Gbyte of disk based memory. Each PE communicates with the other PE's by shared memory and synchronizes by the use of locks. Each PE operates at 0.75 Mips.

As a means of comparing the efficiency of the parallel version of the ROMKF, the results of the Row Method, presented in Section 4.5, are compared to the Table Method proposed by Potter *et al.* in [52, 66] to implement the ROMKF in parallel. The Table Method works by setting up a lookup table in shared memory. The lookup table is constructed by the use of complex mathematical equations describing the dimensions of the support model and the size of the input image. The table contains a list of those pixels that may be processed in parallel and a second list of which parallel iteration the computation should take place. The lookup table is consulted by all PE's. Figure 5.32 shows the total processing time in seconds for filtering a 256x256 field with the ROMKF.

When one processor is used to process the field, the Row Method requires 14.3 seconds which is more than four times faster than the time required by the Table Method. Additionally, the Row Method can process the field in 8.6 seconds when 3 processors are used whereas the Table Method requires 11 processors to equal this performance. In all cases the

Row Method is more rapid than the Table Method. The total processing time continues to decline as the number of processors increases and levels off at 4.6 seconds when 14 processors are used. Additional PE's do not contribute to a reduction in processing time due to increased bus contention.

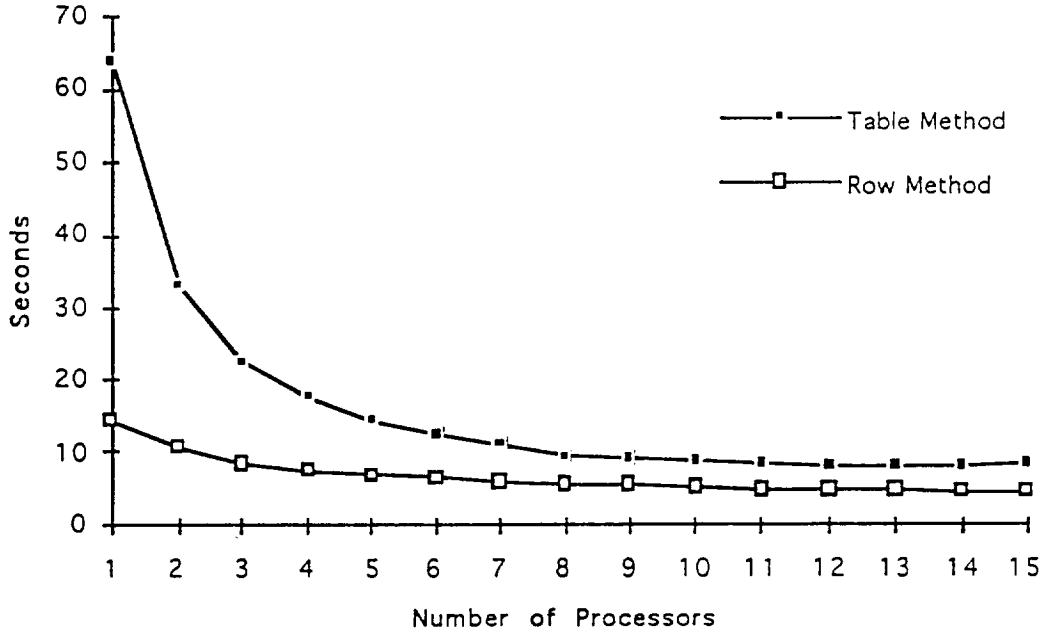

Figure 5.32 - Parallel Execution Time for the ROMKF

The consistently lower processing time of the Row Method ROMKF is mostly due to its efficiency in communications along the shared bus. Efficient message passing is used to initiate and coordinate multi-PE activity. No global tables need be consulted as in the Table Method. The Row Method uses localized control which requires less memory overhead and provides for a reduction in memory contention. The Row Method is also more flexible in that the allocation of PE's is done at runtime rather than precomputed as in the Table Method. This allows the Row Method to be

adapted to different field sizes as well as different model supports without modifications to the algorithm.

