To the Graduate Council:

I am submitting herewith a dissertation written by Lang Hong entitled "3-D SCENE RECONSTRUCTION FROM BINOCULAR SEQUENCES OF NOISY IMAGES USING OPTIMAL INFORMATION FUSION." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Dragana Brzakovic, Major Professor

We have read this dissertation
and recommend its acceptance:

M. A. Abidi

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

3-D SCENE RECONSTRUCTION FROM BINOCULAR SEQUENCES OF NOISY IMAGES USING OPTIMAL INFORMATION FUSION

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Lang Hong

December 1989

DEDICATION

This dissertation is dedicated to my wife, Xiu-mei Luo, and to my son, Steven Hong.

ACKNOWLEDGEMENTS

The author wishes to express his profound appreciation to his major advisor, Dr. Dragana Brzakovic, for the privilege of working with her, as well as the invaluable support, advice, and unwavering patience and encouragement that she has given to him throughout his doctoral program. Many thanks are due to Dr. W. L. Green, Dr. J. M. Bailey, Dr. M. A. Abidi, and Dr. K. Soni for serving on author's dissertation committee, reviewing this dissertation and teaching him so much throughout his graduate study.

Special thanks are extended to the University of Tennessee, Knoxville, and the Department of Electrical and Computer Engineering for providing an excellent graduate study program and research environment and equipment. The author is also grateful to the Martin Marietta Corporation for supporting with a graduate research assistantship through out this research.

The author is deeply indebted to his parents for their sturdy motivation and strong support. Finally, and most importantly, the author would like to thank his wife, Xiu Mei Luo, for her constant support and encouragement throughout his graduate study period.

ABSTRACT

The objective of this work is to develop a vision system for 3-D scene reconstruction from sequences of binocular noisy images. The vision system assumes that the positions of the cameras are known with uncertainties, and uses the camera parameters in the process of 3-D scene reconstruction. In contrast to other vision systems that utilize multiple images, this system does not establish correspondence between pixels in image pairs, but considers all possible correspondences (based on camera parameters) and assigns a confidence value to each possible correspondence. Confidence values are assigned only to significant intensity changes. The confidences are a function of the number of possible correspondences and similarity in intensity changes of candidate pixels. For computational efficiency, assignment of the confidence values is accomplished in a hierarchical fashion, by considering images at the lower resolution first. Each significant intensity change at the lower resolution is associated with a group of pixels at the higher resolution. This association is the function of the multi-resolution image generation process, in this case Gaussian pyramid, and it determines the positions of the pixels to be considered when establishing confidences at the higher resolution. At a particular level of the hierarchy, the vision system incorporates the information provided by image sequences by fusing confidence values assigned to possible correspondences in sequences of image pairs. The results of fusion guide matching at the next level of the resolution hierarchy.

The fusion process employs an optimal distributed fusion algorithm. The

dynamic models for information fusion are derived using 3-D geometrical modeling assuming probabilistic description of 3-D uncertainty volumes. The fusion process is dynamic so that the fused confidences are continuously updated when new information is available. In the fusion process, the confidence values of the correct correspondences are increased and the confidence values of false correspondences are decreased. At the end of the fusion process, the correspondences with high confidence values provide correct 3-D information about the scene. The objects in the scene are reconstructed by using moving-least-square surface interpolation over a sparse set of calculated 3-D points.

The experimental results obtained are very satisfactory. For the binocular noisy image sequences with a noise standard deviation of 20, the percentage of correct matches is over 95% and the average error of calculated 3-D distances to the objects is less than 5%. The vision system is robust to the presence of noise in the images, since it fuses information provided by the image pairs and image sequences. It is also computationally efficient since it employs a resolution hierarchy and utilizes the distributed fusion algorithms which can be implemented in parallel.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION 1

1.1 Review of Relevant Work 2

1.1.1 3-D Reconstruction Using Binocular Vision 2

1.1.2 Structure from Motion 5

1.1.3 Information/Multi-Sensor Fusion 9

1.2 Outline of the Dissertation 12

1.3 Main Contributions of This Work 13

2. VISION SYSTEM ARCHITECTURE 15

2.1 System Architecture and System Description 16

2.2 Hierarchical Data Structure 19

2.2.1 Creation of the Pyramid Structure 19

2.2.2 Top to Bottom Predictions for Matching 25

3. MATCHING AND ASSIGNMENT OF CONFIDENCE VALUES 27

3.1 Feature Extraction 29

3.2 Matching and Assignment of Confidence Values 33

3.2.1 Pairwise and Sequencewise Matching 33

CHAPTER PAGE

3.2.2 Assignment of Confidence Values in the Top First Image Pair 43
3.2.3 Assignment of Confidence Values in Other Image Pairs . . 44
3.2.4 Refinement of Confidence Values 47

4. GEOMETRICAL MODELING 50
4.1 Geometrical Modeling for a Pair of Images 50
4.1.1 A 3-D Uncertainty Volume from a Possible Match 51
4.1.2 Calculation of Gaussian Parameters 57
4.2 Geometrical Modeling for Sequences of Images 59
4.2.1 Transformation and Calculation of Geometrical Structures for Sequences of Images 60
4.2.2 Calculation of the Strength of Links between Spheres . . 67

5. MODEL-BASED INFORMATION FUSION 76
5.1 Concepts of Information Fusion 76
5.2 Dynamic Models for Fusion 78
5.2.1 Dynamic Models for Confidence Value Fusion 78
5.2.2 Dynamic Models for 3-D Information Fusion 82
5.3 Fusion Algorithms 84
5.3.1 An Algorithm for Confidence Value Fusion 85
5.3.2 An Algorithm for 3-D Information Fusion 89

6. EXPERIMENTAL RESULTS 93

CHAPTER PAGE

6.1 Acquisition of Binocular Image Sequences 93

6.2 Preprocessing 100

6.3 Matching, Assignment of Confidence Values and Information Fusion 100

7. CONCLUSIONS 127

BIBLIOGRAPHY 129

APPENDIX 141

MULTI-SENSOR/INFORMATION FUSION ALGORITHMS . . 142

VITA 172

LIST OF FIGURES

FIGURE		PAGE
2.1	The architecture of the vision system.	17
2.2	An image pyramid.	21
2.3	Candidate son nodes for a father node. A weighted sum of sixteen son nodes at l_1 creates one father node at l_2.	23
2.4	Candidate father nodes for a son node. Each son node at l_1 has four father nodes at l_2. Son nodes S_1, S_2, S_3, and S_4 at l_1 have same the father nodes at l_2.	24
3.1	An input image.	30
3.2	The strength map of the zero-crossings of the input image.	31
3.3	The sign map of the zero-crossings of the input image.	32
3.4	Calculation of the window size for matching.	35
3.5	Pattern numbers for 8 and 24 neighborhoods: (a) 8 neighborhood, and (b) 24 neighborhood.	38
3.6	Two examples for calculating pattern differences.	40
3.7	Weighted averaging of the confidence values: (a) a specified range in the neighborhood of $P(i,j)$, and (b) a distribution of weights.	48
4.1	A 3-D uncertainty volume in the geometrical modeling for a pair of images.	52
4.2	Gaussian probability description of the 3-D uncertainty volume.	54

FIGURE **PAGE**

4.3 Spherical Gaussian probability description of the 3-D uncertainty volume. 56

4.4 A 3-D uncertainty region in the geometrical modeling for a pair of images. 58

4.5 Intersected 3-D uncertainty volumes in the geometrical modeling for sequences of binocular images. 61

4.6 Transformation between coordinate systems and prediction of locations of pixels in next image pair where the sphere C_{k+1} can be seen. 65

4.7 Two extracted 3-D uncertainty volumes. 68

4.8 Case two in calculating intersected area S^i. 72

4.9 Case three in calculating intersected area S^i. 73

4.10 Case four in calculating intersected area S^i. 74

4.11 Case five in calculating intersected area S^i. 75

5.1 Determination of the measurement error due to the uncertainties in camera's motion. 81

5.2 A signal flow chart for hierarchical fusion of confidence values. . . 88

5.3 A signal flow chart for 3-D information fusion. 92

6.1 A Cincinnati Milacron $T^3 - 726$ industrial robot with 6 degrees of freedom and a CCD camera mounted on the robot. 94

6.2 The first two noise free image pairs of $\mathcal{S}_1$. Upper: the first pair, and lower: the second pair. 96

6.3 The first two noise free image pairs of $\mathcal{S}_2$. Upper: the first pair, and lower: the second pair. 97

6.4 The first two noisy image pairs of $\mathcal{S}_1$ ($\sigma = 15$). Upper: the first pair, and lower: the second pair. 98

6.5 The first two noisy image pairs of $\mathcal{S}_2$ ($\sigma = 20$). Upper: the first pair, and lower: the second pair. 99

6.6 A 3-level pyramid of signs of zero-crossing of noisy image $L_{1\mathcal{S}_1}$. . 101

6.7 A 3-level pyramid of signs of zero-crossing of noisy image $L_{1\mathcal{S}_2}$. . 102

6.8 The strength map of the noisy image $L_{1\mathcal{S}_1}$. 103

6.9 The sign map of the noisy image $L_{1\mathcal{S}_1}$. 104

6.10 The strength map of the noisy image $L_{1\mathcal{S}_2}$. 105

6.11 The sign map of the noisy image $L_{1\mathcal{S}_2}$. 106

6.12 Possible matches found in the first pair of $\mathcal{S}_1$ and the regions predicted in the second pair at top level. (a) Possible matches found in the first image pair, and (b) regions predicted in the second image pair. 109

6.13 Regions in the first image pairs of $\mathcal{S}_1$ predicted by the possible matches found in the first image pair at the top level. 110

6.14 Error covariances of the fusion process. 111

6.15 Confidence values as a function of time for a correct match in the first image pair at the top level. 113

6.16 Confidence values as a function of time for a false match in the first image pair at the top level. 114

6.17 Correct matches found for $\mathcal{S}_1$ using a 3-level hierarchy. 115

6.18 Correct matches found for $\mathcal{S}_2$ using a 3-level hierarchy. 116

6.19 Correct matches found for $\mathcal{S}_1$ using a 2-level hierarchy. 117

6.20 Correct matches found for $\mathcal{S}_2$ using a 2-level hierarchy. 118

6.21 Reduction of 3-D distance uncertainty. 120

6.22 Selected matches in $\mathcal{S}_1$ whose 3-D distances are given by Table 1. 121

6.23 Selected matches in $\mathcal{S}_2$ whose 3-D distances are given by Table 2. 123

6.24 Reconstructed object in sequence $\mathcal{S}_1$ by moving least square surface interpolation. 125

6.25 Reconstructed object in sequence $\mathcal{S}_2$ by moving least square surface interpolation. 126

A.1 A dynamic multi-sensor system. 144

A.2 A static multi-sensor system. 146

A.3 A dynamic centralized estimator. 156

A.4 A signal flow chart for dynamic multi-sensor/information fusion. 166

A.5 The static distributed estimation. 169

LIST OF SYMBOLS

English Symbols

b	length of base line of stereo cameras
$Cf(i,j)$	confidence value
$Conf(i,j)$	confidence value after refinement
$\mathbf{C}$	covariance matrix in Gaussian probability distribution
C_k	a sphere at the kth moment
C_{k+1}^i	spheres at the $(k+1)th$ moment
$Diff_{P_{zc8}}$	differences of zero-crossing patterns in 8 neighborhood
$Diff1_{P_{zc24}}$	differences of zero-crossing patterns in the subregion 1 of 24 neighborhood
$Diff2_{P_{zc24}}$	differences of zero-crossing patterns in the subregion 2 of 24 neighborhood
$Diff_{P_{zc24}}$	total differences of zero-crossing patterns in 24 neighborhood
$Diff_{S_{zc}}$	differences of zero-crossing strengths
f	length of focus of cameras
H_{k+1}^i	observation matrices in the mesurement models for confidence value fusion

I_l	image at level l of resolution hierarchy
$I_l(i,j)$	a pixel of the image I_l
(i,j)	a specific pixel
$\mathbf{J}_k$	camera rotation matrix at the kth moment
$\mathbf{J}_{a_k}$	Jacobian
$\mathbf{K}^i_{k+1}$	local Kalman gain matrix for local 3-D information updates at the $(k+1)th$ moment
$K^{l^i}_{cf_{k+1}}$	local Kalman gains for local confidence value updates at the $(k+1)th$ moment
L^j_k	the kth image in the left sequence at level j of resolution hierarchy
$\underline{m}$	mean vector
$Nct(i,j)$	number of possible matches found in the top first image pair
$\underline{Q}^i_{k+1}$	vector representations of sphere centers at the $(k+1)th$ moment
P	a 3-D point
$P(i,j)$	a pixel
P_l, P_r	projections of the 3-D point P on the left and right image planes
P'_l	transferred coordinate of P_l
$P_l(i,j), P_r(i,j)$	left and right zero-crossing pattern numbers

$\mathbf{P}_0$	initial error covariance matrix for 3-D information fusion
$\mathbf{P}_{k\|k}$	error covariance matrix of globally-estimated 3-D information at the kth moment
$\mathbf{P}_{k+1\|k+1}$	error covariance matrix of globally-estimated 3-D information at the $(k+1)th$ moment
$\mathbf{P}_{k+1\|k}$	propagated error covariance matrix of globally-estimated 3-D information from the kth to the $(k+1)th$ moment
$\mathbf{P}^i_{k+1\|k+1}$	locally-updated error covariance matrices of local 3-D information at the $(k+1)th$ moment
P^l_{cf0}	initial error covariance of confidence values at level l
$P^l_{cf_{k\|k}}$	error covariance of globally-fused confidence values at level l at the kth moment
$P^l_{cf_{k\|k}}$	error covariance of globally-fused confidence values at level l at the $(k+1)th$ moment
$P^l_{cf_{k+1\|k}}$	propagated error covariance of globally-fused confidence values from the kth to the $(k+1)th$ moment
$P^{l^i}_{cf_{k+1\|k+1}}$	locally-updated error covariances of local confidence value updates at the $(k+1)th$ moment

$\mathbf{Q}_{J_k}$, $\mathbf{Q}_{T_k}$	covariance matrices of $\mathbf{J}_k$ and $\underline{T_k}$
q^i	coefficient proportional to the uncertainties in motion parameters
R_k^j	the kth image in the right sequence at the jth level of resolution hierarchy
R_{K+1}^i	measurement modeling error covariance matrix
r	radius of a sphere
S_{zc_l}, S_{zc_r}	left and right strengths of zero-crossings
S^i	common volume of two spheres
$T_{P_{zc}}^P$	zero-crossing pattern number threshold for pairwise matching
$T_{P_{zc}}^S$	zero-crossing pattern number threshold for sequencewise matching
$T_{S_{zc}}$	zero-crossing strength threshold for matching
$\underline{T_k}$	camera translation vector at the kth moment
$\underline{U}, \underline{U_k}$	Gaussian white noise representing spatial uncertainties
$w(p,q)$	generating kernel for image pyramid creation
w^k	weights in refining confidence values
W_j	searching window at the jth level
W_I	searching window at the bottom level
$\tilde{w}_{k+1}^i$	Gaussian white noise

$\underline{X}_k$	vector representation of a 3-D point at the kth moment
$\underline{X}_0$	initial value for 3-D information fusion
$\underline{\hat{X}}_{k\|k}$	globally-estimated 3-D information at the kth moment
$\underline{\hat{X}}_{k+1\|k+1}$	globally-estimated 3-D information at the $(k+1)th$ moment
$\underline{\hat{X}}_{k+1\|k}$	propagated globally-estiamted 3-D information from the kth to the $(k+1)th$ moment
$\underline{\hat{X}}^i_{k+1\|k+1}$	locally-updated 3-D information at the $(k+1)th$ moment
(x_l, y_l)	a pixel on the left image plane
(x_r, y_r)	a pixel on the right image plane
(x_{l_k}, y_{l_k})	a pixel on the left image plane at the kth moment
(x_{r_k}, y_{r_k})	a pixel on the right image plane at the kth moment
x_{cf_k}	state variable of confidence value at the kth moment
$\overline{x}^l_{cf0}$	initial confidence value at level l
$\hat{x}^l_{k\|k}$	globally-fused confidence value at level l at the kth moment

$\hat{x}^l_{k+1\|k+1}$	globally-fused confidence value at level l at the $(k+1)th$ moment
$\hat{x}^l_{k+1\|k}$	propagated globally-fused confidence value at level l from the kth to the $(k+1)th$ moment
$\hat{x}^{l'}_{k+1\|k+1}$	locally-updated confidence value at level l at the $(k+1)th$ moment
XYZ	reference coordinate system
$X_kY_kZ_k$	reference coordinate system at the kth moment
Z_{min}	minimum distance between the object and the cameras
$\underline{Z}^i_{k+1}$	measurements of 3-D information
$z^i_{cf_{k+1}}$	measurements of confidence values at the $(k+1)th$ moment

Greek Symbols

$\alpha_1, \alpha_2, \alpha_3$	coefficients in calculating confidence values in the top first image pair
β^i_1, β^i_2	weights of noise and uncertainties in geometrical modeling
$\gamma_1, \gamma_2, \gamma_3$	coefficients in calculating confidence values in other image pairs
$\mathbf{\Gamma}(\cdot)$	nearest integer operator

$\Delta_{i_l}, \Delta_{j_l}$	the increments of left regions in region prediction between levels of the hierarchy
$\Delta_{i_r}, \Delta_{j_r}$	the increments of right regions in region prediction between levels of the hierarchy
Δl	range for refining confidence values
$\Delta \mathbf{J}_k$	uncertainty of $\mathbf{J}_k$
$\Delta \underline{T}_k$	uncertainty of $\underline{T}_k$
$\rho_{xy}, \rho_{xz}, \rho_{yz}$	correlation coeficients
σ	standard deviation
$\sigma_x, \sigma_y, \sigma_z$	standard deviations

Math symbols

$diag\{...\}$	diagonal matrix
$E\{\cdot\}$	expectation
$erf(\cdot)$	error fuction
$N(0, \mathbf{C})$	normal distribution with zero-mean and covariance $\mathbf{C}$
$\frac{\partial}{\partial x}$	partial derivative with respect to x
∇^2	Laplacian operator
$G(x, y)$	Gaussian operator
$\nabla^2 G(x, y)$	Laplacian of Gaussian
$\lvert \cdot \rvert$	absolute value

$f * g$	convolution of f and g

CHAPTER 1

INTRODUCTION

Reconstruction of a three-dimensional (3-D) scene has been one of the main research topics in computer vision for many years. Two main methods for 3-D scene recovery, given multiple views of the scene, are binocular vision and structure from motion. As demonstrated by many applications, neither of these methods alone can give satisfactory results in real applications. This work aims at developing a vision system which overcomes the weaknesses of the binocular vision and the structure from motion methods by combining the information provided by these two methods.

To operate in unknown environments, a vision system needs to utilize all possible information to increase its adaptability. Combining the information, called here information fusion, from multiple sources/sensors provides the following:

- Compensation among sources of information, each of which can only deliver reasonably accurate information over a limited operating range
- Complete and reliable information about the scene
- Robustness to failure of an individual information source
- Robustness to noise and uncertainties existing in the information sources

The developed vision system fuses the information provided by image pairs and image sequences to acquire accuracy, reliability, and robustness.

In Sections 1.1, some current methods of binocular vision and structure from motion, and information/multi-sensor fusion are reviewed. The outline of the dissertation is given in Section 1.2 and the main contributions of this work are listed in Section 1.3.

1.1 Review of Relevant Work

In this section, the relevant work in 3-D scene reconstruction from multiple views of the scene and in information/multi-sensor fusion is reviewed.

1.1.1 3-D Reconstruction Using Binocular Vision

The essence of binocular vision methods is the establishment of correspondence between points in two images. This procedure is known as matching. There are two approaches to matching: area-based matching and feature-based matching. The area-based approach matches a neighborhood of a pixel in one image with a neighborhood in the other image. In the feature-based approach, features from each image are used for matching. The feature-based approach, which is employed in this work, is more efficient than the area-based approach. A survey of binocular vision methods can be found in [11].

1.1.1.1 Area-Based Matching

In area-based matching, a point of interest is determined first. Then cross-correlation measures are applied to search for the corresponding point of interest in the other image. The early work in area-based matching was done by Moravec [95], in which area-based correlation and a coarse-to-fine search strategy were used to find the corresponding matching points. Gennery [47,48] employed a high resolution correlator that used the matches provided by the correlation matcher and produced an improved estimate of the matching point based on the statistics of noise in the images. Cohen [26] combined region-based matching and segmentation into a cooperative scheme. The region-based techniques have a disadvantage since they use intensity values at each pixel directly. Therefore, they are sensitive to changes in absolute intensity, contrast, and illumination. Also, the presence of occluding boundaries in the correlation window is likely to confuse the correlation-based matcher.

1.1.1.2 Feature-Based Matching

In the feature-based approach, the most crucial and difficult step is to establish a correspondence between features from two images which correspond to same 3-D physical feature in space. Once the correspondence between the features in two images has been established, the 3-D information can be recovered by triangulation, provided that the positions of the cameras, the focal lengths, and the orientations of the optical axes are known.

The particular matching paradigm to choose depends on the matching features used, the number of cameras used, and the position of the cameras. Marr-Hildreth zero-crossings [85] have been used as matching features by, among others, Grimson [52,53], Mayhew and Frisby [88,89,90], Ayache and Faverjion [8,9], and Kim [72]. Some researchers have used different features, such as peaks of the magnitude of the first derivatives [98], contours [44], Canny edges [16,102], and linear edge segments [6,7,91,132]. In general, a feature in one image has more than one match in the other image; at most, one is the correct match. To remove ambiguity in matching, various constraints need to be introduced. Commonly used constraints are the disparity continuity of Eastman and Waxman [33], Grimson [52,53], Kim [71], and Marr and Poggio [83,84], the figural continuity of Mayhew and Frisby [88,89]; the disparity gradient limit of Koenderink and VanDoorn [73]; the analytic disparity fields of Eastman and Waxman [33]; the local planar/quadric patches of Hoff and Ahuja [62,63]; and the reproducing kernel-based spline smoothing of Chen and Boult [19]. Relaxation algorithms have been also applied to achieve stereo matching between features. The relaxation algorithms utilize matching information of neighborhood points in a cooperative manner and iteratively update the initial matches. The most well-known relaxation algorithms are those of Marr-Poggio [83], Barnard-Thompson [10], and Kim-Aggarwal [72]. Some researchers have introduced an additional camera into stereo vision to add a geometrical constraint for removing ambiguities in establishing possible matches [7,100,135].