The ROMKF algorithm used in the reduction of distortions in depth and displacement fields can benefit greatly by the use of parallel processing. The Row Method ROMKF provides a practical, real situation application in which parallel processing is very effective. The large data set, due to the size of the depth and displacement fields, may be processed sequentially with the time required being proportional to the number of data points, N. The Row Method provides for rapid execution of the ROMKF routine by recognizing the non-dependence between calculations of fixed locations in the field and by performing many of the operations in parallel.

5.9. Summary

From the results it is clear that restoration of depth and displacement fields by establishing a model to describe the underlying field and the observation process can provide results superior to conventional methods. An oversmoothing problem introduced by a spatially invariant system model applied to a field that contains discontinuities is solved by the use of multiple models tuned to the directions of the discontinuities. Considering the nonlinear relation between the displacement and the depth field, filtering the displacement field and then calculating the depth map is preferred. A highly successful parallel processing version is presented in detail to address the time critical needs of on-line processing.

CHAPTER 6

6. Summary and Future Research Areas

6.1. Research Summary

In this thesis, a method to restore distorted depth and displacement fields based on modeling the underlying field and the observation equation is presented. The modeling provides a means of obtaining more accurate and reliable field results than current non-model based estimation algorithms. Significant reduction in the variance of the fields was obtained by the use of the ROMKF with 1x1x1 NSHP support region. The improvements due to the ROMKF restoration are greater when smaller patch regions are utilized in the displacement estimation algorithm. This is due to the smoothing effects of the larger patch regions. The model parameters were estimated from the distorted fields and were found to be stable over a wide range of intensity image noise levels.

Provisions are presented by which direct or indirect observations may be entered into the observation equation. Direct observation deals with an observation equation based on the actual corrupted input intensity images. Indirect observation uses measurements on the distorted depth and displacement fields obtained through an external source, such as those provided through a stereo region matching algorithm. Direct observations based on single intensity values are shown to be more sensitive to intensity image noise than observations based on patch regions in the image. The averaging effect of the patch region operations reduces the sensitivity to image noise.

The problems encountered with oversmoothing of edge boundaries in the depth and displacement fields by the use of a single spatially invariant system model were overcome by the use of multiple models tuned to the directions of the discontinuities. The multiple model approach allows for distortion reduction while maintaining clear discontinuities or sharp edges in the restored fields. Results from processing images representative of robotic scenarios, i.e., images that contain distinct object boundaries such as in the block images, clearly show the multiple model's ability to preserve discontinuities in the field.

Prefiltering the intensity images prior to the depth and displacement estimation stage is shown to yield a measurable noise reduction in the restored field. However, filtering of the displacement field estimates produces a more significant improvement in the results (regardless of whether the input intensity images are prefiltered or unfiltered). A comparison was made to show that the restoration of the displacement field estimates provides superior results to those obtained by just filtering the input intensity images.

Since image processing deals with large data sets due to the size of the images processed, a parallel version of the restoration procedure based on the issue of data independence is presented. The results show that dramatic savings in total computational time are possible with multiple CPU concurrent processing. This is particularly important for applications that have time constraints imposed on the processing of data.

6.2. Recommendations for Future Research

Issues of adaptive parameter estimation may provide better modeling to the actual formation of the depth and displacement fields. In this thesis, the spatial support coefficients were found by a least squares fit to the entire field or based on a multiple model approach. A continuous adaptive identification method as is being developed by Koch [36] may provide better modeling of the underlying field than either of these approaches alone. Additionally, the temporal coefficient was fixed for the entire field. A method that tracks the values over a sequence of fields would allow for adaptive temporal coefficients based on the estimates of the correlation on a per pixel basis.

The issues involved with blurring of input intensity images and estimation of depth and displacement need to be addressed. Effective results may be forthcoming if the blurring effects were to be accounted for in a prefiltering stage to avoid the non-linear effects which they generate in the depth calculations. Extensions to the observation equations presented in Chapter 3 to account for these effects should be investigated.

Although parallel versions of the ROMKF have been presented, a significant amount of computation is required to obtain the estimation of the displacements fields between stereo images. To be effective in "real-time" scenarios, this operation must be carried out more rapidly. Since the SSD algorithm requires only simple calculations, a parallel version based on a highly pipelined architecture would seem appropriate for robotic assembly tasks.