Feature-based matching has several advantages over area-based matching. First, the features are associated with intensity changes rather than absolute values of intensity and therefore, better characterize the physical changes in the scene. Second, the features are more localizable. For instance, zero-crossings can be theoretically located at sub-pixel precision. Third, feature-based matching, in general, is faster than area-based matching, since there are fewer pixels involved. However, feature matching is very sensitive to the noise in the images, because the extraction of features involves second-order differentiation.

1.1.2 Structure from Motion

An image sequence taken when the camera moves provides important cues for scene reconstruction, which is also called structure from motion. When the motion of a camera is unknown, both motion parameters and the structure of the scene can be recovered. A review of the computation of motion from sequences can be found in [3]. The problem of structure from motion takes different forms based on the magnitude of motion involved, i.e., finite motion or infinitesimal motion between two successive frames. For finite motion, given a point-to-point correspondence between two successive frames, vectors are superimposed on the image plane, indicating where each point moves from one frame to the next frame. The vector field is called the displacement field. If the motion is infinitesimal, optical flow is studied.

1.1.2.1 Structure from Finite Motion

Structure from finite motion can be regarded as a stereo vision problem, wherein the feature correspondences between images taken at different times are established. However, the knowledge of motion parameters, given or recovered, can aid segmentation of image features corresponding to distinct objects and, therefore, ease the feature matching task. Ullman [123,124] used an orthographic imaging model to precisely estimate the structure and motion of an object under rigid motion. Ullman showed that the structure of the scene could be uniquely determined (up to a reflection) from the orthographically projected locations of four non-coplannar points. Although Ullman's orthographical model is adequate, it is not appropriate for real world applications where perspective projections are involved. The use of perspective transformation [5,34,68,96,107,122] increases the complexity of the problem substantially, since it results in a set of nonlinear equations. Almost all researchers in this area have concentrated on how to solve these nonlinear equations. The work of Huang, Tsai and Fang [34,68,122] stands out for being mathematically precise and the first to address the uniqueness problem in this area. Fang and Huang [34] proved that the nonlinear system developed using 5 points has a unique solution which can be iteratively solved. Tsai and Huang [122] proposed a method for finding the motion of a planar surface patch from 2-D perspective views. They also developed a method for determining motion parameters in the case of curved surfaces, using seven points. Recently, a large amount of work in this area has been reported, among which

are [37,55,75,80,128,129,130,136].

Although a great deal of effort has been focused on motion parameters and structure from finite motion, establishing the correspondence between images and finding the numerically stable solutions to the nonlinear equations remain the most serious problems in the area.

1.1.2.1 Structure from Infinitesimal Motion

When the motion between two moments is infinitesimal, the instantaneous changes in intensities in the images are analyzed to yield a dense velocity field called optical flow. The three-dimensional motion parameters and the structure of the scene can be computed based on various assumptions. No correspondences are required. The commonly used assumptions in calculating optical flow are the following: (1) optical flow field is smooth and neighboring points have similar velocities, (2) optical flow is constant over an entire segment of the image, (3) optical flow is the result of restricted motion, for instance, planar motion. Horn and Schunck [65] imposed the optical flow smoothness constraint to calculate optical flow iteratively. Yachida [134] extended Horn and Schunck's method such that the smoothness constraint was imposed in both spatial and temporal neighborhoods.

Under orthography, Hoffman [64] developed a relationship between optical flow and derivatives and local surface orientation and illustrated that if the acceleration of the optical flow is known, the computation of structure is feasible. There is much recent work concerning the recovery of the structure and motion

parameters of a moving object from its changing retinal image, under perspective projection. Longuet-Higgins and Prazdny [79] developed the relation between the optical flow and the motion parameters as well as the relationship between the gradient of the surface in view and the derivatives of the optical flow. Prazdny [105] proposed an approach in which the velocity field is decomposed into rotational and translational components. An error function of three parameters is used to evaluate the estimated motion. Minimization of the error yields the best estimate. Bruss and Horn [13,66] discussed the formulation of an iterative least-square error approach to estimate 3-D motion from optical flow. Recently reported work on calculating motion parameters and structure can also be found in [1,40,50,69,110,127,133].

Since the computation of the optical flow as well as the calculating motion and structure from optical flow require the evaluation of first and second order partial derivatives of the image brightness values, this scheme is very sensitive to the noise. Furthermore, as pointed out by Verri and Poggio [125], the optical flow, in general, is not the same as the displacement field and is often ill-suited for computing structure from motion.

In summary, the following difficulties remain in existing methods for 3-D scene reconstruction:

- Both binocular vision and 3-D structure from finite motion methods must solve the correspondence problem, which has been proven not to have a general solution.

- Both binocular vision and structure from motion (finite and infinitesimal) methods are very sensitive to noise in images.

- The structure from finite motion methods have to solve nonlinear equations. The numerical methods for solving these nonlinear equations have inherent shortcomings.

1.1.3 Information/Multi-Sensor Fusion

It has been realized by many researchers that accurate and robust information, such as 3-D information about the scene, can only be obtained by combining multiple information cues. Studying the way of combining information leads to a new and active area in computer vision call information/multi-sensor fusion. The literature of information/multi-sensor fusion in computer vision can be grouped into the following approaches:

- Fusing feature maps approaches

  These approaches have been mainly used to integrate range and intensity images. Because information from intensity images and from range images is dissimilar, it needs to be in the same form in order to be combined. The edge maps are usually adopted as a common ground for fusion. Surveys for these approaches can be found in [81,92,93]. Some typical works using thses approaches can be found in [28,29,49].

- Rule-based approaches

  These approaches use a set of rules to combine sensory data, and they are

set up for each specific application. Allen [4] implemented a hypothesize-and-test system in which the hypotheses were formed on the basis of sparse information from a vision sensor, and then refined by information from an active touch sensor. In Flynn's work [38], sonar and infrared sensors were combined according to a set of rules that determined when data from either sensor were valid and also which sensor was reliable when conflicting information was obtained. Other rule-based algorithms can be found in [29,42,43,61,76,99,118,119,126]. The advantage of using these approaches is their easy implementation.

- AI approaches:

  A good survey of AI approaches for information fusion can be found in [42]. In these approaches, the inference process is the fundamental mechanism for combining information. A significant aspect of most AI systems for information fusion is the means by which they manage their overall workload, by focusing processing attention and controlling which inferences are drawn, and when it is appropriate to make inferences [43,76,99].

- Geometric approaches

  A geometrical approach for multisensor integration was introduced by Chen [20]. Integration is performed using the spherical model on a single set of spatial information of the scene. Another geometrical fusion approach was proposed by Nakamura and Xu [97]. This approach is based on the geometrical analysis of the sensing uncertainty. The optimal fusion minimizes the

geometrical volume of the ellipsoid, which models the uncertainty among all the possible linear combinations of sensory information.

- Statistical approaches

  The statistical decision theory has been used as a means of sensor fusion by some researchers. Durrant-Whyte [30,31] used a decision team theory in sensor fusion where each sensor was treated as a decision maker. Richardson and Kenneth [106] adopted a Bayesian decision approach to combine uncertain information. The likelihood method was employed by Sher [114] to combine the outputs of several detectors for the same feature of an image. Porrill, *et al.* [103,104], utilized a statistical combination method for sensor fusion. Mann [82] used concurrent computing for high and low level multi-sensor integration. Poggio, *et al.* [101,102] utilized Markov Random Fields developed by Geman and Geman [46], to combine multiple visual cues under the guidance of intensity edges [41].

- Recursive filtering approaches

  Ayache and Faugeras [6,8,35,36] used extended Kalman filtering to optimally combine uncertainties in geometrical features. The similar work can be found in [86,87,51].

- Linear static estimation approaches

  Luo and Lin [77,78] employed a static Bayesian estimator for robot multisensor fusion, and the necessary prior information was provided by local Fisher estimators. Least square methods were used to combine different

sensory data when prior information was not available [32,112,113]. This approach gives the optimal combination of sensory data.

1.2 Outline of the Dissertation

The objective of this work is to develop a vision system which can reconstruct a 3-D scene from sequences of binocular noisy images. The system employs a resolution hierarchy to achieve computational efficiency and information fusion to obtain accurate and robust information of the scene, which make this vision system superior to the existing vision systems [27,47,87,95].

In Chapter 2, the vision system is overviewed and the architecture of the system is described. The creation of a hierarchical structure is also detailed in Chapter 2. Matching and confidence value assignment are introduced in Chapter 3. The models which are required in information fusion are derived in Chapter 4. Information fusion–confidence value fusion and 3-D information fusion are presented in Chapter 5. Chapter 6 presents the experimental results and, finaly, the conclusions of this work are given in Chapter 7. General algorithms, static and dynamic, for optimal information/multi-sensor fusion are developed in the Appendix.

1.3 Main Contributions of This Work

The main contributions of this work include the following:

- Development of a vision system which reliably reconstructs a 3-D representation of a scene from sequences of binocular noisy images. The vision system has the following unique features:

    1. It optimally fuses information provided by binocular image pairs and sequences of images by employing the dynamic, optimal information fusion algorithms developed in this work.
    2. It assigns confidence values to all possible matches instead of solving the correspondence problem between pairs of images. The confidence values are updated by the information obtained from the binocular sequences in the process of fusion. The 3-D points with high confidence values form a reconstructed 3-D scene.
    3. It reduces the uncertainties inherent in the calculation of 3-D information from binocular images by fusing the 3-D information obtained at different locations and at different moments.
    4. It is robust to the presence of noise in the images, since the fusion of information provided by the sequences cancels the effect of noise. Aside from this, it is robust to the uncertainties in the motion parameters by modeling the uncertainties into error covariances.

5. It compensates the effect of discrete photoelements in the cameras, which causes quantization errors in determining 3-D properties of the scene, by the geometrical modeling introduced in this work.

6. It is computationally efficient. This is reflected in two aspects: the utilization of a coarse-to-fine matching control strategy and the parallel nature of the information fusion algorithms.

- Development of algorithms, static and dynamic, for optimal multi-sensor/information fusion for a general sensor system where each sensor is associated with its own coordinate system and the communication networks between sensors are with uncertainties.

CHAPTER 2

VISION SYSTEM ARCHITECTURE

In this work, a vision system is developed which reconstructs a 3-D scene from sequences of binocular noisy images. The vision system employs a resolution hierarchy and utilizes dynamic, optimal information fusion. This system has the following unique features:

- Optimal information fusion
- Calculation of 3-D information without solving the correspondence problem pairwise
- Robustness to the presence of noise in the images.
- Computational efficiency
- Reduction of the uncertainties inherent in the calculation of 3-D information
- Compensation of the effects of discrete photoelements in cameras

In Section 2.1, an overall description of the vision system architecture is presented, and the discussion of the hierarchical data structure is given in Section 2.2.

2.1 System Architecture and System Description

The architecture of the vision system which uses a resolution hierarchy is shown in Figure 2.1, where level 1 is the top level and level I is the bottom level. The resolution hierarchy in this work utilizes a pyramid data structure whose image size at at the jth level is half that of the $(j+1)th$ level. The inputs at each level are feature maps (strength and sign maps of zero-crossings). At level j ($j = 1, ..., I$), there are two sequences, left (L_k^j) and right (R_k^j), k=1,2,...,M, of zero-crossing strength maps and two sequences of zero-crossing sign maps. Matching is performed pairwise between each image pair and sequencewise between the first and second images, the first and third images, and so on, in left and right sequences. The term *pair* refers to two images taken by the binocular cameras at the same moment.

Starting with the first image pair at the top level, matching begins with a priori information about minimum distance from the scene to the cameras, finds all the possible matches under a set of specified criteria, and assigns the confidence values to each match based on the results of matching. Each possible match of the first image pair determines a 3-D geometrical structure, which is used to predict regions in the second, third, ..., and Mth image pairs of the top level, wherein matching is to be performed, provided the motion parameters of the cameras are known. The confidence values of the possible matches in the first image pair are updated in the information fusion unit using confidence values obtained by pairwise matching and sequencewise matching at the top

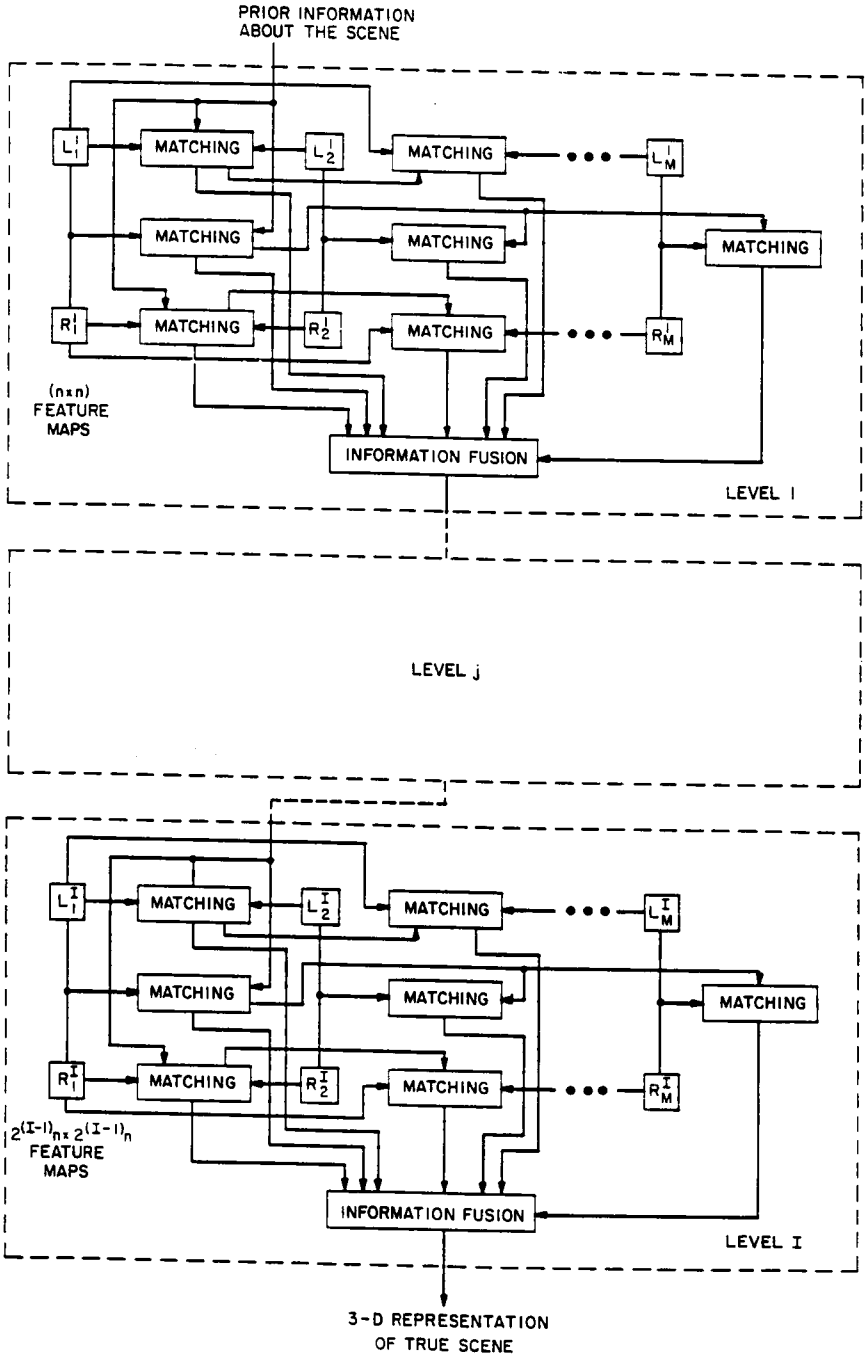


Figure 2.1: The architecture of the vision system.

level. Then, each updated possible match predicts a region in the first image pair at the lower level (second level) of the hierarchy, and the updated confidence value of the possible match at the top level is assigned to the possible matches found in the predicted region at the second level. In general, the number of possible matches increases when going down the hierarchy because the sizes of the images increase from top to bottom. The confidence values at the second level are updated in a similar way in the information fusion unit by the information provided by pairwise and sequencewise matching at this level. This procedure repeats downwards until the bottom level of the hierarchy is reached. At the end of information fusion at the bottom level, the possible matches with high confidence values are considered to be correct matches, and the corresponding 3-D information is calculated using triangulation. However, due to the finite resolution of cameras, matches determine 3-D information with uncertainties. These uncertainties are reduced by fusing the 3-D information provided by the image pairs that were taken at different locations and at different times. The 3-D information fusion is done only at the bottom level.

The hierarchical data structure is discussed in Section 2.2. Feature extraction and assignment of confidence values are detailed in Chapter 3, and the information fusion, including confidence value fusion and 3-D information fusion, is presented in Chapter 5. The models for information fusion are derived using the geometrical modeling introduced in Chapter 4.

2.2 Hierarchical Data Structure

A resolution hierarchy (pyramid structure) is employed by this vision system in order to increase computational efficiency and to reduce noise effects. Since the sizes of the images are smaller at the higher levels of the pyramid structure, the size of the window, wherein the possible matches are found, is accordingly smaller: $W_j = W_I/(2^{I-j})$, where W_j is the window size at the jth level and W_I is the window size at the bottom level (original image resolution level). Therefore, matching at high levels of the hierarchy requires less computation. Once the possible matches are found at higher levels, they are used to guide the matching at lower levels. Thus, computational efficiency is acquired by employing resolution hierarchy. Due to the low-pass filtering in generating the pyramid data structure, fine details are removed from the higher levels. This low-pass filtering is especially welcome when dealing with noisy images. At high levels, since noise is reduced, possible matches can be more accurately found than at the bottom level. Then, at lower levels matching is done in small regions, predicted by the possible matches found at upper levels. Together with information fusion, introduced in Chapter 5, hierarchical structure makes the vision system more robust to noise.

2.2.1 Creation of the Pyramid Structure

In general, an image pyramid is a data structure in which the input image forms the base of the pyramid. Each subsequent level of the pyramid is formed by taking averages over regions of the level below such that each level is of a lower

resolution than its predecessor. There are different ways to create the pyramid structure, the scheme that is used to create the pyramid structure in this work is the Gaussian pyramid, initially proposed by Burt [14]. The Gaussian pyramid is a sequence of images in which each subsequent image is the equivalent of a low-pass filtered image of its predecessor. A Gaussian pyramid is created by using an array I_m of dimension $2^m \times 2^m$, representing the original image, as the base of the pyramid. Each subsequent level of the pyramid, $I_{m-1} \ldots I_0$, is a square array half the dimension of its predecessor. These arrays are lower resolution representations of the original image. The top level, I_0, of the pyramid is a 1×1 array. Each level of the image pyramid, I_{m-1} through I_0, is created from the image at the level below by a half-overlapping averaging method. Figure 2.2 shows the structure of the image pyramid. An element (node) of the array I_l $(l < m)$ is obtained by a weighted average of nodes in I_{l+1} within a 4×4 neighborhood [120], as

$$I_l(i,j) = \sum_{p=-2}^{2} \sum_{q=-2}^{2} w(p,q) I_{l+1}(2i + p - \frac{p}{2|p|} + \frac{1}{2}, 2j + q - \frac{q}{2|q|} + \frac{1}{2}) \quad for\ p,q \neq 0. \quad (2.1)$$

The 4×4 generating kernel, $w(m,n)$, is chosen such that it is

- Separable: $w(p,q) = \hat{w}(p)\hat{w}(q)$
- Normalized: $\sum \hat{w}(p) = 1$
- Symmetric: $\hat{w}(p) = \hat{w}(-p)$.