Appendix

Appendix 2-D Kalman Filtering of Intensity Images

Kalman filtering is based on modeling the dynamics of a system with a state space model. In applications involving images, a scan line ordering was used by Woods and Radewan [67] and later by Woods and Ingle [65] to derive the Kalman filter for two dimensions. The recursive structure of the Kalman filter establishes a causality in the data. A raster scan pixel ordering still maintains only one direction of recursion. The state and observation equations for 2-D images are:

$$\mathbf{s}(m,n) = \mathbf{F}\mathbf{s}(m-1,n) + \mathbf{G}\mathbf{u}(m,n) + \mathbf{E}\mathbf{w}(m,n) \tag{A.1}$$

$$r(m,n) = \mathbf{H}\mathbf{s}(m,n) + v(m,n) \tag{A.2}$$

In the system state equation, $\mathbf{s}(m,n)$ represents the state vector at location (m,n), $\mathbf{w}(m,n)$ is a forcing or noise term to account for uncertainties in the modeling, $\mathbf{u}(m,n)$is a deterministic input or control, $\mathbf{F}$ is the transition matrix, and $\mathbf{G}$ and $\mathbf{E}$ are system matrices. In the observation equation, $\mathbf{H}$ is the observation matrix and $v(m,n)$ accounts for observation noise. $\mathbf{w}(m,n)$ and $v(m,n)$ are assumed to be uncorrelated zero-mean white Gaussian noise processes with covariances $\mathcal{Q}_w$ and $\mathcal{Q}_v$ respectively. The state error covariance matrix, $\mathbf{P}$, is defined as:

$$\mathbf{P}(m,n) = \mathcal{E}\left[\left(\mathbf{s}(m,n) - \hat{\mathbf{s}}(m,n)\right)\left(\mathbf{s}(m,n) - \hat{\mathbf{s}}(m,n)\right)^{T}\right] \tag{A.3}$$

where $\mathcal{E}[\cdot]$ denotes the expectation operator.

The Kalman filter is accomplished in two steps: extrapolation or prediction and update. The subscripts (-) and (+) are used to denote

immediately before and after the discrete measurement, respectively. The state and error covariance prediction equations are:

$$\hat{\mathbf{s}}_-(m,n) = \mathbf{F}\hat{\mathbf{s}}_+(m-1,n) + \mathbf{G}\mathbf{u}(m,n) \tag{A.4}$$

$$\mathbf{P}_-(m,n) = \mathbf{F}\mathbf{P}_+(m-1,n)\mathbf{F}^T + \mathbf{D}\mathcal{Q}_w\mathbf{D}^T \tag{A.5}$$

The state estimate and error covariance update equations are evaluated as:

$$\hat{\mathbf{s}}_+(m,n) = \hat{\mathbf{s}}_-(m,n) + \mathbf{K}(m,n)\left[r(m,n) - \mathbf{H}\hat{\mathbf{s}}_-(m,n)\right] \tag{A.6}$$

$$\mathbf{P}_+(m,n) = \left[\mathbf{I} - \mathbf{K}(m,n)\mathbf{H}\right]\mathbf{P}_-(m,n) \tag{A.7}$$

where $\mathbf{K}(m,n)$ is the Kalman gain found by:

$$\mathbf{K}(m,n) = \mathbf{P}_-(m,n)\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_-(m,n)\mathbf{H}^T + \mathcal{Q}_v\right] \tag{A.8}$$

The error covariance may also be found by the recursive Riccati equation given by:

$$\begin{aligned}\mathbf{P}_-(m,n) = \mathbf{F}\big[&\mathbf{P}_-(m-1,n) - \\ &\mathbf{P}_-(m-1,n)\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_-(m-1,n)\mathbf{H}^T + \mathcal{Q}_v\right)^{-1}\mathbf{H}\mathbf{P}_-(m-1,n)\big]\mathbf{F}^T \\ &+ \mathbf{E}\mathcal{Q}_w\mathbf{E}^T\end{aligned} \tag{A.9}$$

Gelb [22] and Anderson [5] provide detailed information on the derivation of the 1-D Kalman filter and its applications. Current applications to image restoration may be found in [8].