For $\hat{w}(-2) = \hat{w}(2) = a$, and $\hat{w}(-1) = \hat{w}(1) = b$, the normalization constraint implies $a + b = 0.5$. If the pattern of weights is traced from any level l node, it can be shown [14] that there is always an equivalent weighting function W_l which

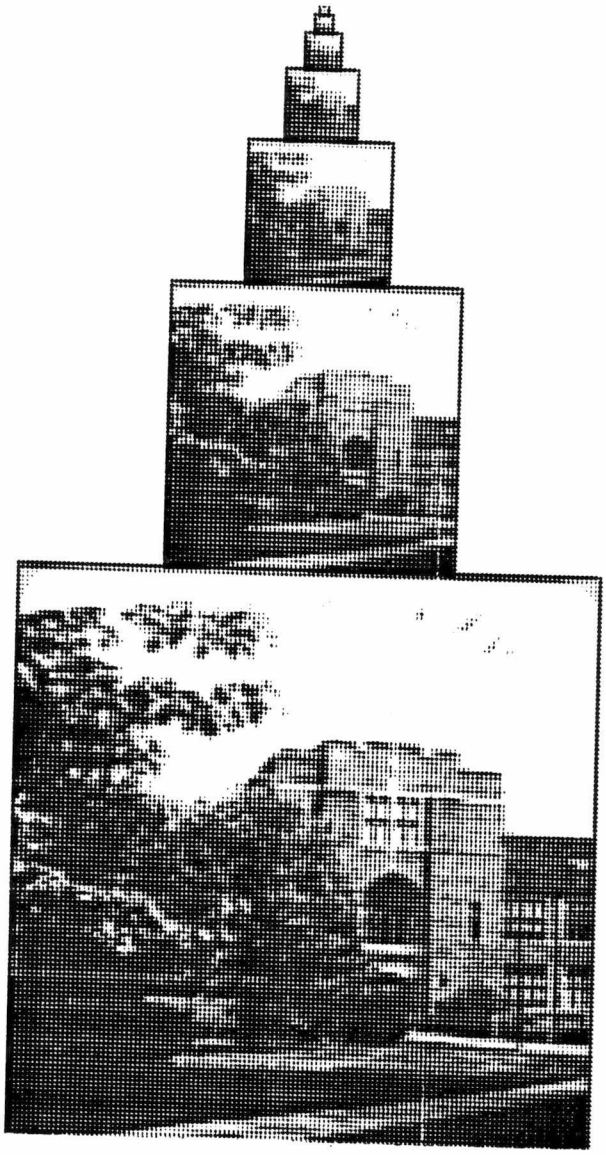


Figure 2.2: An image pyramid.

could have been applied directly to the image I_m to obtain the I_l values. The shape of these equivalent weighting functions is such that if a and b are chosen to have values 0.125 and 0.375, respectively, the weighting functions then closely approximate the Gaussian probability density function. This is the reason that the image pyramid thus created is known as the Gaussian pyramid and may be viewed as a multi-resolution low-pass filter, as shown in [14]. When the input array is a feature map, the pyramid is called the feature pyramid. For the Gaussian image pyramid, each father node at level l ($l < m$) has a 4×4 array of *candidate son* nodes on level $l+1$, as shown in Figure 2.3. The father node is really a weighted average of the sixteen son nodes. Node $I_l(i,j)$, for example, has 16 candidate son nodes, denoted by $I_{l+1}(i',j')$, where

$$\begin{aligned} i' &= 2i-1,\ 2i,\ 2i+1,\ 2i+2, \\ j' &= 2j-1,\ 2j,\ 2j+1,\ 2j+2. \end{aligned} \tag{2.2}$$

Conversely, for each son node at level l below the apex of the pyramid, there exists a 2×2 array of *candidate father nodes* at the level $(l-1)$. Since Equation (2.2) does half-overlapping averaging, four son nodes at level l have the same father nodes at level $(l-1)$ as shown in Figure 2.4. For node $I_l(i,j)$, there are four candidate father nodes denoted by $I_{l-1}(i'',j'')$, where

$$\begin{aligned} i'' &= \Gamma(i/2)-1,\ \Gamma(i/2), \\ j'' &= \Gamma(j/2)-1,\ \Gamma(j/2), \end{aligned} \tag{2.3}$$

where $\Gamma(\cdot)$ is a nearest integer operator. It should be noted that for nodes along the image border the missing relatives are defined by extending the image beyond

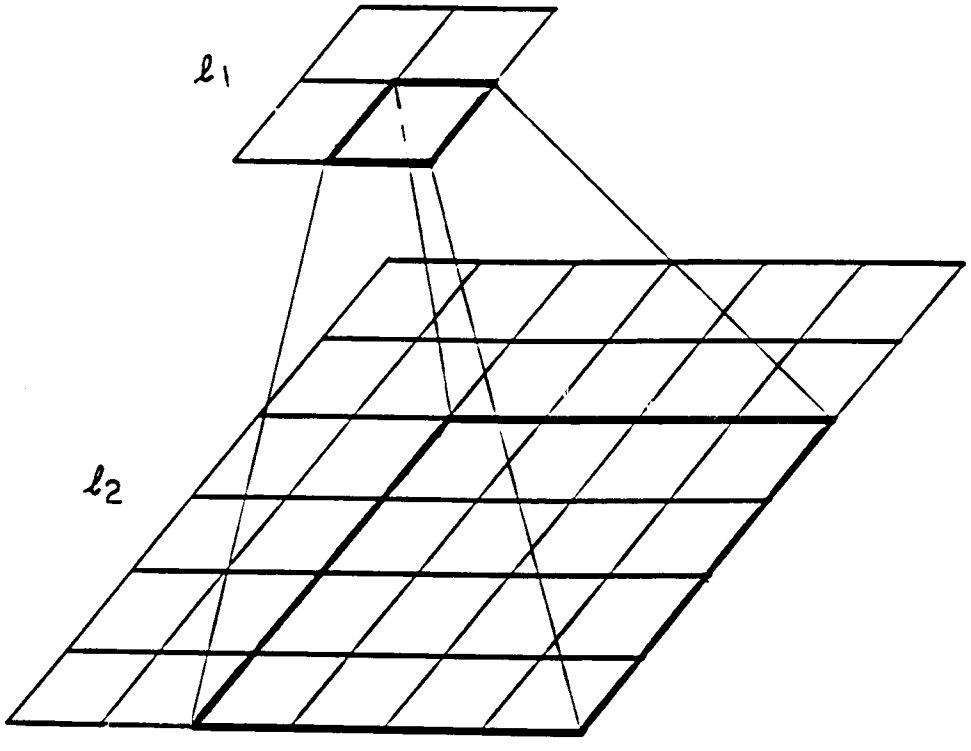


Figure 2.3: Candidate son nodes for a father node. A weighted sum of sixteen son nodes at l_1 creates one father node at l_2.

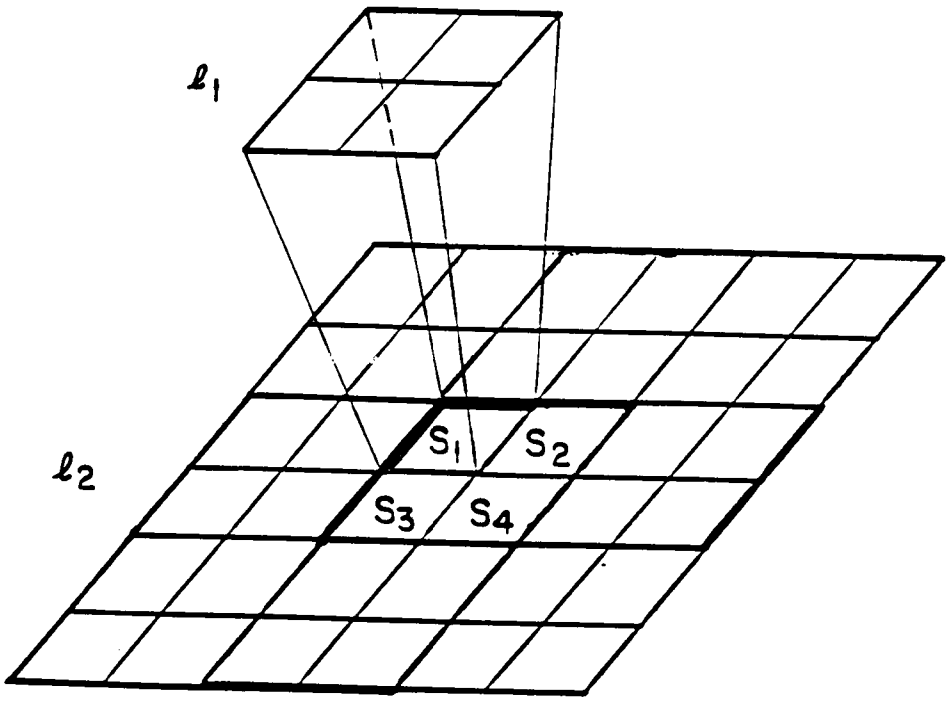


Figure 2.4: Candidate father nodes for a son node. Each son node at l_1 has four father nodes at l_2. Son nodes S_1, S_2, S_3, and S_4 at l_1 have same the father nodes at l_2.

the border. This is done by reflecting the top and bottom rows and the right and left columns of the image outwards.

This vision system utilizes a truncated pyramid whose top level size is $n \times n$ instead of 1×1. The determination of the size of the top level is a tradeoff between retaining fine details of the images and supressing noise – larger top level size (fewer levels in the pyramid) retains more image detail, but is also more sensitive to noise. In this system, the top level with either size 64×64 (3-level pyramid) or size 128×128 (2-level pyramid) is used, depending on the degree of noise in the images.

2.2.2 Top to Bottom Predictions for Matching

As mentioned in Section 2.1, updated possible matches at the upper level predict the regions in the first pair at the current level wherein matching is performed. This prediction is addressed here.

Because of the half overlapping used in the weighted average, Equation (2.1), and because of the shift in the location of zero-crossings, locations of the zero-crossings at the lth level and the locations of the corresponding zero-crossings at the $(l+1)th$ level do not satisfy Equation (2.2). Therefore, an adjustment is necessary when predicting the regions at the lower level. Assuming that a possible match is found at the lth level, with coordinates (i_l, j_l) for the left point of the match and (i_r, j_r) for the right point, the regions in the left and right

images of the first pair at the $(l+1)th$ level are predicted as

$$\begin{aligned} i'_l &= 2i_l - 1 - \Delta i_l,\ ...,\ 2i_l - 1,\ 2i_l,\ 2i_l + 1,\ 2i_l + 2,\ ...,\ 2i_l + 2 + \Delta i_l \\ j'_l &= 2j_l - 1 - \Delta j_l,\ ...,\ 2j_l - 1,\ 2j_l,\ 2j_l + 1,\ 2j_l + 2,\ ...,\ 2j_l + 2 + \Delta j_l, \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} i'_r &= 2i_r - 1 - \Delta i_r,\ ...,\ 2i_r - 1,\ 2i_r,\ 2i_r + 1,\ 2i_r + 2,\ ...,\ 2i_r + 2 + \Delta i_r \\ j'_r &= 2j_r - 1 - \Delta j_r,\ ...,\ 2j_r - 1,\ 2j_r,\ 2j_r + 1,\ 2j_r + 2,\ ...,\ 2j_r + 2 + \Delta j_r, \end{aligned} \tag{2.5}$$

where Δi_l, Δj_l, Δi_r, and Δj_r are the increments of the regions at the $(l+1)th$ level that incorporate the shifts of the zero-crossings. The matches found in the regions specified by Equations (2.4) and (2.5) at the $(l+1)th$ level then predict the regions of the $(l+2)th$ level using the same adjustments as those in Equations (2.4) and (2.5). In this way, the predictions of the regions in the first image pairs, from top to bottom, are corrected level by level to compensate for the shifts of zero-crossings.

CHAPTER 3

MATCHING AND ASSIGNMENT OF CONFIDENCE VALUES

Given a pair of images, the corresponding 3-D scene can be reconstructed if the correspondences between two images taken at different locations are known. However, finding correspondences has been proved to be very difficult, especially for noisy images. To solve this problem, the concept of confidence value is introduced in this work. Unlike stereo vision, where the establishment of the correspondences is attempted in each image pair, this vision system uses binocular matching only to find all the possible matches and assigns each possible match a corresponding confidence value. Because of the noise in the images, the possible matches with high confidence values from one image pair are not necessarily correct matches; but they have better chances to be correct matches. Similarly, the possible matches with low confidence values from one image pair are not necessarily false matches. Since the information about the scene contained in the image sequences is in part redundant, and the noise is independent over the sequences, the information about the scene is retained and the effect of noise is cancelled, when fusing the information provided by the image sequences and image pairs together. Not only is the effect of noise in the images cancelled, but

the effects of other factors which cause false matching in stereo vision, such as change of illumination, are effectively reduced.

The confidence values of possible matches are updated by the information about the scene contained in binocular image sequences. In the process of information fusion, the confidence values of the correct matches are increased[1], while the confidence values of the false matches are decreased. After fusion, the matches with high confidence values are considered correct matches. Therefore, finding correct matches is much more robust in the sense that fewer correct matches are discarded and fewer false matches are accepted, on the average.

In this chapter, finding possible matches and assigning confidence values are discussed. Information fusion is the subject of Chapter 5. In Section 3.1, features used for matching are discussed, and extraction of features from images is studied. The criteria for matching and assigning confidence values are the main topics of Section 3.2.

[1]The word *increased* or *decreased* is understood in the following way: If the confidence values for the correct matches are high, they are unlikely to be decreased in the fusion process, while if the confidence values for the correct matches are low (this is very likely in the case of noisy images), they are likely to be increased by fusing the information from sequences. For false matches, the situation is the same: the confidence values of false matches are not likely to be increased by information fusion.

3.1 Feature Extraction

In this work, the features chosen for matching are the Marr-Hildreth zero-crossings [85], because of their ability to compromise between accurately localizing significant changes in the intensities and effectively smoothing the noise in the images. For each level of the resolution hierarchy, the zero-crossings are extracted by convolving the images $I(x, y)$ with the Laplacian of Gaussian

$$\nabla^2 G(x,y) = \frac{1}{\sqrt{2\pi}\sigma^3}(\frac{x^2+y^2}{\sigma^2} - 2)e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{3.1}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{3.2}$$

and

$$G(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x^2+y^2)/2\sigma^2}, \tag{3.3}$$

and by finding the positions where the function changes the sign.

In this work, the value of σ is constant for all levels of the hierarchy and is based on the degree of noise in the images. Besides locations of zero-crossings, strengths and signs of zero-crossings are also obtained. The strength of a zero-crossing is determined by summing the absolute values of $\nabla^2 G(x, y) * I(x, y)$ on both sides of the zero-crossing. The sign of a zero-crossing is assigned based on whether the intensity changes from low to high or from high to low across the zero-crossing. An example of extracting zero-crossings is shown in Figures 3.1 - 3.3, where Figure 3.1 is the original image, Figure 3.2 shows the strength map of the zero-crossings, and Figure 3.3 shows the sign map of the zero-crossings

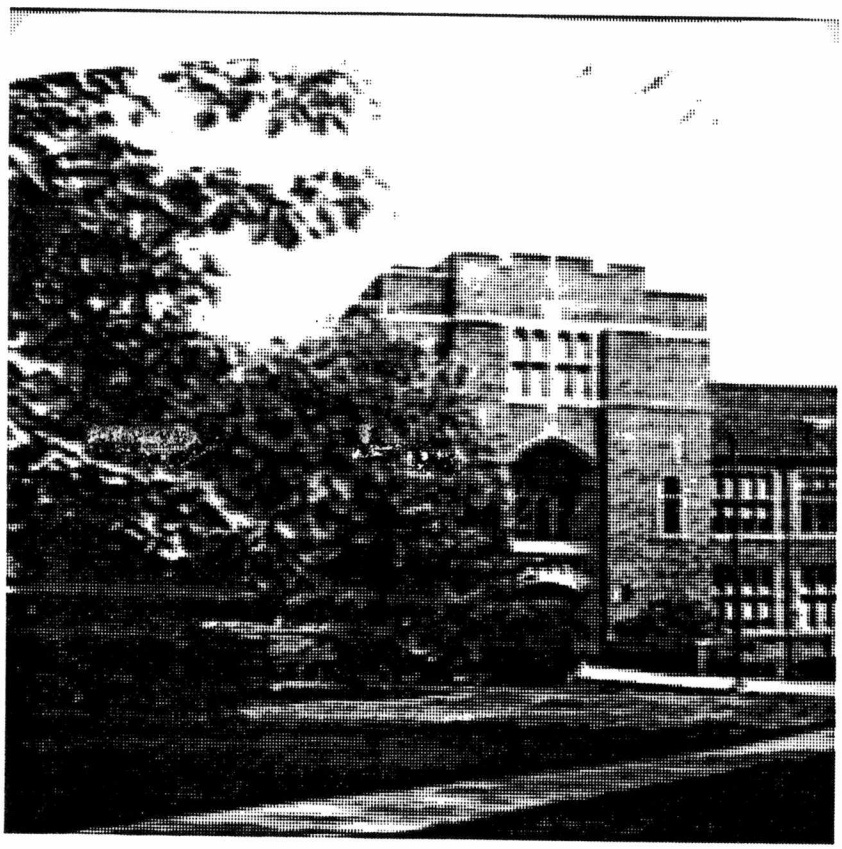


Figure 3.1: An input image.

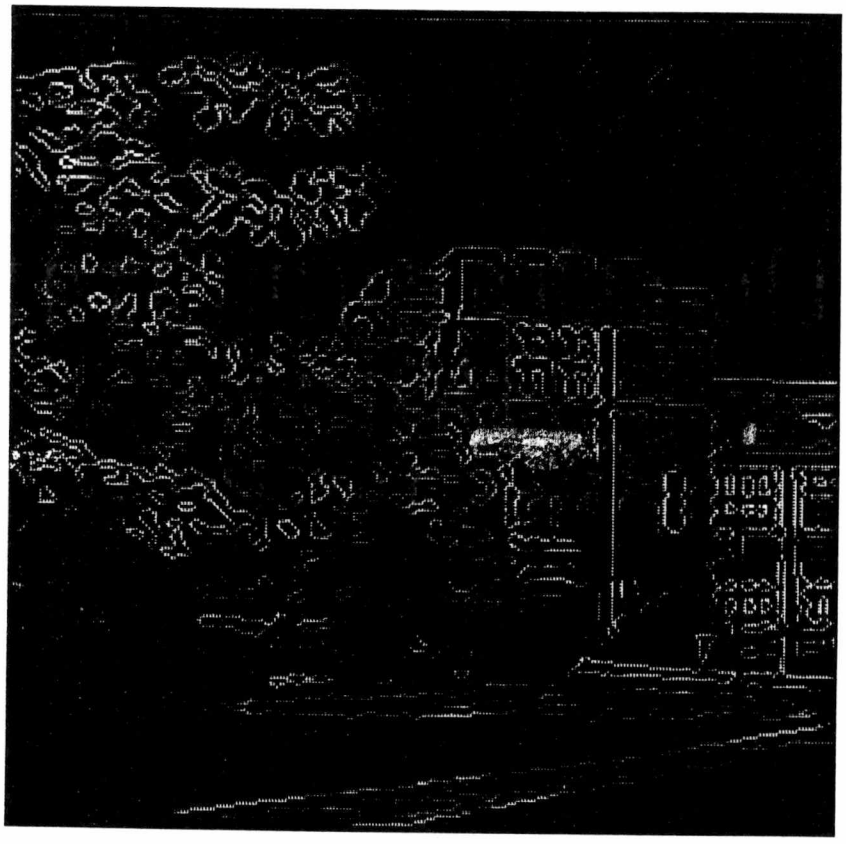


Figure 3.2: The strength map of the zero-crossings of the input image.

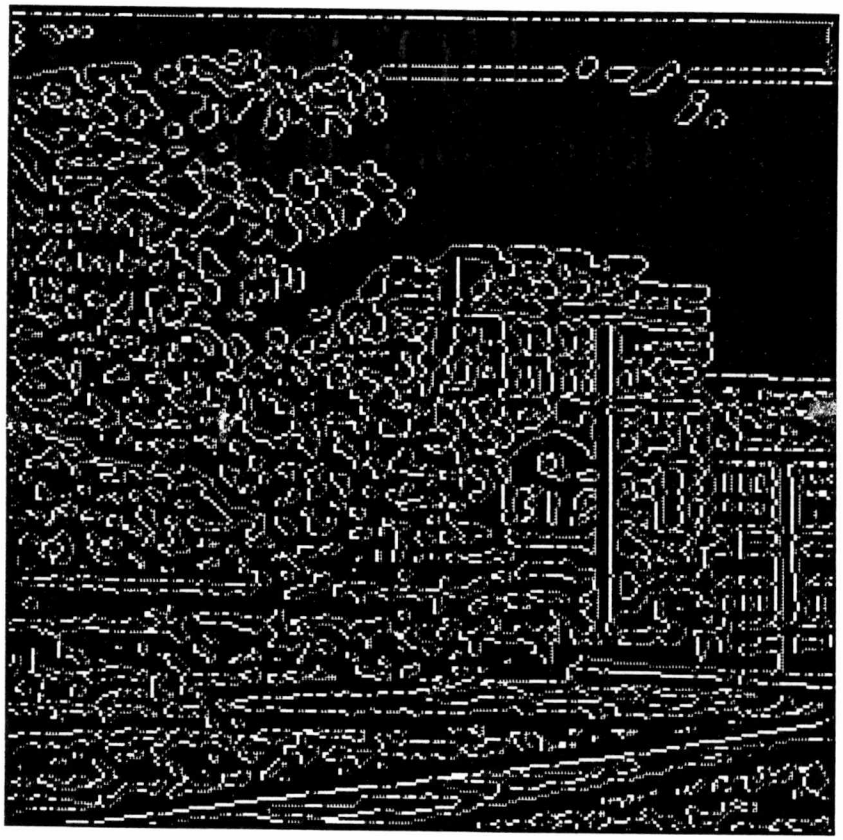


Figure 3.3: The sign map of the zero-crossings of the input image.

where the positive zero-crossings are in white and the negative zero-crossings are in gray.

It has been pointed out by some researchers that zero-crossing edge detection algorithms can detect locations, so-called phantom or spurious edges, which do not correspond to significant intensity changes in an image and cause various problems in matching [25]. However, in this vision system, since final matching results are acquired by fusing large amount of information from the sequences, a few phantom edges do not affect the final results.

3.2 Matching and Assignment of Confidence Values

3.2.1 Pairwise and Sequencewise Matching

In this vision system, the objective of matching is to find all the possible matches under specified criteria and to assign confidence values to each possible match. Matching is carried out pairwise and sequencewise, Figure 2.1. In this work, the term *pairwise matching* means that the matching is performed between left and right images, and the term *sequencewise matching* indicates that the matching is carried out between images in the left sequence or the right sequence. Since the matching in the second and later image pairs is used to update the matching in the first image pair, sequencewise matching is performed between the first and second pairs, the first and third pairs, and the first and fourth pairs, etc. The incorporation of sequencewise matching reduces the number of possible false matches in pairwise matching.