Bibliography

1. Aggarwal, J. K. and Nandhakumar, N. "On the Computational of Motion from Sequences of Images." *Proceedings of the IEEE* **76** (Aug 1988): pp. 917-935.

2. Aggarwal, J. K. and Wang, Y. F. "Analysis of a Sequence of Images using Point and Line Correspondences"." in *Proceedings of the 1987 IEEE Conference on Robotics and Automation*. Raleigh, NC, by IEEE: March, 1987, pp. 1275-1280.

3. Albus, J. S. and Hong, T. H. "Motion, Depth, and Image Flow." in *Proceedings of the 1990 IEEE International Conference on Robotics and Automation*. Cincinnati, OH, by IEEE: May 13-18, 1990, pp. 1161-1170.

4. Anandan, P. "Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion." in *IUS Workshop*. by DARPA: Dec, 1985, pp. 236-246.

5. Anderson, B. D. O. and Moore, J. B. *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall Inc., 1979.

6. Angwin, D. L. *Adaptive Image Restoration Using Reduced Order Model Based Kalman Filters, Image Processing Laboratory*. PhD Thesis, Rensselaer Polytechnic Institute, IPL TR-107, 1989.

7. Angwin, D. L. and Kaufman, H. "Image Restoration Using Reduced Order Models." *Signal Processing* **16** (Jan 1989): pp. 21-28.

8. Angwin, D. L. and Kaufman, H. "Nonhomogeneous Images Identification and Restoration Procedures". In *Digital Image Restoration,* ed. A. K. Katsaggelos. Springer-Verlag, 1991. pp. 177-208.

9. Baker, H. H. "Edge Based Stereo Correlation." in *Proc. ARPA Image Understanding Workshop*. University of Maryland, by ARPA: April, 1980, pp.168-175.

10. Ballard, D. H. and Kimball, O. A. "Rigid Body Motion from Depth and Optical Flow." *Computer Vision, Graphics, and Image Processing* **22** (1983): pp. 95-115.

11. Biemond, J., Looijenga, L. and Boekee, D. E. "A Pel-Recursive Wiener-Based Displacement Estimation Algorithm." *Signal Processing* **13** (1987): pp. 399-412.

12. Bridwell, N. J. and Huang, T. S. "A Discrete Spatial Representation for Lateral Motion Stereo." *Computer Vision, Graphics, and Image Processing* **21** (1983): pp. 33-57.

13. Burt, P. J. and Adelson, E. H. "The Laplacian Pyramid as a Compact Image Code." in *Conference on Pattern Recognition and Image Processing*. Dallas, TX, by IEEE: 1982, pp. 671-679.

14. CIRSSE. *Annual Report*. Rensselaer Polytechnic Institute: 1991.

15. Damour, K. T. *Parallel Implementation of the Kalman Filter for Noise and Blur Removal from Digital Images*. Master's Thesis, Rensselaer Polytechnic Institute, 1988.

16. Daniilidis, K. and Nagel, H.-H. "Analytical Results on Error Sensitivity of Motion Estimation from Two Views." *Image and Vision Computing* 8, No. 4. (Nov 1990): pp. 297-303.

17. Driessen, H. in preparation, Ph.D. Thesis, TU Delft, Netherlands, 1992.

18. Driessen, J. N. and Biemond, J. "Motion Field Estimation by 2-D Kalman Filtering". In *Signal Processing V: Theories and Applications,* ed. L. Torres, E. Masgrau and M. A. Lagunas. Elsevier Sciences Publishers B.V., 1990. pp 975-978.

19. Driessen, J. N., Boroczky, L. and Biemond, J. "Pel-recursive Motion Field Estimation from Image Sequences." *Journal of Visual Communication and Image Representation* **2**, No. 3. (Sept. 1991): pp. 259-280.

20. Efstratiadis, S. N. and Katsaggelos, A. K. "A Multiple-frame Pel-recursive Wiener-based Displacement Estimation Algorithm." *SPIE Visual Communications and Image Processing IV* **1199** (1989): pp. 51-60.

21. Enkelmann, W. "Obstacle Detection by Evaluation of Optical Flow Fields from Image Sequences." *Image and Visual Computing* **9** (June 1991): pp. 160-168.

22. Gelb, A. , ed. *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.

23. Gonzalez, R. C. and Woods, R. E. *Digital Image Processing*. Reading, MA: Addison-Wesley Publishing Co., 1992.

24. Heeger, D. J. and Jepson, A. D. "Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation." *Neural Computation* **2** (1990): pp. 127-135.

25. Heeger, D. J. and Jepson, D. "Visual Perception of Three-Dimensional Motion." *Neural Computation* **2** (1990): pp. 129-137.

26. Horn, B. K. P. *Robot Vision*. Cambridge, MA: The MIT Press, 1986.

27. Horn, B. K. P. and Schunck, B. G. "Determining Optical Flow." *Artificial Intelligence* **17** (1981): pp. 185-203.

28. Jain, A. K. "Advances in Mathematical Models for Image Processing." *Proceedings of the IEEE* **69**, No. 5. (May 1981): pp. 502-528.

29. Kalivas, D. S. and Sawchuk, A. A. "Motion Compensated Enhancement of Noisy Image Sequences." in *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing*. by IEEE: 1990, pp. 2121-2124.

30. Katsaggelos, A. K., Driessen, J. N., Efstratiadis, S. N. and Lagendijk, R. L. "Spatio-temporal Motion Compensated Noise Filtering of Image Sequences." *SPIE Visual Communications and Image Processing IV* Proc. SPIE **1199** (1989): pp. 61-70.

31. Kaufman, H. "Maximum Liklihood Identification of Image Parameters." Personal Communication. 1987.

32. Kaufman, H., Woods, J. W., Dravida, S. and Tekalp, A. M. "Estimation and Identification of Two Dimensional Images." *IEEE Transactions on Automatic Control* **28** (July 1983): pp. 745-756.

33. Kearney, J. K., Thompson, W. B. and Boley, D. L. "Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, No. 2. (March 1987): pp. 229-244.

34. Kim, S. P., Bose, N. K. and Valenzuela, H. M. "Recursive Reconstruction of High Resolution Image from Noisy Undersampled Multiframes." *IEEE Transaction on Acoustics, Speech, and Signal Processing* **38**, No. 6. (1990): pp. 1013-1027.

35. Kleihorst, R. P., Lagendijk, R. L. and Biemond, J. "Non-linear Filtering of image Sequences using Order Statistics." Personal Communication. 1992.

36. Koch, S. *Image Restoration Using Extended Kalman Filter*. Proposal for Ph.D. Thesis, Rensselaer Polytechnic Institute, 1990.

37. Lagendijk, R. L. *Iterative Identification and Restoration of Images*. Ph.D. Thesis, Techische Universiteit Delft, 1990.

38. Lagendijk, R. L., Biemond, J. and Boekee, D. E. "Simultaneous Image Identification and Restoration Using the EM-Algorithm." in *Proc. SPIE Conference on Visual Communication and Image Processing*. Cambridge, MA, by 1988, pp. 2-9.

39. Laing, P., Chang, Y. L. and Hackwood, S. "Adaptive Self-Calibration of Vision-Based Robot Systems." *IEEE Transactions on Systems, Man, and Cybernetics* **19**, No. 4. (July/Aug 1989): pp. 811-824.

40. Lee, D. "Some Computational Aspects of Low-Level Computer Vision." *Proceedings of the IEEE* **76**, No. 8. (Aug 1988): pp. 890-898.

41. Martinez, D. M. *Model-Based Motion Estimation and its Application to Restoration and Interpolation of Motion Pictures*. MIT, EECS Dept.: Technical Report No. 530, 1987.

42. Matthies, L. *Dynamic Stereo Vision*. CMU: Report No. CS-89-195, 1989.

43. Matthies, L., Kanade, T. and Szeliski, R. "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences." *International Journal of Computer Vision* **3**(1989): pp. 209-236.

44. Nagel, H. H. "On the Estimation of Optical Flow: Relations between Different Approaches and Some New Results." *Artificial Intelligence* **33** (1987): pp. 299-324.

45. Nagel, H. H. and Enkelmann, W. "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences." *IEEE Transaction on Pattern Analysis and Machine Intelligence* **PAMI-8**, No. 5. (September 1986): pp. 565-593.