For the pairwise matching, in the *first* image pair at the *top* level of the resolution hierarchy, for a given pixel $P_l(i,j)$ in the left image, all the possible matches within a specified window W on the epipolar line e_r [52,53,72], Figure 3.4, in the right image are found and assigned corresponding confidence values. For the cases where the optical axes of cameras are parallel, Figure 3.4, as in the case of this vision system, the epipolar lines e_l and e_r are parallel to each other and to the X axis of reference coordinate system XYZ. For general cases where the optical axes are not parallel to each other, the epipolar lines can be calculated by camera calibration [36]. The size of the window W is determined by the estimated minimum distance between the cameras and the scene [72]:

$$W = P_l^{'} - P_{r|_{Z_{min}}}, \tag{3.4}$$

where P_l and P_r are projections of the 3-D point $P(x_p, y_p)$ on the left and right image planes. The point $P_l^{'}$ is the transferred coordinate of P_l, $P_l^{'} = P_l$, Figure 3.4, and Z_{min} is the estimated minimum distance. The values of $P_l^{'}$ and P_r are determined by

$$P_l^{'} = \frac{(x_p + b/2)f}{Z_{min}2^{I-1}}, \tag{3.5}$$

and

$$P_r = \frac{(x_p - b/2)f}{Z_{min}2^{I-1}}, \tag{3.6}$$

where I is the number of levels in the hierarchy. The coefficients b and f are the length of baseline and the length of focus. Therefore, the size of the window is given by

$$W = \frac{bf}{Z_{min}2^{I-1}}. \tag{3.7}$$

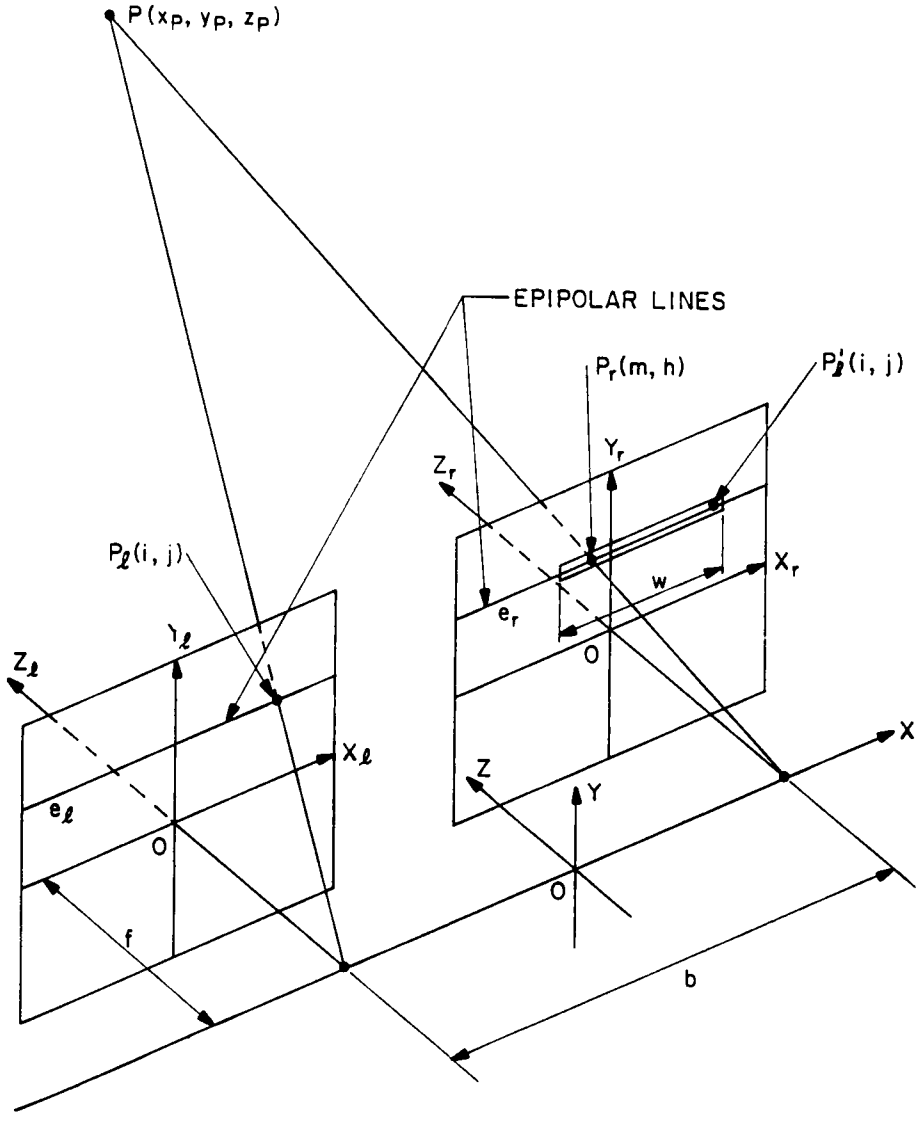


Figure 3.4: Calculation of the window size for matching.

In the *second* and *later* image pairs and in the image pairs at *lower* levels, matching is done in the predicted regions. The prediction of regions for matching in the second and later image pairs is described in Chapter 4, and the prediction of regions in the first image pairs in the lower levels is discussed in Chapter 2.

In the following, the procedure for pairwise matching is described. For sequencewise matching, one needs only to change the variables in the left image to the variables in the first image and change the variables in the right image to the variables in the *kth* image. Sequencewise matching is done in both left and right sequences.

Feature point matching, for pairwise matching, is based on:

1. Signs of the zero-crossings

   For two feature points to be possible matches, the signs of the zero-crossings at these two points should be the same (both positive or both negative).

2. Strengths of the zero-crossings

   For two feature points to be possible matches, the absolute value of the difference of two zero-crossing strengths should also be less than a threshold:

$$Diff_{S_{zc}}(i,j) = |S_{zc_l}(i,j) - S_{zc_r}(i,j')| \leq T_{S_{zc}}, \tag{3.8}$$

   where $T_{S_{zc}}$ is the given threshold for the strength difference, and S_{zc_l} and S_{zc_r} stand for the left and right strength maps of zero-crossings. In the first image pair at the top level,

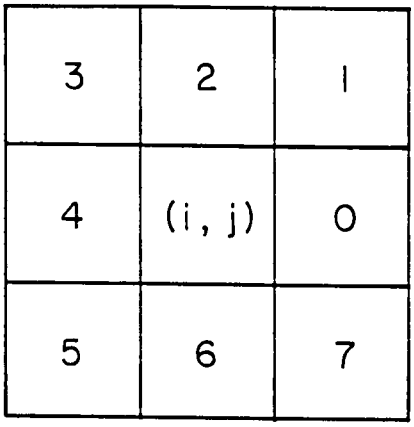
$$j - W \leq j' \leq j. \tag{3.9}$$

Otherwise, j' lies in the predicted regions.

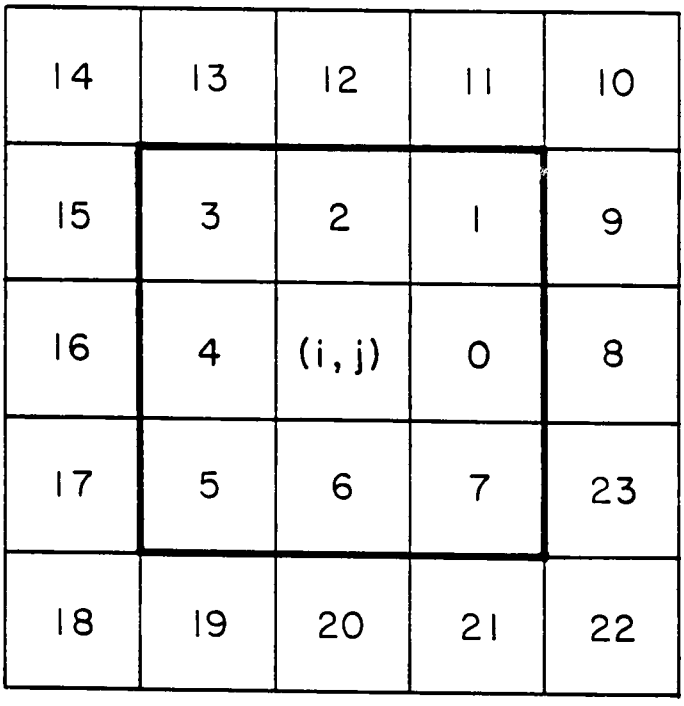
3. Patterns of the zero-crossings

For two points to be possible matches, the patterns of the zero-crossings in the neighborhoods of these two points should be similar. The similarity of the zero-crossing patterns is measured by the difference of their pattern numbers which are shown in Figure 3.5. The pattern numbers were originally proposed in [72], in which only 8 neighborhood was studied and only two directional components of zero-crossings in the 8 neighborhood were used. In this work, we propose two sets of pattern numbers for two kinds of neighborhoods, 8 neighborhood and 24 neighborhood, and study multi-directional components of zero-crossings in the neighbohoods. For the 8 neighborhood, the pattern numbers shown in Figure 3.5 (a) are used. The pattern numbers shown in Figure 3.5 (b) are used for the 24 neighborhood, where the neighborhood is divided into two subneighborhoods: subneighborhood 1, which is the same as the 8 neighborhood; and subneighborhood 2, which is the band in Figure 3.5 (b) numbered from 8 to 23.

For the 8 neighborhood, assume that the number of pixels occupied by zero-crossings in the neighborhood of (i, j) is NP_l, and the number of pixels occupied by zero-crossings in the neighborhood of (i, j') is NP_r. The pattern difference, $Diff^i_{P_{zc}}$, for each zero-crossing point is calculated as



(a)



(b)

Figure 3.5: Pattern numbers for 8 and 24 neighborhoods: (a) 8 neighborhood, and (b) 24 neighborhood.

follows:

$$\text{when } NP_l > NP_r : \begin{cases} Diff^i_{P_{zc8}} = |P_{l_i} - P_{r_k}|, & i = 1, ..., NP \\ Diff^i_{P_{zc8}} = |8 - Diff^i_{P_{zc}}| \text{ if } Diff^i_{P_{zc}} > 4, \end{cases} \tag{3.10}$$

$$\text{when } NP_l < NP_r : \begin{cases} Diff^k_{P_{zc8}} = |P_{r_k} - P_{l_i}|, & k = 1, ..., NP \\ Diff^k_{P_{zc8}} = |8 - Diff^k_{P_{zc}}| \text{ if } Diff^i_{P_{zc}} > 4, \end{cases} \tag{3.11}$$

where P_{r_k} is the nearest pixel of P_{l_i} occupied by the zero-crossing counted counterclockwise, and $NP = max(NP_l, NP_r)$. For instance, in Figure 3.6 (a), $NP = 3$, and the three pattern number differences are

$$Diff^1_{P_{zc8}} = |P_{l_1}(i-1, j+1) - P_{r_1}(i-1, j')| = |1-2| = 1,$$

$$Diff^2_{P_{zc8}} = |P_{l_2}(i-1, j) - P_{r_1}(i-1, j')| = |2-2| = 0,$$

and

$$Diff^3_{P_{zc8}} = |P_{l_3}(i+1, j) - P_{r_2}(i+1, j'|) = |6-6| = 0.$$

In Figure 3.6 (b), $NP = 4$, and the four pattern number differences are

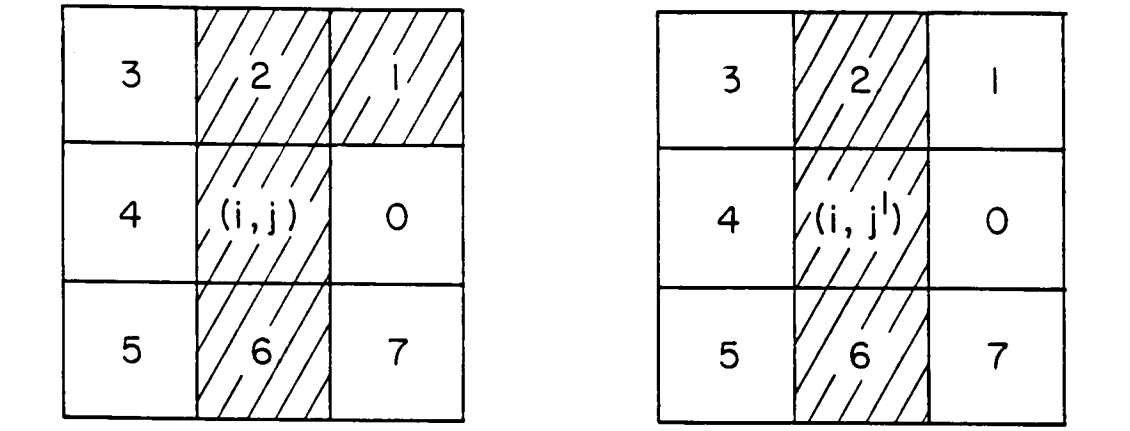
$$Diff^1_{P_{zc8}} = |P_{r_1}(i-1, j'+1) - P_{l_1}(i-1, j+1)| = |1-1| = 0,$$

$$Diff^2_{P_{zc8}} = |P_{r_2}(i-1, j') - P_{l_1}(i-1, j+1)| = |2-1| = 1,$$

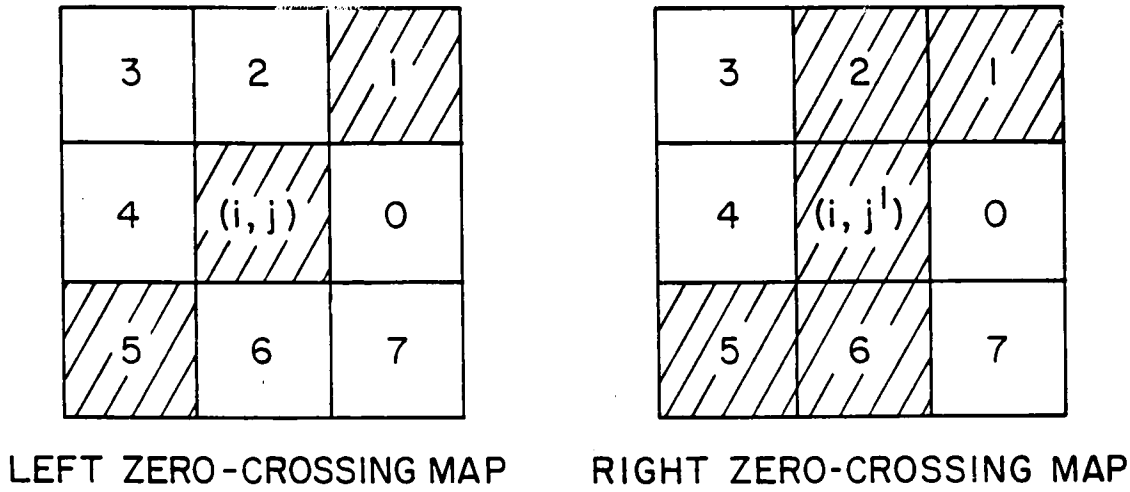
$$Diff^3_{P_{zc8}} = |P_{r_3}(i+1, j'-1) - P_{l_2}(i+1, j-1)| = |5-5| = 0,$$

and

$$Diff^4_{P_{zc8}} = |P_{r_4}(i+1, j') - P_{l_2}(i+1, j-1)| = |6-5| = 1.$$



(a)



(b)

Figure 3.6: Two examples for calculating pattern differences.

The difference of pattern numbers for points (i, j) and (i, j') is then defined by

$$Diff_{P_{zc8}} = \sum_{i=1}^{NP} Diff^{i}_{P_{zc8}}. \tag{3.12}$$

For the example in Figure 3.6 (a), $Diff_{P_{zc8}} = 1 + 0 + 0 = 1$, and in Figure 3.6 (b), $Diff_{P_{zc8}} = 0 + 1 + 0 + 1 = 2$.

For the 24 neighborhood, the pattern differences in two subneighborhoods are calculated separately and added together. The calculation of the pattern difference for the subneighborhood 1 is the same as that in the 8 neighborhood

$$Diff1_{P_{zc24}} = Diff_{P_{zc8}}. \tag{3.13}$$

The pattern differences for the pixels in subneighborhood 2 are given as follows:

$$\text{when } NP2_l > NP2_r : \begin{cases} Diff2^{i}_{P_{zc24}} = |P2_{l_i} - P2_{r_k}|, & i = 1, ..., NP2 \\ Diff2^{i}_{P_{zc24}} = |16 - Diff2^{i}_{P_{zc24}}| & \text{if } Diff2^{i}_{P_{zc24}} > 8, \end{cases} \tag{3.14}$$

$$\text{when } NP2_l < NP2_r : \begin{cases} Diff2^{k}_{P_{zc24}} = |P2_{r_k} - P2_{l_i}|, & k = 1, ..., NP2 \\ Diff2^{k}_{P_{zc24}} = |16 - Diff2^{k}_{P_{zc24}}| & \text{if } Diff2^{k}_{P_{zc24}} > 8, \end{cases} \tag{3.15}$$

where $NP2_l$ and $NP2_r$ are the numbers of pixels occupied by the zero-crossings in subneighborhood 2 and $NP2 = max(NP2_l, NP2_r)$. The pattern difference in subneighborhood 2 is then:

$$Diff2_{P_{zc24}} = \sum_{i=1}^{NP2} Diff2^{i}_{P_{zc24}}. \tag{3.16}$$

The pattern difference for the 24 neighborhood is obtained by adding the pattern differences in two subneighborhoods:

$$Diff_{P_{zc24}} = Diff1_{P_{zc24}} + Diff2_{P_{zc24}}. \tag{3.17}$$

Two zero-crossing patterns are said to be similar if

$$Diff_{P_{zc8}} \leq T^S_{P_{zc}}, \quad \text{for 8 neighborhood,} \tag{3.18}$$

or

$$Diff_{P_{zc24}} \leq T^P_{P_{zc}}, \quad \text{for 24 neighborhood,} \tag{3.19}$$

where $T^S_{P_{zc}}$ is the given threshold for sequencewise pattern matching and $T^P_{P_{zc}}$ is the threshold for pairwise pattern matching. The 8 neighborhood matching is used in sequencewise matching, and the 24 neighborhood matching is used in pairwise matching. Therefore, pairwise matching is more rigorous than sequencewise matching in this system.

The zero-crossings which are aligned with the epipolar lines are not matched since they have inherent ambiguities. In this vision system, horizontal zero-crossings with lengths greater than or equal to 3 pixels are excluded from matching. The matching of zero-crossing patterns actually implements the constraint of figural continuity [89].

In order to reduce computation, these three matching criteria are used in a sequential manner: if the first criterion (signs of zero-crossings) is met, then proceed to the second criterion (difference of strengths), otherwise, terminate the

matching; if the second criterion is met, continue to the third criterion (pattern differences), otherwise, terminate the matching.

The procedure for sequencewise matching is similar to that of pairwise matching, except that the matching is performed between the first image and the kth image in the left or right sequence.

3.2.2 Assignment of Confidence Values in the Top First Image Pair

In the top first image pair, two feature points are possible matches if their signs are the same, their zero-crossing strength difference is less than $T^P_{S_{zc}}$, and their zero-crossing pattern difference is less than $T^P_{P_{zc}}$ (the superscript P denotes pairwise matching). Each possible match[2] is assigned a confidence value:

$$Cf(i,j) = \frac{\alpha_1}{1 + Diff_{S_{zc}}(i,j)} + \frac{\alpha_2}{1 + Diff_{P_{zc}}(i,j)} + \frac{\alpha_3}{1 + Nct(i,j)} \tag{3.20}$$

where $Diff_{S_{zc}}$ is the difference of the strengths of zero-crossings, $Diff_{P_{zc}}$ is the difference of the patterns of the zero-crossings, and $Nct(i,j)$ is the number of possible matches found in the window W. The coefficients α_1, α_2 and α_3, which satisfy

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \tag{3.21}$$

normalize the range of the confidence value $Cf(i,j)$ to $0 - 1$, and specify the weight distribution among the three terms in Equation (3.19).

[2]Exactly, the confidence value is assigned to the left feature point (i,j) of the match.

3.2.3 Assignment of Confidence Values in Other Image Pairs

There are two cases regarding the assignment of confidence values, and they will be discussed separately: the assignment of confidence values in the first image pair at the lower levels, and the assignment of confidence values in the second, third, ..., image pairs at all levels (including the top level.)

Assignment of confidence values in the first image pairs at lower levels

In this case, matching is done only pairwise in the regions predicted by the upper level. Two feature points are possibly matched if their signs are the same, if the difference of the strengths is less than $T^P_{S_{zc}}$, and if the difference of the zero-crossing patterns is less $T^P_{P_{zc}}$. The confidence values to be assigned are determined by

$$Cf(i,j) = \frac{\gamma_1}{1 + Diff_{S_{zc}}(i,j)} + \frac{\gamma_2}{1 + Diff_{P_{zc}}(i,j)}, \tag{3.22}$$

where

$$\gamma_1 + \gamma_2 = 1. \tag{3.23}$$

Since matching is performed in predicted regions and the sizes of the regions are much smaller than the size of the matching window used in the first top pair (this is one of the advantages of using hierarchical structure), the number of matches found in the regions plays no important role in determining the confidence values.

Assignment of confidence values in the second and later image pairs at all levels

Matching in the second and later image pairs is performed in the regions predicted by the possible match in the first image pair (Chapter 5 details the prediction). Sequencewise matching goes into effect in determining the possible matches in this case. Two feature points in the kth image pair are said to be possible matches:

- If they have the same signs of zero-crossings, and these signs are the same as the signs of the possible match found in the first image pair.

- If the differences of the strengths of the zero-crossings pairwise and sequencewise satisfy the following:

$$(Diff_{S_{zc}})|_k \leq T^P_{S_{zc}}, \tag{3.24}$$

$$(Diff_{S_{zc}})|^l_{1-k} \leq T^S_{S_{zc}}, \tag{3.25}$$

and

$$(Diff_{S_{zc}})|^r_{1-k} \leq T^S_{S_{zc}}. \tag{3.26}$$

In this instance, $T^S_{S_{zc}}$ is the sequencewise threshold which is greater than or equal to the pairwise threshold $T^P_{S_{zc}}$; $(Diff_{S_{zc}})|_k$ is the difference of zero-crossing strengths of two feature points in the kth image pair; $(Diff_{S_{zc}})|^l_{1-k}$

is the difference of zero-crossing strengths between the left point of the possible match in the first image pair and left feature point in the kth image pair; and $(Diff_{S_{zc}})|_{1-k}^{r}$ is the difference of zero-crossing strengths between the right point of the possible match in the first image pair and right feature point in the kth image pair.

- If the differences of the zero-crossing patterns, pairwise and sequencewise, also satisfy the following:

$$(Diff_{P_{zc24}})|_k \leq T_{P_{zc}}^{P}, \tag{3.27}$$

$$(Diff_{P_{zc8}})|_{1-k}^{l} \leq T_{P_{zc}}^{S}, \tag{3.28}$$

and

$$(Diff_{P_{zc8}})|_{1-k}^{r} \leq T_{P_{zc}}^{S}, \tag{3.29}$$

where $T_{P_{zc}}^{S}$, $T_{P_{zc}}^{S} > T_{P_{zc}}^{P}$, is the sequencewise threshold for pattern differences; $(Diff_{P_{zc24}})|_k$ is the difference of zero-crossing patterns of two feature points in the kth image pair using 24 neighborhood matching; $(Diff_{P_{zc8}})|_{1-k}^{l}$ is the difference of zero-crossing patterns between the left point of the possible match in the first image pair and left feature point in the kth image pair using 8 neighborhood matching; and $(Diff_{P_{zc8}})|_{1-k}^{r}$ is the difference of zero-crossing patterns between the right point of the possible match in the first image pair and right feature point in the kth image pair also using 8 neighborhood matching.

From the above, it can be seen that sequencewise matching is less rigorous than pairwise matching because images change more sequencewise than pairwise due to the motion of the cameras. Once two feature points are found to be possible matches, the confidence value is calculated based only on the pairwise matching in the kth image pair:

$$Cf(i,j) = \frac{\gamma_1}{1 + (Diff_{S_{zc}}(i,j))|_k} + \frac{\gamma_2}{1 + (Diff_{P_{zc}}(i,j))|_k}, \tag{3.30}$$

where

$$\gamma_1 + \gamma_2 = 1. \tag{3.31}$$

3.2.4 Refinement of Confidence Values

Usually, a zero-crossing belongs to a connected group of zero-crossings. Consequently, given a zero-crossing image, matching between zero-crossings becomes matching between pixel groups. Therefore, the confidence values assigned to the possible matches should reflect the matching of pixels in their neighborhood. The final confidence value of a possible match is the weighted average of the confidence values in the neighborhood of the zero-crossing. For instance, in Figure 3.7 (a), the final confidence value of a possible match at the pixel $P(i,j)$ in the left image (for the pairwise matching) is averaged with weighted confidence values assigned to the points lying in the range of Δl pixels on both sides of $P(i,j)$, Figure 3.7 (a). The weights for averaging are described by

$$w^k = \frac{1}{(\Delta l + 1)^2}(\Delta l + 1 - sgn(k)k), \tag{3.32}$$

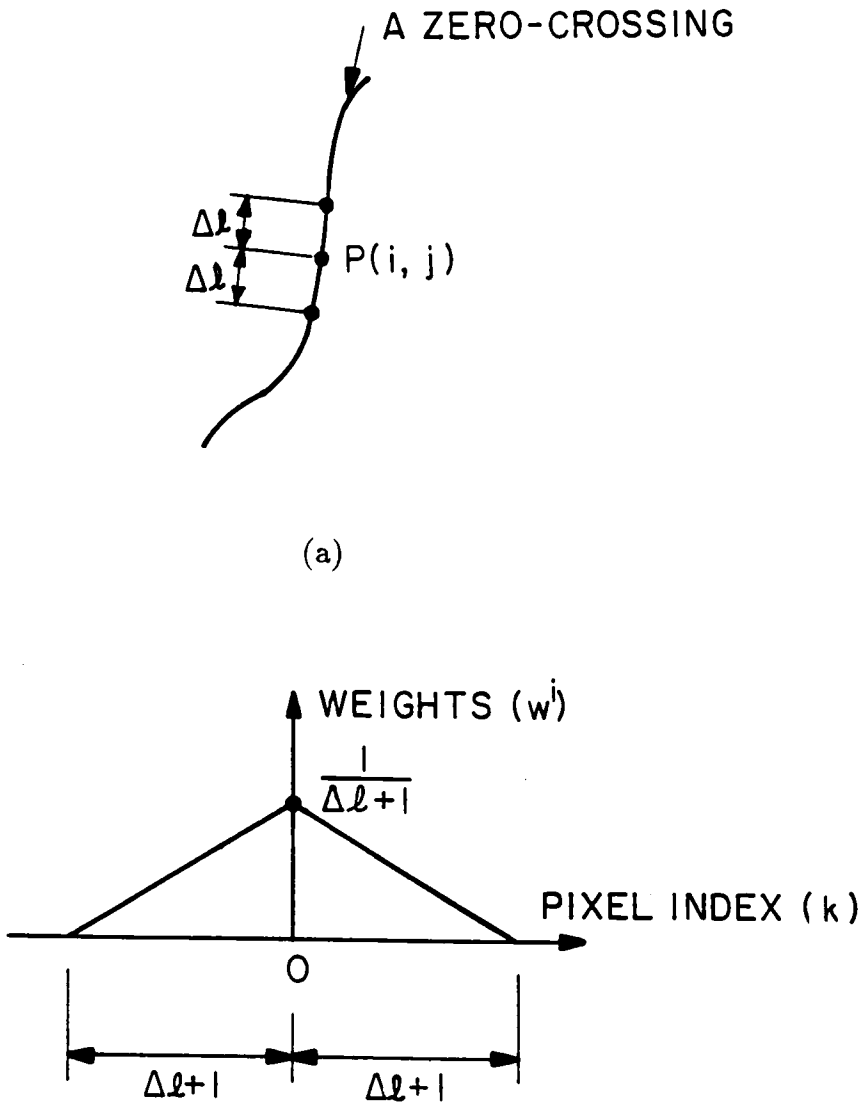


Figure 3.7: Weighted averaging of the confidence values: (a) a specified range in the neighborhood of $P(i,j)$, and (b) a distribution of weights.

where $sgn(k)$ is

$$sgn(k) = \begin{cases} 1, & \text{when } k \geq 0, \\ -1, & \text{when } k < 0. \end{cases} \tag{3.33}$$

The weight distribution, shown in Figure 3.7 (b), gives more weight to the points closer to the point $Q(i,j)$. This reflects the Markov effect: a pixel in an image (a random field) can only be affected by the finite number of pixels in its neighborhood. The final confidence value is then

$$Conf(i,j) = \sum_{k=-\Delta l}^{\Delta l} w^k Cf^k(i'',j''), \tag{3.34}$$

where $k, -\Delta l \leq k \leq \Delta l$ is the index of the pixels in the range of the neighborhood, w^k are the weights given by Equation (3.32), satisfying

$$\sum_{k=-\Delta l}^{\Delta l} w^k = 1, \tag{3.35}$$

and $Cf^k(i'',j'')$ are the confidence values assigned to the pixels in the neighborhood of $P(i,j)$ within the given range.

The confidence values $Conf(i,j)$ of the possible matches are updated by the information provided by the binocular image sequences, which is detailed in Chapter 5.

CHAPTER 4

GEOMETRICAL MODELING

This chapter describes geometrical modeling for a pair of images and for sequences of images. Geometrical modeling reveals the inherent information relationship between image pairs in the sequences and enables us to effectively use the information provided by the sequences to update the confidence values of possible matches and to reduce the uncertainties in calculating 3-D information. In Section 4.1, the geometrical modeling for a single pair of images is presented, and the geometrical modeling for sequences of images is introduced in Section 4.2.

4.1 Geometrical Modeling for a Pair of Images

The study of the geometrical modeling for a pair of images is divided into two parts: deriving a 3-D uncertainty volume from a possible match, Section 4.1.1; and calculating the Gaussian parameters of the probabilistic representation of the uncertainty volume, Section 4.1.2.

4.1.1 A 3-D Uncertainty Volume from a Possible Match

In practice, cameras have only finite resolution. Thus, an inevitable error occurs when 3-D information is obtained from a possible match because of the spatial quantization effect in forming the images [117,108]. For instance, because of the discrete nature of pixels, a possible match, (x_l, y_l) and (x_r, y_r) in Figure 4.1, actually determines a 3-D volume $ABCDA'B'C'D'$, called 3-D uncertainty volume, instead of a 3-D point. In other words, a pixel on the image plane is a small area, instead of a point. For instance, the pixels, (x_l, y_l) and (x_r, y_r), on the left and right image planes of Figure 4.1, are actually small areas defined by

$$x_l - 0.5 \leq x_l^{'} \leq x_l + 0.5 \ , \quad y_l - 0.5 \leq y_l^{'} \leq y_l + 0.5$$

and

$$x_r - 0.5 \leq x_r^{'} \leq x_r + 0.5 \ , \quad y_r - 0.5 \leq y_r^{'} \leq y_r + 0.5 \ .$$

Therefore, hereafter, a pixel stands for a small area. The error obviously occurs when 3-D information is calculated from the match, (x_l, y_l) and (x_r, y_r), since the true 3-D point can be anywhere in the 3-D uncertainty volume.

To compensate for the error and to make use of the information provided by the sequences, one needs to effectively describe the 3-D uncertainty volume, when given a possible match in a image pair. Since the 3-D uncertainty volumes are irregular in shape, to effectively describe them, a probabilistic description is proposed here. The weight distribution of the 3-D points with same brightness in the uncertainty volume in forming the intensities of pixels (x_l, y_l) and (x_r, y_r) is assumed Gaussian. Consequently, the distribution of weights of the 3-D points,

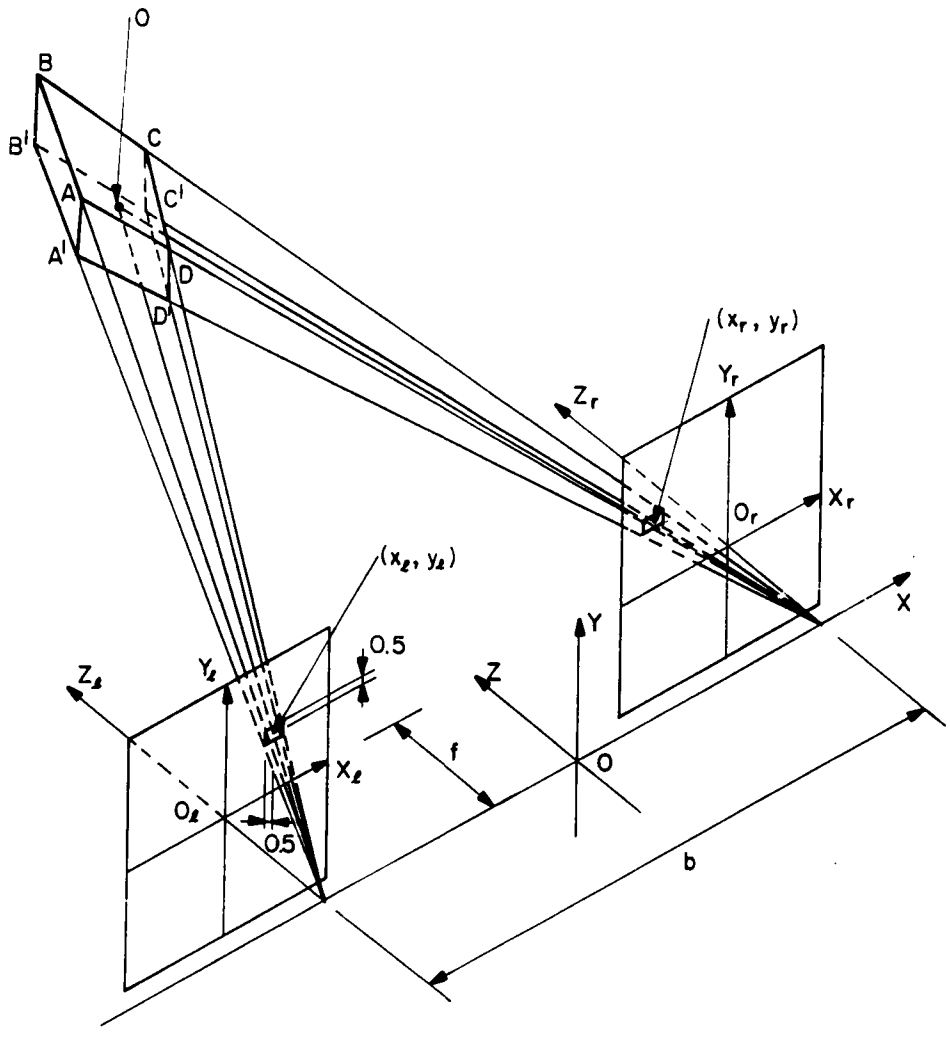


Figure 4.1: A 3-D uncertainty volume in the geometrical modeling for a pair of images.

in pixel intensity formation, over the uncertainty region is characterized by a Gaussian probability distribution [115,116],

$$p(x,y,z) = \frac{1}{(2\pi)^{3/2}\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}(\underline{X}-\underline{m})\mathbf{C}^{-1}(\underline{X}-\underline{m})^T}, \tag{4.1}$$

where

$$\underline{X} = [x,y,z]^T$$

is a variable vector,

$$\underline{m} = [m_x, m_y, m_z]^T \tag{4.2}$$

$$= [E\{x\}, E\{y\}, E\{z\}]^T \tag{4.3}$$

is a mean vector, and

$$\mathbf{C} = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y & \rho_{xz}\sigma_x\sigma_z \\ \rho_{xy}\sigma_y\sigma_x & \sigma_y^2 & \rho_{yz}\sigma_y\sigma_z \\ \rho_{xz}\sigma_z\sigma_x & \rho_{yz}\sigma_z\sigma_y & \sigma_z^2 \end{bmatrix} \tag{4.4}$$

$$= \begin{bmatrix} E\{(x-m_x)^2\} & E\{(x-m_x)(y-m_y)\} & E\{(x-m_x)(z-m_z)\} \\ E\{(y-m_y)(x-m_x)\} & E\{(y-m_y)^2\} & E\{(y-m_y)(z-m_z)\} \\ E\{(z-m_z)(x-m_x)\} & E\{(z-m_z)(y-m_y)\} & E\{(z-m_z)^2\} \end{bmatrix} \tag{4.5}$$

is a covariance matrix. The equal probability points of $P(x,y,z)$ in Equation (4.1) constitute a surface which is an ellipsoid, as shown in Figure 4.2. Therefore, the uncertainty volume $ABCDA'B'C'D'$ is completely described by Gaussian parameters: $\underline{m}$ and $\mathbf{C}$. With the aid of this probability description of the 3-D uncertainty volume, the true 3-D point specified by the possible match (x_l, y_l) and (x_r, y_r) is given by

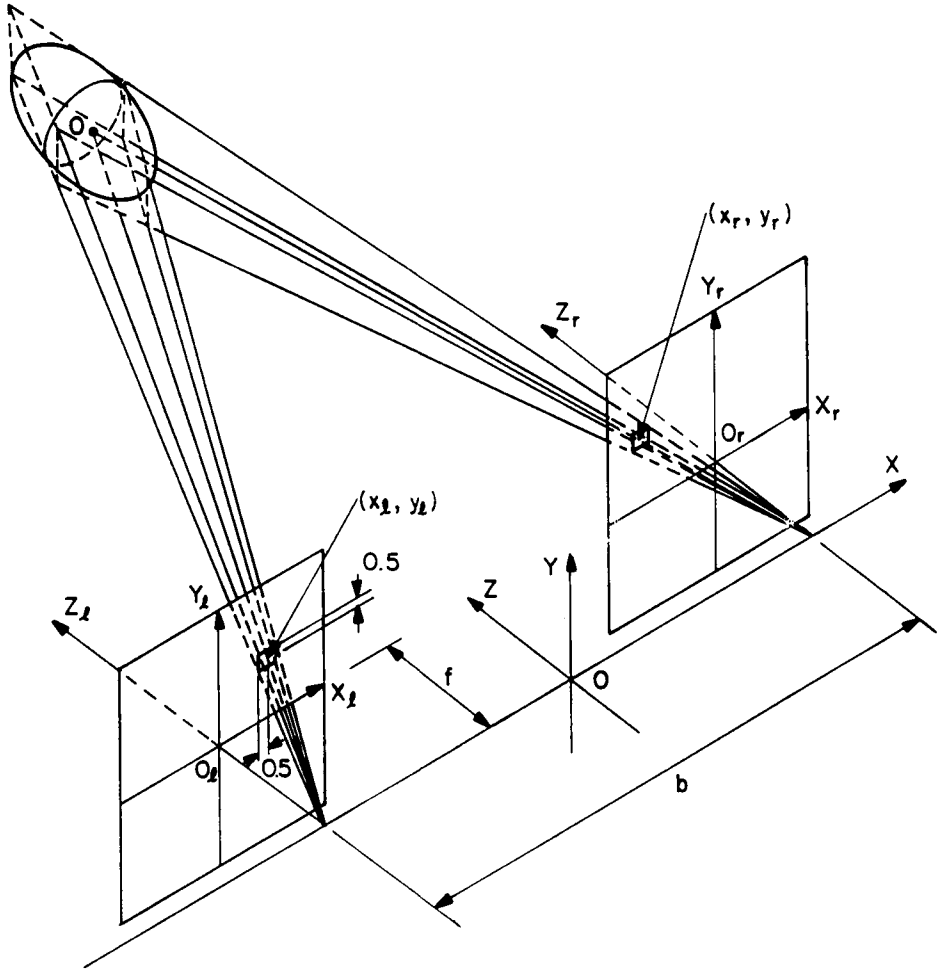


Figure 4.2: Gaussian probability description of the 3-D uncertainty volume.

$$\underline{X} = [x, y, z]^T \tag{4.6}$$

$$= \underline{m} + \underline{U}, \tag{4.7}$$

where

$$\underline{m} = [x_m, y_m, z_m]^T$$

is the center of the ellipsoid, and $\underline{U}$ is Gaussian white noise

$$\underline{U} \sim N(0, \mathbf{C}),$$

which represents the spatial uncertainties involved in the calculation of the 3-D point $\underline{X}$. For simplicity, in this work a special case of Gaussian distribution, spherical Gaussian distribution

$$p(x, y, z) = \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-\frac{1}{2}[\frac{(x-m_x)^2}{\sigma^2} + \frac{(y-m_y)^2}{\sigma^2} + \frac{(z-m_z)^2}{\sigma^2}]}, \tag{4.8}$$

is used. The equal probability surface described by Equation (4.8) forms a sphere, as shown in Figure 4.3. The 3-D point determined under the assumption of spherical Gaussian distribution is

$$\underline{X} = \underline{m} + \underline{U}, \tag{4.9}$$

where

$$\underline{m} = [x_m, y_m, z_m]^T,$$

and

$$\underline{U} \sim N(0, \mathbf{C}), \quad \mathbf{C} = diag\ \{\sigma^2, \sigma^2, \sigma^2\}.$$

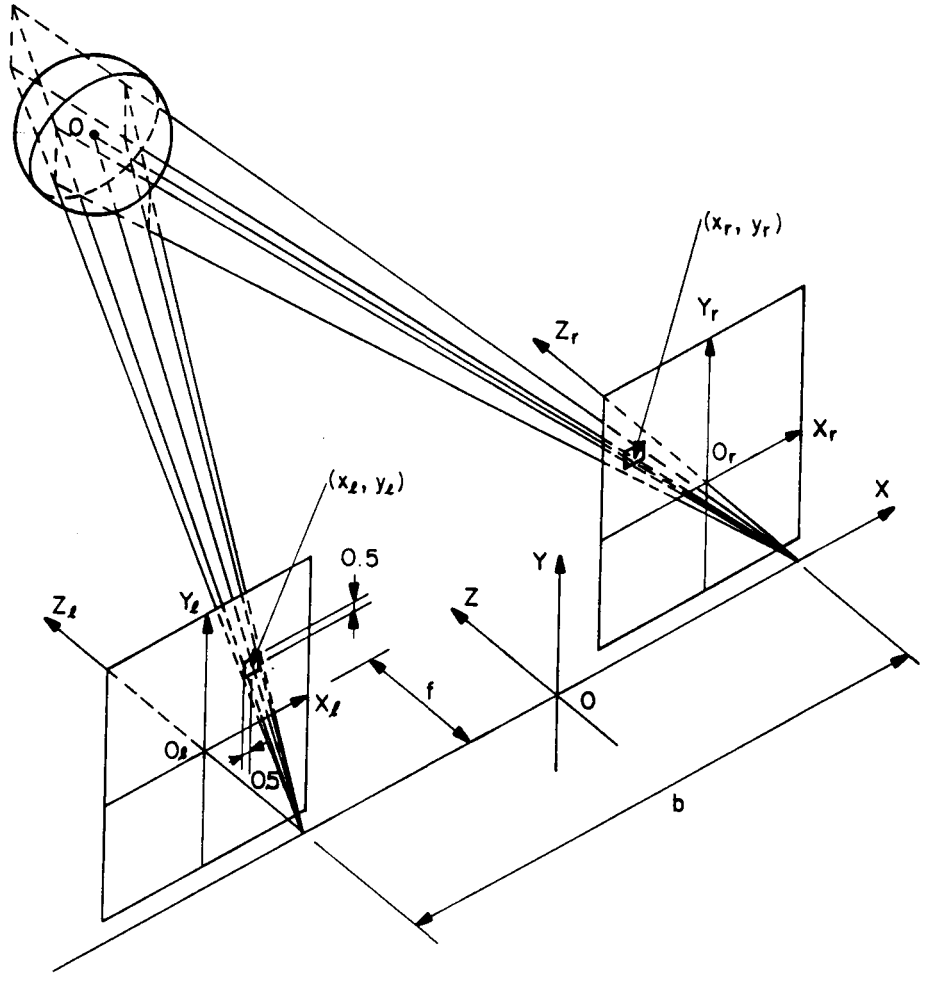


Figure 4.3: Spherical Gaussian probability description of the 3-D uncertainty volume.