46. Netravali, A. N. and Robbins, J. D. "Motion-compensated television coding: Part 1." *Bell Systems Technical Journal* **58**, No. 3. (March 1979): pp. 631-670.

47. Netravali, A. N. and Robbins, J. D. "Motion-Compensated Coding: Some New Results." *Bell System Technical Journal* **59**, No. 9. (Nov 1980): pp. 1735-1745.

48. Nomura, A., Miike, H. and Koga, K. "Field Theory Approach for Determining Optical Flow." *Pattern Recognition Letters* **12**, No. 3. (March 1991): pp. 183-190.

49. Olsen, S. I. "Stereo Correspondence by Surface Reconstruction." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, No. 3. (1990): pp. 309-315.

50. Papanikolopoulos, N. P. and Khosla, P. K. "Using Control and Vision for Space Applications." in *Third Annual Conference on Intelligent Robotic Systems for Space Exploration*. Troy, NY, by IEEE Computer Society: Nov 18-19, 1991, pp. 46-60.

51. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*. Communication and Information Theory, ed. S. W. Director. New York, NY: McGraw-Hill Publishing, 1984.

52. Potter, D. J., Damour, K. T. and Gilder, M. G. "Two-Dimensional Kalman Filtering using Parallel Processors." in *Proceedings of the International Supercomputer Institute*. Boston, MA, by International Supercomputing Institute: May, 1988, pp. 390-395.

53. Schalkoff, R. J. *Digital Image Processing and Computer Vision*. New York, NY: Wiley, 1989.

54. Schunck, B. G. "The Image Flow Constraint Equation." *Computer Vision, Graphics, and Image Processing* **35** (1986): pp. 20-46.

55. Schunck, B. G. "Image Flow Segmentation and Estimation by Constraint Line Clustering." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, No. 10. (Oct. 1989): pp. 1010-1027.

56. Simpson, B. *Reduced Model Kalman Filter*. Master's Thesis, Rensselaer Polytechnic Institute, 1989.

57. Stuller, J. and Krishnamurthy. "Kalman Filter Formulation of Low-Level Television Image Motion Estimation." *Computer Vision, Graphics, and Image Processing* **21** (1983): pp. 169-204.

58. Szeliski, R. "Bayesian Modeling of Uncertainty in Low-level Vision." *International Journal of Computer Vision* **5**, No. 3. (1990): pp. 271-301.

59. Tekalp, A. M. *Identification and Restoration of Noisy and Blurred Images*. Ph.D. Thesis, Rensselaer Polytechnic Institute, 1984.

60. Tekalp, A. M., Kaufman, H. and Woods, J. W. "Edge-adaptive Kalman Filtering for Image Restoration with Ringing Suppression." *IEEE Transactions on Acoustic, Speech, and Signal Processing* **37**, No. 6. (June 1989): pp. 892-899.

61. Terzopoulos, D. "Regularization of Inverse Visual Problems Involving Discontinuities." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, No. 4. (July 1986): pp. 413-424.

62. Van Trees, H. L. *Detection, Estimation, and Modulation Theory: Part 1*. New York, NY: John Wiley & Sons, 1968.

63. Wildes, R. P. "Direct Recovery of Three-Dimensional Scene Geometry from Binocular Stereo Disparity." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, No. 8. (August 1991): pp. 761-774.

64. Wohn, K. Y., Wu, J. and Brockett, R. W. "A Contour-Based Recovery of Image Flow: Iterative Transformation Method." *IEEE*

Transactions on Pattern Analysis and Machine Intelligence **1 3**, No. 8. (August 1991): pp. 746-760.

65. Woods, J. W. and Ingle, V. K. "Kalman Filtering in Two-Dimensions - Further Results." *IEEE Transactions on Acoustic, Speech, and Signal Processing* **29** (April 1981): pp. 188-197.

66. Woods, J. W., Potter, D. J. and Kaufman, H. "Parallel Realization of 2-D Recursive Kalman Filters." in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. Dallas, TX, by IEEE: 1987,

67. Woods, J. W. and Radewan, C. H. "Kalman Filtering in Two Dimensions." *IEEE Transactions on Information Theory* **IT-23**, No. 4. (1977): pp. 473-482.