4.1.2 Calculation of Gaussian Parameters

Every match determines a 3-D uncertainty volume specified by Equation (4.8). Therefore, the Gaussian parameters $\underline{m}$ and $\mathbf{C}$ can be calculated from the coordinates of each match (x_l, y_l) and (x_r, y_r). From Equation (4.8), it follows that a fixed probability determines a sphere

$$q^2 = [\frac{(x-m_x)^2}{\sigma^2} + \frac{(y-m_y)^2}{\sigma^2} + \frac{(z-m_z)^2}{\sigma^2}], \tag{4.10}$$

where q is given by

$$p = -\frac{1}{\sqrt{2\pi}} + 2erf(q) - \sqrt{\frac{2}{\pi}} q e^{-\frac{q^2}{2}}, \tag{4.11}$$

p is the probability of a 3-D point lying within the ellipsoid specified by Equation (4.10), and $erf()$ is the error function [116]

$$erf(q) = \frac{2}{\sqrt{\pi}} \int_0^q e^{-v^2} dv. \tag{4.12}$$

Equation (4.10) specifies a sphere with a radius of $q\sigma$, centered at (m_x, m_y, m_z). To simplify the calculations of (m_x, m_y, m_z) and σ, in terms of x_l, y_l and x_r, y_r, we assume that the surfaces $ABCD$ and $A'B'C'D'$ are parallel. Therefore, the parameters of the sphere can be readily derived from a 3-D region $ABCD$. With reference to Figure 4.4, where only the projection of the uncertainty volume onto the XZ plane is depicted, the parameters are calculated as follows:

a) Mean vector $\underline{m}$:

$$m_x = \frac{b}{2} \frac{x_l(i) + x_r(m)}{x_l(i) - x_r(m)}, \tag{4.13}$$

$$m_y = \frac{b}{2} \frac{y_l(i) + y_r(m)}{x_l(i) - x_r(m)}, \tag{4.14}$$

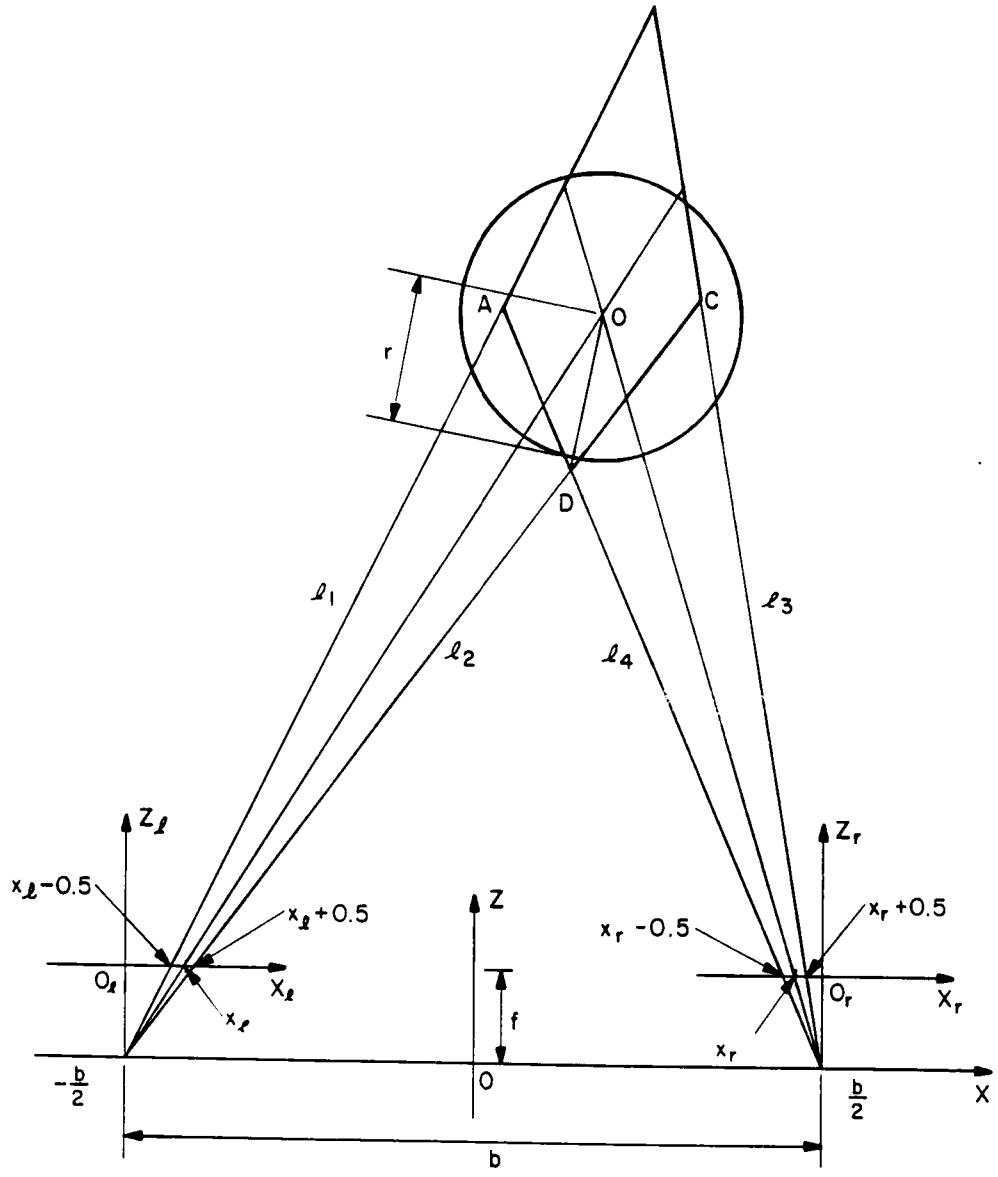


Figure 4.4: A 3-D uncertainty region in the geometrical modeling for a pair of images.

$$m_z = b\frac{f}{x_l(i) - x_r(m)}. \tag{4.15}$$

b) Variance σ^2:

from Equation (4.10), we can directly write the variance σ^2 as

$$\sigma^2 = (r/q)^2, \tag{4.16}$$

where r is the radius of the sphere, defined by

$$r = 0.9\overline{OD}, \tag{4.17}$$

and where

$$\overline{OD} = \sqrt{(x_D - m_x)^2 + (y_D - m_y)^2 + (z_D - m_z)^2}. \tag{4.18}$$

The coordinates of point D are functions of x_l, y_l and x_r, y_r:

$$x_D = \frac{b}{2}\frac{x_l + x_r}{x_l - x_r + 1}, \tag{4.19}$$

$$y_D = \frac{b}{2}\frac{y_l + y_r + 1}{x_l - x_r + 1}, \tag{4.20}$$

$$z_D = b\frac{f}{x_l - x_r + 1}. \tag{4.21}$$

4.2 Geometrical Modeling for Sequences of Images

When cameras move with known motion parameters, the geometrical structures determined by each pair of images at different moments intersect each other. Because of the finite number of photoelements in the cameras, the number of regions and the shapes of the regions involved in intersecting change with

the cameras' motion. All of these create complexity in modeling the geometrical structures. In this section, the inherent relationship between image pairs is revealed and the strengths of links between image pairs are calculated. These links play an important role in building a dynamic system model for information fusion.

4.2.1 Transformation and Calculation of Geometrical Structures for Sequences of Images

A geometrical structure for sequences of binocular images is shown in Figure 4.5, where the sphere C_k is represented with respect to the coordinate system $X_kY_kZ_k$, and C_{k+1}^i, $i = 1, ...4$, are represented in the coordinate system $X_{k+1}Y_{k+1}Z_{k+1}$. To reduce the accumulated error, a relative reference coordinate system fixed with cameras is adopted. When cameras move (described by a rotation matrix $\mathbf{J}_k$ and a translation vector $\underline{T}_k$) the reference coordinate system also moves. To derive the relationship of information between image pairs, we need to transform the location of sphere C_k (center) from the coordinate system $X_kY_kZ_k$ to the coordinate system $X_{k+1}Y_{k+1}Z_{k+1}$, denoted as C_{k+1}, and to calculate the spheres intersecting the sphere C_{k+1}.

4.2.1.1 Transformation between Coordinate Systems

The transformation between coordinate systems is specified by the motion parameters, rotation matrix $\mathbf{J}_k$ and translation vector $\underline{T}_k$, as

$$\mathbf{J}_k = \mathbf{J}_{0k} + \Delta\mathbf{J}_k, \quad \Delta\mathbf{J}_k \sim N(0, \mathbf{Q}_{J_k}), \tag{4.22}$$

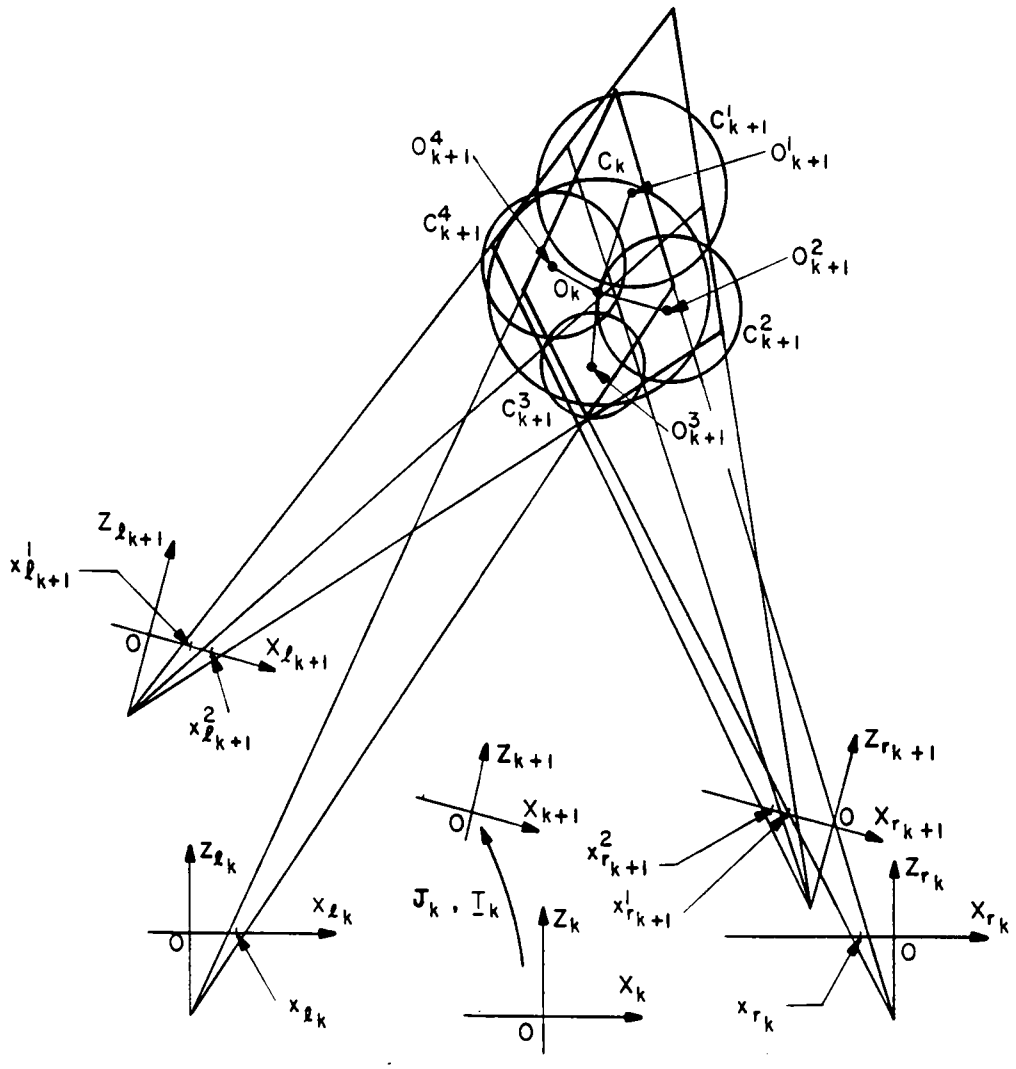


Figure 4.5: Intersected 3-D uncertainty volumes in the geometrical modeling for sequences of binocular images.

$$\underline{T}_k = \underline{T}_{0k} + \Delta\underline{T}_k, \quad \Delta\underline{T}_k \sim N(0, \mathbf{Q}_{T_k}), \tag{4.23}$$

where $\mathbf{J}_{0k}$ and $\underline{T}_{,k}$ are nominal values, and $\Delta\mathbf{J}_k$ and $\Delta\underline{T}_k$ are uncertainties of the motion parameters characterized by $\mathbf{Q}_{J_k}$ and $\mathbf{Q}_{T_k}$. Since ΔJ_k is a 3×3 matrix, its covariance matrix is composed of the covariances of the 9 elements of ΔJ_k. Therefore, the above covariance matrix $\mathbf{Q}_{J_k} \in \mathbf{R}^{9\times 9}$ should be understood as

$$\mathbf{Q}_{J_k} = E\{[\Delta J_{11}, \Delta J_{12}, ..., \Delta J_{32}, \Delta J_{33}]^T [\Delta J_{11}, \Delta J_{12}, ..., \Delta J_{32}, \Delta J_{33}]\}.$$

The transformation of the center of sphere C_k from the coordinate system $X_kY_kZ_k$ to the coordinate system $X_{k+1}Y_{k+1}Z_{k+1}$ is performed by

$$\underbrace{\begin{bmatrix} x_{O_{k+1}} \\ y_{O_{k+1}} \\ z_{O_{k+1}} \end{bmatrix}}_{\underline{O}_{k+1}} = \underbrace{\begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix}_k}_{\mathbf{J}_k} \underbrace{\begin{bmatrix} x_{O_k} \\ y_{O_k} \\ z_{C_k} \end{bmatrix}}_{\underline{O}_k} + \underbrace{\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}_k}_{\underline{T}_k}. \tag{4.24}$$

Since $\mathbf{J}_k$ and $\underline{T}_k$ are with uncertainties, the transformed center is also with uncertainties:

$$\underline{O}_{k+1} = \tilde{\underline{O}}_{k+1} + \Delta\underline{O}_{k+1}, \quad \Delta\underline{O}_{k+1} \sim (0, \mathbf{Q}_{O_{k+1}}), \tag{4.25}$$

where $\tilde{\underline{O}}_{k+1}$ is the mean value of the center vector determined by the nominal values of $\mathbf{J}_k$ and $\underline{T}_k$

$$\tilde{\underline{O}}_{k+1} = [\tilde{x}_{O_{k+1}}, \tilde{y}_{O_{k+1}}, \tilde{z}_{O_{k+1}}]^T \tag{4.26}$$

$$= \mathbf{J}_{0k}\underline{O}_k + \underline{T}_{0k}, \tag{4.27}$$

and $\mathbf{Q}_{O_{k+1}}$ is the covariance matrix specified in the following. To calculate the covariance matrix $\mathbf{Q}_{O_{k+1}}$, Equation (4.24) is expanded into a Taylor series at the

mean values of the transformed center $\underline{\tilde{O}}_{k+1}$. Using a procedure similar to the transformation given in Appendix A.2,

$$\Delta x_{O_{k+1}} = x_{O_{k+1}} - \tilde{x}_{O_{k+1}} \tag{4.28}$$

$$= \underline{Q}_k^T \begin{bmatrix} \Delta J_{11_k} \\ \Delta J_{12_k} \\ \Delta J_{13_k} \end{bmatrix} + \Delta T_{1_k}, \tag{4.29}$$

$$\Delta y_{O_{k+1}} = y_{O_{k+1}} - \tilde{y}_{O_{k+1}} \tag{4.30}$$

$$= \underline{Q}_k^T \begin{bmatrix} \Delta J_{21_k} \\ \Delta J_{22_k} \\ \Delta J_{23_k} \end{bmatrix} + \Delta T_{2_k}, \tag{4.31}$$

and

$$\Delta z_{O_{k+1}} = z_{O_{k+1}} - \tilde{z}_{O_{k+1}} \tag{4.32}$$

$$= \underline{Q}_k^T \begin{bmatrix} \Delta J_{31_k} \\ \Delta J_{32_k} \\ \Delta J_{33_k} \end{bmatrix} + \Delta T_{3_k}. \tag{4.33}$$

Putting Equations (4.28), (4.30), and (4.32) together gives

$$\underbrace{\begin{bmatrix} \Delta x_{O_{k+1}} \\ \Delta y_{O_{k+1}} \\ \Delta z_{O_{k+1}} \end{bmatrix}}_{\Delta \underline{O}_{k+1}} = \underbrace{\begin{bmatrix} \underline{O}_K^T & & \\ & \underline{O}_k^T & \\ & & \underline{O}_k^T \end{bmatrix}}_{\mathbf{J}_{a_k}} \underbrace{\begin{bmatrix} \Delta J_{11_k} \\ \Delta J_{12_k} \\ \Delta J_{13_k} \\ \Delta J_{21_k} \\ \Delta J_{22_k} \\ \Delta J_{23_k} \\ \Delta J_{31_k} \\ \Delta J_{32_k} \\ \Delta J_{33_k} \end{bmatrix}}_{\Delta J_k} + \underbrace{\begin{bmatrix} \Delta T_{1_k} \\ \Delta T_{2_k} \\ \Delta T_{3_k} \end{bmatrix}}_{\Delta \underline{T}_k}. \tag{4.34}$$

The covariance of the transformed center is then

$$E\{\Delta \underline{O}_{k+1} \Delta \underline{O}_{k+1}^T\} = \mathbf{J}_{a_k} \mathbf{Q}_{J_k} \mathbf{J}_{a_k}^T + \mathbf{Q}_{T_k} \tag{4.35}$$

$$= \mathbf{Q}_{O_{k+1}}, \tag{4.36}$$

where $\mathbf{J}_{a_k}$ is called Jacobian. Equation (4.36) geometrically represents an ellipsoid, Figure 4.6, and is taken into account in the measurement models in Chapter 5.

4.2.1.2 Calculation of Spheres Intersecting C_{k+1}

To calculate the spheres intersecting the sphere C_{k+1}, the pixels on the image planes at the $(k+1)th$ moment, where the sphere C_{k+1} can be seen, need to be calculated first. The calculation is detailed in the following.

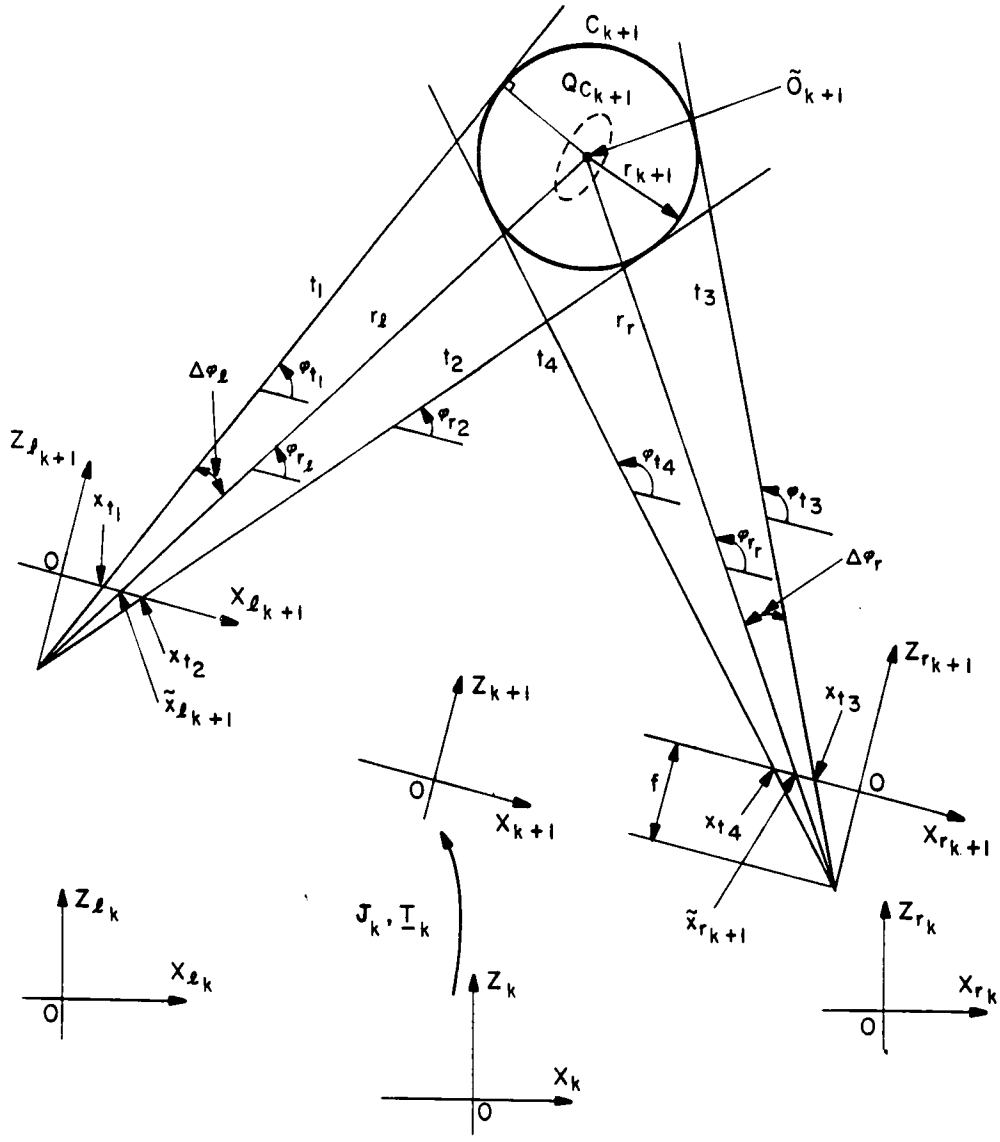


Figure 4.6: Transformation between coordinate systems and prediction of locations of pixels in next image pair where the sphere C_{k+1} can be seen.

The sphere C_{k+1} is projected onto the left and right image planes at the $(k+1)th$ moment. The projections of the sphere C_{k+1} are ellipses. To simplify, we approximate the ellipses by circles whose radii are obtained by calculating x_{t1}, x_{t2}, x_{t3}, and x_{t4} in Figure 4.6. Since

$$r_l = \sqrt{(\tilde{x}_{O_{k+1}} + \frac{b}{2})^2 + (\tilde{z}_{O_{k+1}} + f)^2}, \tag{4.37}$$

$$\Delta\varphi_l = sin^{-1}\frac{r_{k+1}}{r_l}, \tag{4.38}$$

$$\varphi_{r_l} = tan^{-1}\frac{\tilde{z}_{O_{k+1}} + f}{\tilde{x}_{O_{k+1}} + \frac{b}{2}}, \text{ and} \tag{4.39}$$

$$\varphi_{t1} = \varphi_{r_l} + \Delta\varphi_l \tag{4.40}$$

the intersection of tangent $t1$ with the X axis is

$$x_{t1} = \frac{f}{tan\ \varphi_{t1}}. \tag{4.41}$$

Similarly,

$$x_{t2} = \frac{f}{tan\ \varphi_{t2}}, \quad \varphi_{t2} = \varphi_{r_L} - \Delta\varphi_l. \tag{4.42}$$

The values of x_{t3} and x_{t4} are obtained in the same way:

$$x_{t3} = \frac{f}{tan\ \varphi_{t3}}, \quad \varphi_{t3} = \varphi_{r_r} - \Delta\varphi_r, \tag{4.43}$$

$$x_{t4} = \frac{f}{tan\ \varphi_{t4}}, \quad \varphi_{t4} = \varphi_{r_r} - \Delta\varphi_r, \tag{4.44}$$

where

$$\varphi_{r_r} = tan^{-1}\frac{\tilde{z}_{O_{k+1}} + f}{\tilde{x}_{O_{k+1}} - \frac{b}{2}}, \tag{4.45}$$

and

$$\Delta\varphi_r = sin^{-1}\frac{r_{k+1}}{r_r}. \tag{4.46}$$

The projections of the center of the sphere C_{k+1} are

$$\tilde{x}_{l_{k+1}} = \frac{f(\tilde{x}_{O_{k+1}} + \frac{b}{2})}{\tilde{z}_{O_{k+1}}}, \tag{4.47}$$

and

$$\tilde{x}_{r_{k+1}} = \frac{f(\tilde{x}_{O_{k+1}} - \frac{b}{2})}{\tilde{z}_{O_{k+1}}}. \tag{4.48}$$

Thus, the radii of the circles of projections are $|\tilde{x}_{l_{k+1}} - x_{t_1}|$ and $|\tilde{x}_{r_{k+1}} - x_{t_4}|$, and the pixels where the sphere C_{k+1} can be seen are those lying within the circles. Then the spheres which intersect the sphere C_k are calculated by the method discussed in Section 4.1.

4.2.2 Calculation of the Strength of Links between Spheres

Once the center of sphere C_k has been transformed into the coordinate system $X_{k+1}Y_{k+1}Z_{k+1}$ and the spheres intersecting C_{k+1} have been determined, the strength of the links between the two spheres can be readily obtained by calculating the size of the common volume of these two spheres. For clarity, we examine two spheres only. These spheres are extracted from Figure 4.5 and shown in Figure 4.7. To simplify the computation, we calculate the area of the intersection S^1 of two 2-D circles, $C_{2D_{k+1}}$ and $C^1_{2D_{k+1}}$, since the common volume of the two intersected spheres is proportional to the 2-D intersection area.

The procedure for calculating the strength of the link between spheres C_{k+1} and C^i_{k+1}, $i = 1, ..., N$, where N is the number of spheres intersecting the sphere C_{k+1}, involves the following steps:

1) Calculation of the centers and the radii of two spheres

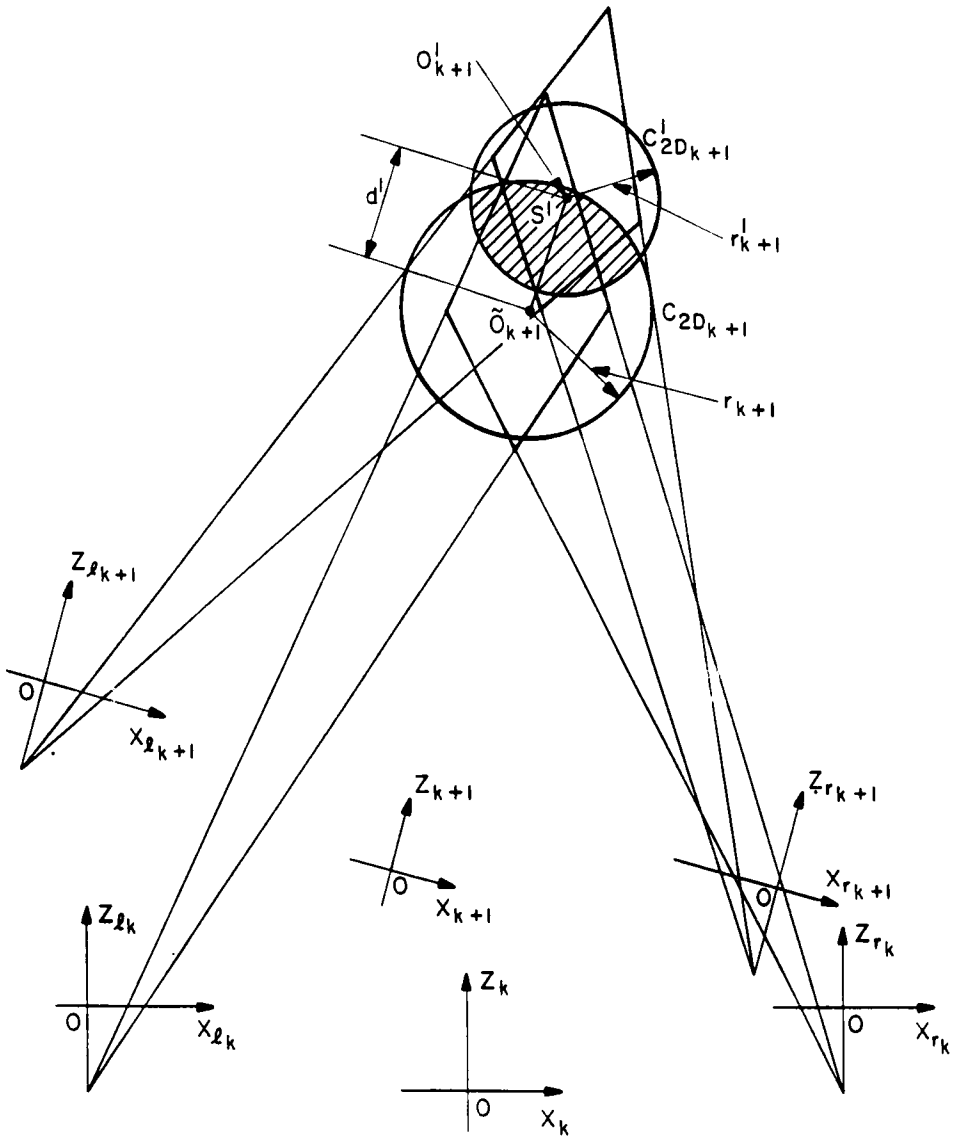


Figure 4.7: Two extracted 3-D uncertainty volumes.

The mean of the transformed center of sphere C_{k+1} is given by Equation (4.26), and the radius of C_{k+1} is

$$r_{k+1} = r_k = \sigma_k \; q. \tag{4.49}$$

The centers and the radius of C^i_{k+1} can be calculated using the method discussed in Section 4.1. They are

$$\underline{O}^i_{k+1} = [m_{x^i_{k+1}}, m_{y^i_{k+1}}, m_{z^i_{k+1}}]^T, \quad r^i_{k+1} = \sigma^i_{k+1} \; q. \tag{4.50}$$

2) Calculation of 2-D intersected area S^i

A local coordinate system is used to calculate the intersected area S^i for each pair of intersected spheres C_{k+1} and C^i_{k+1}. The origin of the local coordinate system is located in the center $\tilde{O}_{k+1}$ of the sphere C_{k+1}, and the X axis is aligned with the connection between two centers $\tilde{O}_{k+1}$ and O^i_{k+1}, as shown in Figure 4.8. There are five cases to be considered.

Case 1: $d^i \geq r_{k+1} + r^i_{k+1}$.

In this case, there is no intersection between two spheres. Therefore, $S^i = 0$.

Case 2: $d^i < r_{k+1} + r^i_{k+1}$ and $\phi^i_{k+1} > \frac{\pi}{2}$, Figure 4.8.

The intersected area S^i is given by

$$S^i = S_{PMN} + S_{MQN}, \tag{4.51}$$

where

$$S_{PMN} = S_{MPNO^i_{k+1}} - S_{MNO^i_{k+1}} \tag{4.52}$$

$$= (r^i_{k+1})^2(\pi - \phi^i_{k+1}) - (d^i - |a|)|b|, \tag{4.53}$$

and

$$S_{MQN} = S_{MQN\tilde{O}_{k+1}} - S_{MN\tilde{O}_{k+1}} \tag{4.54}$$

$$= r^2_{k+1}\pi_k - |a||b|. \tag{4.55}$$

Here, a and b are the coordinates of point M:

$$a = \frac{r^2_{k+1} + (d^i)^2 - (r^i_{k+1})^2}{2d^i}, \quad b = \pm\sqrt{r^2_{k+1} - a^2}, \tag{4.56}$$

and

$$\phi_{k+1} = tan^{-1}(\frac{b}{a}), \quad \phi^i_{k+1} = \pi - tan^{-1}(\frac{b}{d^i - a}). \tag{4.57}$$

Case 3: $d^i < r_{k+1} + r^i_{k+1}$ and $\phi^i_{k+1} = \frac{\pi}{2}$, Figure 4.9.

In this case, the area S^i is

$$S^i = r^2_{k+1}\phi_{k+1} - |a||b| + \frac{1}{2}\pi(r^i_{k+1})^2 \tag{4.58}$$

where a and b are given by Equation (4.56).

Case 4: $d^i < r_{k+1} + r^i_{k+1}$ and $\phi^i_{k+1} < \frac{\pi}{2}$, Figure 4.10.

Similarly, we have

$$S^i = r_{k+1}^2 \phi_{k+1} - |a||b| + (r_{k+1}^i)^2(\pi - \phi_{k+1}^i) + (|a| - d^i)|b|, \tag{4.59}$$

where a and b are also given by Equation (4.56).

Case 5: $r_k > r_{k+1}^i + d^i$, Figure 4.11.

In this special case,

$$S^i = \pi(r_{k+1}^i)^2. \tag{4.60}$$

In Chapter 5, the strength of the links between spheres developed in this chapter is used to build the dynamic system model for information fusion.

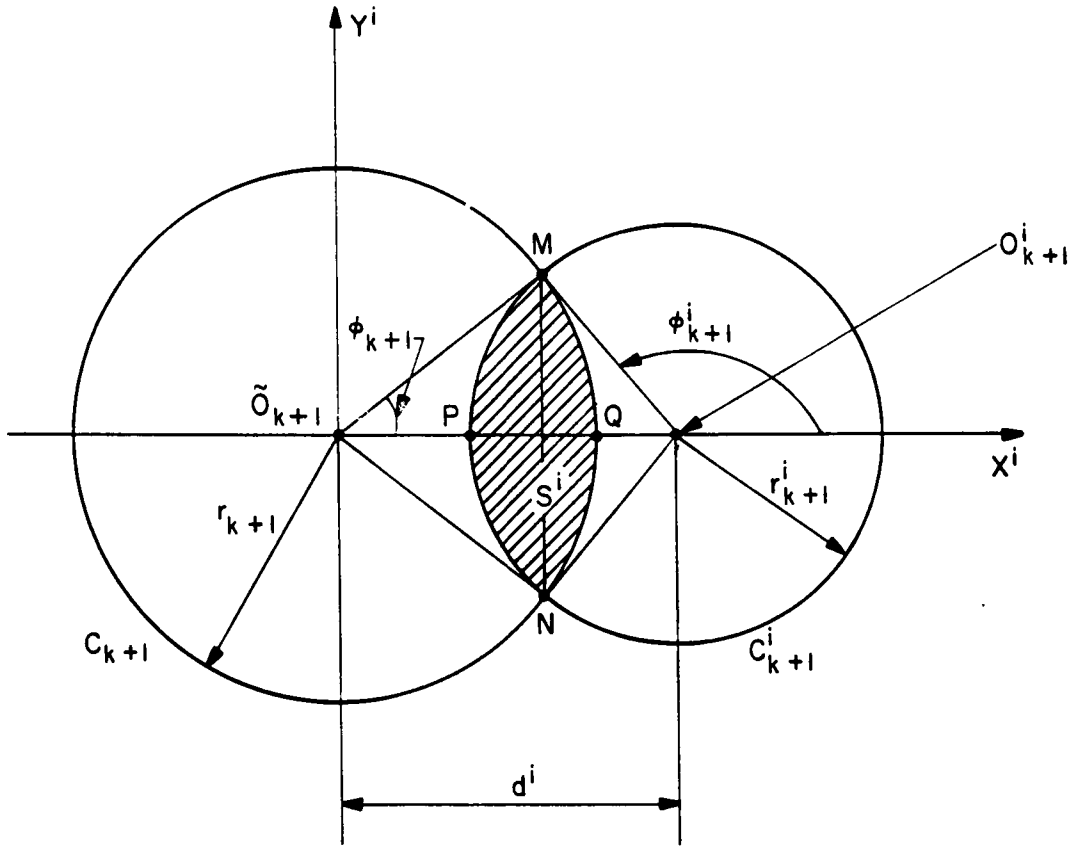


Figure 4.8: Case two in calculating intersected area S^i.

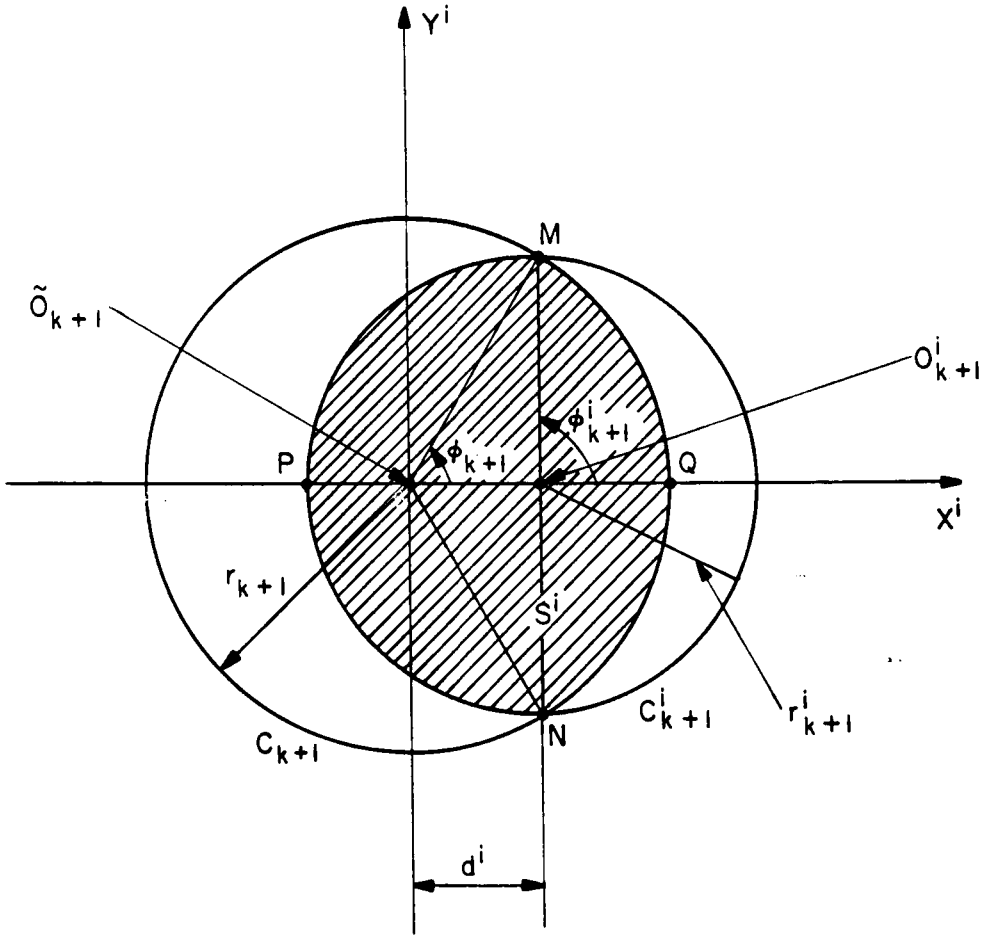


Figure 4.9: Case three in calculating intersected area S^i.

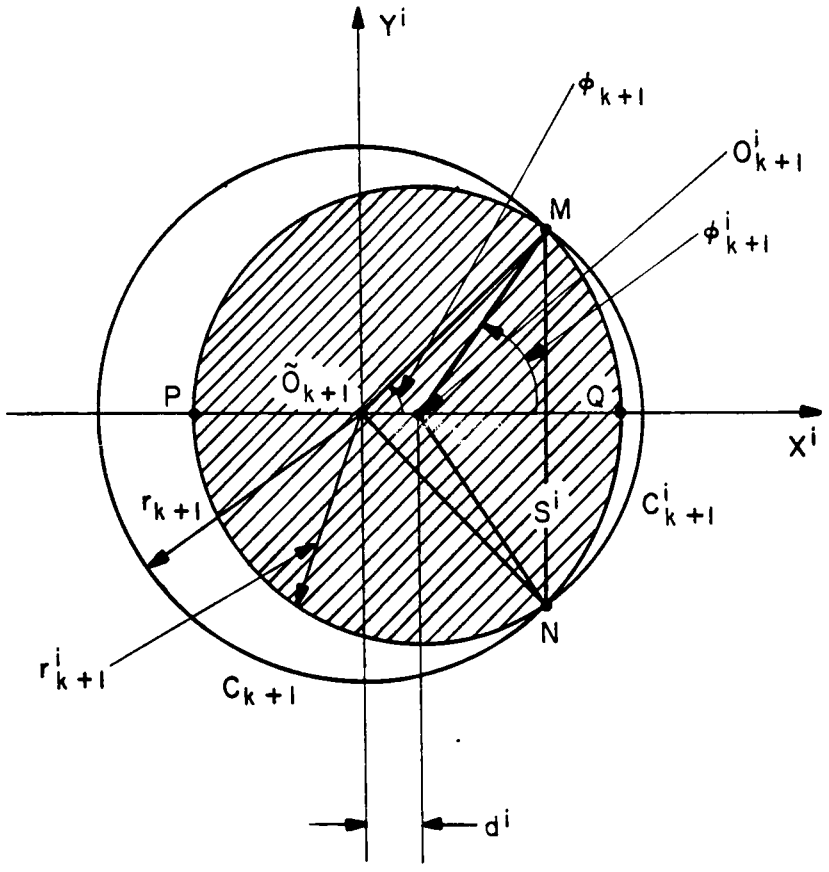


Figure 4.10: Case four in calculating intersected area S^i.

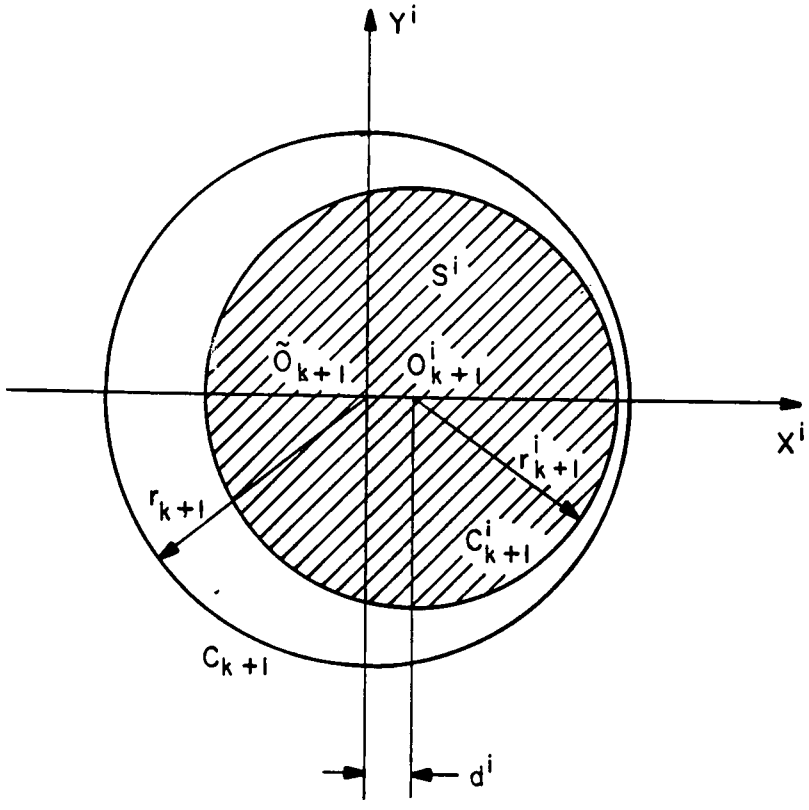


Figure 4.11: Case five in calculating intersected area S^i.

CHAPTER 5

MODEL-BASED INFORMATION FUSION

In this chapter, the confidence values of each possible match, obtained by the matching procedure described in Chapter 3, are updated, and the spatial uncertainties in the 3-D information, due to finite resolution of cameras, are reduced by fusing the information provided by the binocular image pairs and image sequences. General concepts of information fusion are discussed in Section 5.1. The dynamic models for confidence value fusion and for 3-D information fusion are derived in Section 5.2; while Section 5.3 concentrates on fusion algorithms.

5.1 Concepts of Information Fusion

As described in Chapter 4, each possible match determines a 3-D uncertainty volume. If at the kth moment, a possible match is correct, i.e., two pixels in the left and right images correspond to the same 3-D point (or collection of 3-D points), this 3-D point (or 3-D points) will be seen again at the $(k+1)th$ moment at predicted places, since the motion of the cameras is known. For instance, in Figure 4.5, if (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k})[1]

[1]Since Figure 4.5 only shows a 2-D projection of a 3-D structure, one can only see x_{l_k} and x_{r_k}. Similarly, for other pixels, only x components can be seen.

are correctly matched, the 3-D points in the sphere C_k will be seen at $(x^1_{l_{k+1}}, y^1_{l_{k+1}})$, $(x^2_{l_{k+1}}, y^2_{l_{k+1}})$ and $(x^1_{r_{k+1}}, y^1_{r_{k+1}})$, $(x^2_{r_{k+1}}, y^2_{r_{k+1}})$, since the sphere C_k is intersected by spheres C^1_{k+1}, C^2_{k+1}, C^3_{k+1} and C^4_{k+1}. Therefore, the results of matching at $(x^1_{l_{k+1}}, y^1_{l_{k+1}})$, $(x^2_{l_{k+1}}, y^2_{l_{k+1}})$ and $(x^1_{r_{k+1}}, y^1_{r_{k+1}})$, $(x^2_{r_{k+1}}, y^2_{r_{k+1}})$ will increase the confidence value assigned to the match between (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k}). Otherwise, if the match between (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k}) is not correct, the results of matching at $(x^1_{l_{k+1}}, y^1_{l_{k+1}})$, $(x^2_{l_{k+1}}, y^2_{l_{k+1}})$ and $(x^1_{r_{k+1}}, y^1_{r_{k+1}})$, $(x^2_{r_{k+1}}, y^2_{r_{k+1}})$ will decrease the confidence value of the match between (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k}). At this stage, a concept of multi-sensor fusion is introduced by considering the intersected 3-D structure shown in Figure 4.5 as a micro multi-sensor system, where the sensors are logical ones [60]. The object in the sphere C_k is observed by four logical sensors: C^1_{k+1}, C^2_{k+1}, C^3_{k+1}, and C^4_{k+1}. The dynamic model and the algorithm for the model-based fusion of confidence values are presented in the following sections.

Fusing the 3-D information provided by images taken at different times and positions reduces the 3-D information uncertainties, caused by finite camera resolution. For instance, the 3-D point determined by the match (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k}) has spatial uncertainty characterized by

$$\underline{X_k} = \underline{O_k} + \underline{U_k}, \quad \underline{U_k} \sim N(0, \mathbf{C}_k), \tag{5.1}$$

where $\underline{O_k}$ is a vector of the center of sphere C_k and the covariance matrix $\mathbf{C_k}$ is proportional to the radius of sphere C_k. The larger the radius of the sphere, the larger the spatial uncertainty. At the $(k+1)th$ moment, the point $\underline{X_k}$ is observed

at different locations, and the 3-D points calculated from these different locations are also with 3-D uncertainties:

$$\underline{X}^i_{k+1} = \underline{O}^i_{k+1} + \underline{U}^i_{k+1}, \quad \underline{U}^i_{k+1} \sim N(0, \mathbf{C}^i_{k+1}), \ i = 1, 2, 3, 4. \tag{5.2}$$

Since these spheres intersect each other and at the $(k+1)th$ moment the spheres have smaller radii, when fusing all these spatial information together, the uncertainty associated with $\underline{X}_k$ is reduced. Especially, when cameras approach the object, the fused 3-D information has less and less uncertainties. Even when cameras do not move towards the object, fusion of the 3-D information obtained at different locations reduces the 3-D uncertainties. The fusion of 3-D information is detailed in the following sections.

5.2 Dynamic Models for Fusion

The algorithms used for information fusion in this work are model-based fusion algorithms, which have the capability of dynamic information fusion. The algorithms can always incorporate new information, scheduled and unscheduled, to give the best fusion results. Deriving suitable dynamic models, the main topic of this section, is crucial in these algorithms.

5.2.1 Dynamic Models for Confidence Value Fusion

A dynamic model consists of a state model and N measurement models, where N is the number of spheres intersecting C_k, Figure 4.5.

5.2.1.1 State Model

Since the reference coordinate system moves with the cameras, at the $(k+1)th$ moment, all the previously gathered information needs to be transformed to the current coordinate system. When the sphere C_k is transformed from the coordinate system $X_kY_kZ_k$ to the coordinate system $X_{k+1}Y_{k+1}Z_{k+1}$, denoted as C_{k+1}, the confidence value associated with C_k is unchanged. Therefore, the confidence values of the sphere C_k at different moments satisfy the following state model:

$$x_{cf_{k+1}} = x_{cf_k}, \tag{5.3}$$

where x_{cf_k} is the confidence value of the sphere C_k in the coordinate system $X_k\,Y_k\,Z_k$, and $x_{cf_{k+1}}$ is the confidence value of the sphere C_{k+1} in the coordinate system $X_{k+1}\,Y_{k+1}\,Z_{k+1}$.

5.2.1.2 Measurement Models

There are N measurement models, each of which represents a logical sensor. The number of measurement models, N, depends on the number of the spheres intersecting the sphere C_{k+1} at the $(k+1)th$ moment. To derive the ith measurement model, we consider two spheres: C_{k+1} and C^i_{k+1}, Figures 4.8 – 4.11. The confidence value for the sphere C_{k+1} is $x_{cf_{k+1}}$, and the confidence value for C^i_{k+1} is $z^i_{cf_{k+1}}$ which is obtained by applying the matching procedure described in Chapter 3 to the predicted pixels, for instance, pixels $(x^1_{l_{k+1}}, y^1_{l_{k+1}})$, $(x^2_{l_{k+1}}, y^2_{l_{k+1}})$ and $(x^1_{r_{k+1}}, y^1_{r_{k+1}})$, $(x^2_{r_{k+1}}, y^2_{r_{k+1}})$ in Figure 4.5. The densities of the confidence values of sphere C_{k+1} and sphere C^i_{k+1} are

$$\frac{x_{cf_{k+1}}}{\pi r^2_{cf_{k+1}}} \quad \text{and} \quad \frac{z^i_{k+1}}{\pi (r^i_{k+1})^2}$$

and the portions of the confidence values $x_{cf_{k+1}}$ and $z^i_{cf_{k+1}}$, over the common region S^i, are

$$\frac{x_{cf_{k+1}}}{\pi r^2_{k+1}} S^i \quad \text{and} \quad \frac{z^i_{cf_{k+1}}}{\pi (r^i_{k+1})^2} S^i.$$

Considering the uncertainties in the motion parameters and the differences in the sizes of the spheres, the following holds:

$$\frac{z^i_{cf_{k+1}}}{\pi (r^i_{k+1})^2} S^i = \frac{x_{cf_{k+1}}}{\pi r^2_{k+1}} S^i + \left(\frac{r^2_{k+1} - (r^i_{k+1})^2}{r^2_{k+1}(r^i_{k+1})^2}\right)\frac{S^i}{\pi} + \tilde{w}^i_{k+1}, \tag{5.4}$$

where the term $\left(\frac{r^2_{k+1}-(r^i_{k+1})^2}{r^2_{k+1}(r^i_{k+1})^2}\right)\frac{S^i}{\pi}$ compensates for the difference in the sizes of the spheres. The term $\tilde{w}^i_{k+1}$ in Equation (5.4) represents the uncertainties in the motion parameters and it is Gaussian white noise described by:

$$\tilde{w}^i_{k+1} \sim N(0, \tilde{Q}^i_{k+1}). \tag{5.5}$$

The error covariance $\tilde{Q}^i_{k+1}$, representing the uncertainties in the camera motion parameters and the noise in the images, is given by

$$\tilde{Q}^i_{k+1} = \beta^i_1 q^i + \beta^i_2 \sigma_{k+1} \tag{5.6}$$

where q^i is the portion of the ellipsoid $\mathbf{Q}_{C_{k+1}}$ on the connection of two sphere centers, Figure 5.1. The coefficient σ_{k+1} is the standard deviation of the noise of images at the $(k+1)th$ moment. The coefficients β^i_1 and β^i_2 determine a weight distribution between q^i and σ_{k+1}. Finally, the ith measurement model is acquired by rearranging Equation (5.4) into

$$z^i_{cf_{k+1}} = H^i_{k+1} x_{cf_{k+1}} + w^i_{k+1}, \tag{5.7}$$

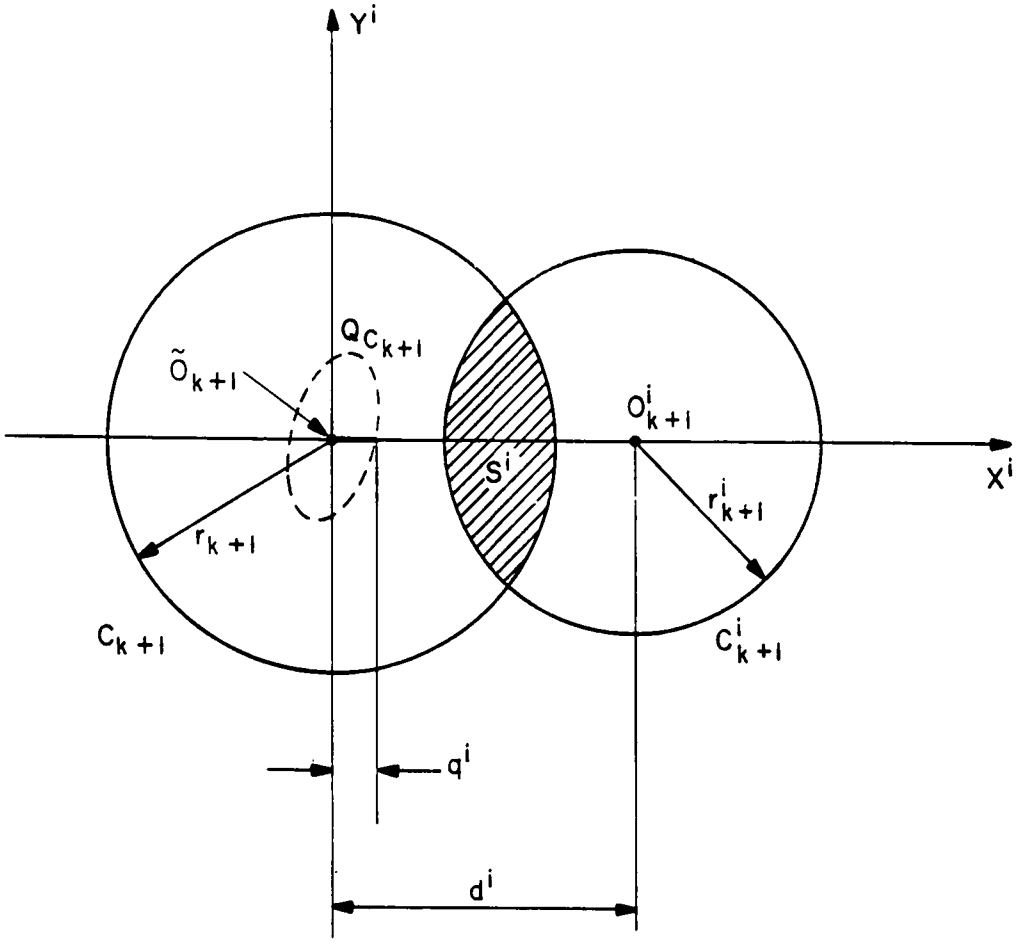


Figure 5.1: Determination of the measurement error due to the uncertainties in camera's motion.

where

$$H^i_{k+1} = \frac{r^i_{k+1}}{r^2_{k+1}}, \tag{5.8}$$

and

$$w^i_{k+1} \sim N(m^i_{k+1}, R^i_{k+1}) = N(\frac{r^2_{k+1} - (r^i_{k+1})^2}{r^2_{k+1}}, \frac{\pi(r^i_{k+1})^2}{S^i}\alpha^i q^i). \tag{5.9}$$

Equation (5.9) shows that the smaller the common region S^i, the larger the measurement error, which means that the smaller S^i, the less effect of the measurement z^i_{k+1} on the fusion of confidence value. When $S^i = 0$, the measurement error is infinite, i.e., the measurement z^i_{k+1} contains no useful information to update the confidence value x_{k+1}.

To summarize, the dynamic model for each confidence value is

$$\begin{aligned} x_{cf_{k+1}} &= x_{cf_k} \\ z^i_{cf_{k+1}} &= H^i_{k+1} x_{cf_{k+1}} + w^i_{k+1}, \quad i = 1, ..., N. \end{aligned} \tag{5.10}$$

5.2.2 Dynamic Models for 3-D Information Fusion

The dynamic model for 3-D information fusion is directly related to the motion of cameras and reflects the advantage of imposing surface continuity constraint in 3-D space.

5.2.2.1 State Model

As discussed in Section 5.1, the 3-D point calculated from the match (x_{l_k}, y_{l_k}) and (x_{r_k}, y_{r_k}) has an associated spatial uncertainty:

$$\underline{X}_k = [x_k, y_k, z_k]^T \tag{5.11}$$

$$= \underline{O}_k + \underline{U}_k, \tag{5.12}$$

where $\underline{O}_k$ is the position vector of the center of C_k, and

$$\underline{U}_k \sim N(0, \mathbf{C}_k), \quad \mathbf{C}_k = diag\{\sigma_k^2, \sigma_k^2, \sigma_k^2\}. \tag{5.13}$$

When transformed from $X_kY_kZ_k$ to $X_{k+1}Y_{k+1}Z_{k+1}$, the 3-D point represented in the later coordinates satisfies the following state model:

$$\underline{X}_{k+1} = \mathbf{J}_k\underline{X}_k + \underline{T}_k, \tag{5.14}$$

where $\mathbf{J}_k$ and $\underline{T}_k$ are given by Equations (4.21) and (4.22). This state model is time varying, and the system matrix $\mathbf{J}_k$ is with uncertainty.

5.2.2.2 Measurement Models

Each possible match found in the predicted circles at the $(k+1)th$ moment, for instance, the possible matches among $(x^1_{l_{k+1}}, y^1_{l_{k+1}})$, $(x^2_{l_{k+1}}, y^2_{l_{k+1}})$ and $(x^1_{r_{k+1}}, y^1_{r_{k+1}})$, $(x^2_{r_{k+1}}, y^2_{r_{k+1}})$, determines a 3-D point with spatial uncertainty:

$$\underline{X}^i_{k+1} = \underline{O}^i_{k+1} + \underline{U}^i_{k+1}, \underline{U}^i_{k+1} \sim N(0, \mathbf{C}^i_{k+1}). \tag{5.15}$$

The measurement models which relate the 3-D points in spheres C_k and C^i_{k+1} are given by

$$\underline{X}^i_{k+1} = \underline{X}_{k+1} + \underline{\tilde{w}}^i_{k+1}, \quad i = 1, ..., N, \tag{5.16}$$

where $\underline{X}^i_{k+1}$ is 3-D information provided by C^i_{k+1} to update $\underline{X}_{k+1}$, and the measurement errors $\underline{\tilde{w}}^i_{k+1}$ are

$$\underline{\tilde{w}}^i_{k+1} \sim N(\underline{O}^i_{k+1} - \overline{\underline{O}}_{k+1}, \mathbf{C}^i_{k+1}). \tag{5.17}$$

Since the spheres C^i_{k+1} are determined by the pixels in the predicted circles, the centers of C^i_{k+1} lie within the sphere C_{k+1}. Therefore, $\underline{X}^i_{k+1}$ and $\underline{X}_{k+1}$ are considered to be on a smooth surface. This consideration satisfies the 3-D surface continuity constraint. To use conventional notation, we denote measurements by $\underline{Z}^i_{k+1}$. Therefore, Equation (5.16) becomes

$$\underline{Z}^i_{k+1} = \underline{X}_{k+1} + \underline{\tilde{w}}^i_{k+1}, \quad i = 1, ..., N. \tag{5.18}$$

Consequently, the dynamic model for 3-D information fusion is

$$\begin{aligned} \underline{X}_{k+1} &= \mathbf{J}_k \underline{X}_k + \underline{T}_k, \\ \underline{Z}^i_{k+1} &= \underline{X}_{k+1} + \underline{\tilde{w}}^i_{k+1}, \quad i = 1, ..., N. \end{aligned} \tag{5.19}$$

5.3 Fusion Algorithms

The distributed model-based fusion algorithms are developed here for information fusion. By model-based we mean that the behavior of the information to be updated can be described by a dynamic (or static) model (state model), and the relationship among sources of information to be fused can be specified by a set of models (measurement models). The developed model-based algorithms are

- Optimal in the sense of minimizing a specific cost function, such as fusion error covariances

- Computationally efficient
- Capable of dynamic fusion
- Easy to implement
- Flexible in manipulating uncertainties and noise

The algorithm for confidence value fusion is presented in Section 5.3.1, while the algorithm for 3-D information fusion is developed in Section 5.3.2. More general algorithms for multi-sensor/information fusion are presented in Appendix A.4.

5.3.1 An Algorithm for Confidence Value Fusion

To achieve computational efficiency, a distributed fusion algorithm is employed. In this algorithm, the confidence value of C_{k+1} is locally updated by each confidence value obtained in $C^i_{k+1}, i = 1, ..., N$. Local updates, which are very time consuming, are independent of each other and can be performed in parallel. Therefore, computational efficiency is acquired. The global fusion result is obtained by optimally fusing N locally [2] updated confidence values. The following algorithm is applied to each level of resolution hierarchy. Only initial conditions are different for different levels in the algorithm. The image sequences are assumed to have M image pairs. The zero moment is the moment when the first image pair was taken. For the top level of the hierarchy ($l = 1$), the initial confidence value, $\overline{x}^1_{cf_0}$, is the confidence value obtained by applying the matching

[2] In the algorithms to be presented, the concept *local* is associated with the spheres C^i_{k+1}, while the term *global* or *central* pertains to the sphere C_{k+1}.

procedure to the first image pair; the initial error covariance $P^1_{cf_0}$ depends on the degree of the noise in the images. Since the algorithm converges rapidly, choosing the value of $P^1_{cf_0}$ is not critical. For other levels of the hierarchy, the initial confidence value is the fused confidence value at the upper level, and the initial error covariance is the fused error covariance at the upper level, i.e., $\overline{x}^l_{cf_0} = \hat{x}^{l-1}_{cf_{M|M}}$ and $P^l_{cf_0} = P^{l-1}_{cf_{M|M}}$, Figure 5.2. The algorithm for hierarchy level l involves the following steps, beginning with initial conditions:

$$\hat{x}^l_{cf_{0|-1}} = \overline{x}^l_{cf_0}, \quad P^l_{cf_{0|-1}} = P^l_{cf_0}. \tag{5.20}$$

- Propagation of globally-fused confidence value of the kth moment and the associated error covariance from t_k to t_{k+1}:

$$\hat{x}^l_{cf_{k+1|k}} = \hat{x}^l_{cf_{k|k}}, \tag{5.21}$$

$$P^l_{cf_{k+1|k}} = P^l_{cf_{k|k}}. \tag{5.22}$$

- Local update of the propagated, globally-fused confidence value, $\hat{x}^l_{cf_{k+1|k}}$, by each confidence value assigned to C^i_{k+1}. The local gain is given by

$$\begin{aligned} K^{l^i}_{cf_{k+1}} &= P^l_{cf_{k+1|k}} H^{l^i}_{cf_{k+1}} [(H^{l^i}_{cf_{k+1}})^2 P^l_{cf_{k+1|k}} + R^{l^i}_{k+1}]^{-1} && (5.23) \\ &= P^{l^i}_{cf_{k+1|k+1}} H^{l^i}_{k+1} (R^{l^i}_{k+1})^{-1}. && (5.24) \end{aligned}$$

The global confidence value $\hat{x}^l_{cf_{k+1|k}}$ is locally updated by the local confidence values $z^{l^i}_{cf_{k+1}}$

$$\hat{x}^{l^i}_{cf_{k+1|k+1}} = (1 - K^{l^i}_{cf_{k+1}} H^{l^i}_{k+1}])\hat{x}^l_{cf_{k+1|k}} + K^{l^i}_{cf_{k+1}} (z^{l^i}_{cf_{k+1}} - m^{l^i}_{k+1}), \tag{5.25}$$

and the global error covariance is locally minimized by

$$(P^{l^i}{}_{cf_{k+1|k+1}})^{-1} = (P^{l}_{cf_{k+1|k}})^{-1} + (H^{l^i}{}_{k+1})^2(R^{l^i}{}_{k+1})^{-1}. \tag{5.26}$$

- Global fusion of N locally updated confidence values results in the globally updated confidence value:

$$\begin{aligned}\hat{x}^{l}_{k+1|k+1} = \; & P^{l}_{cf_{k+1|k+1}}[(P^{l}_{cf_{k+1|k}})^{-1}\hat{x}^{l}_{cf_{k+1|k}} + \sum_{i=1}^{N}((P^{l^i}{}_{cf_{k+1|k+1}})^{-1}\hat{x}^{l^i}_{cf_{k+1|k+1}} - \\ & (P^{l}_{cf_{k+1|k}})^{-1}\hat{x}^{l}_{cf_{k+1|k}})]\end{aligned} \tag{5.27}$$

$$\begin{aligned} = \; & P^{l}_{cf_{k+1|k+1}}[\sum_{i=1}^{N}(P^{l^i}{}_{cf_{k+1|k+1}})^{-1}\hat{x}^{l^i}_{cf_{k+1|k+1}} - \\ & (N-1)(P^{l}_{cf_{k+1|k}})^{-1}\hat{x}^{l}_{cf_{k+1|k}}],\end{aligned} \tag{5.28}$$

and the global updating yields a global minimum error covariance:

$$(P^{l}{}_{cf_{k+1|k+1}})^{-1} = (P^{l}{}_{cf_{k+1|k}})^{-1} + \sum_{i=1}^{N}[(P^{l^i}{}_{cf_{k+1|k}})^{-1} - (P^{l}{}_{cf_{k+1|k}})^{-1}] \tag{5.29}$$

$$= \sum_{i=1}^{N}(P^{l^i}{}_{cf_{k+1|k+1}})^{-1} - (N-1)(P^{l}{}_{cf_{k+1|K}})^{-1}. \tag{5.30}$$

A signal flow chart for the hierarchical fusion of confidence values is given in Figure 5.2. The detailed derivation of the algorithm is presented in Appendix A.4.

At the lowest level of resolution hierarchy (input image resolution level), $t = M$ and $l = I$, the matches with high confidence values are considered correct matches. The 3-D scene can then be recovered from these correct matches.

Discussion

The aforementioned algorithm for fusing confidence values is

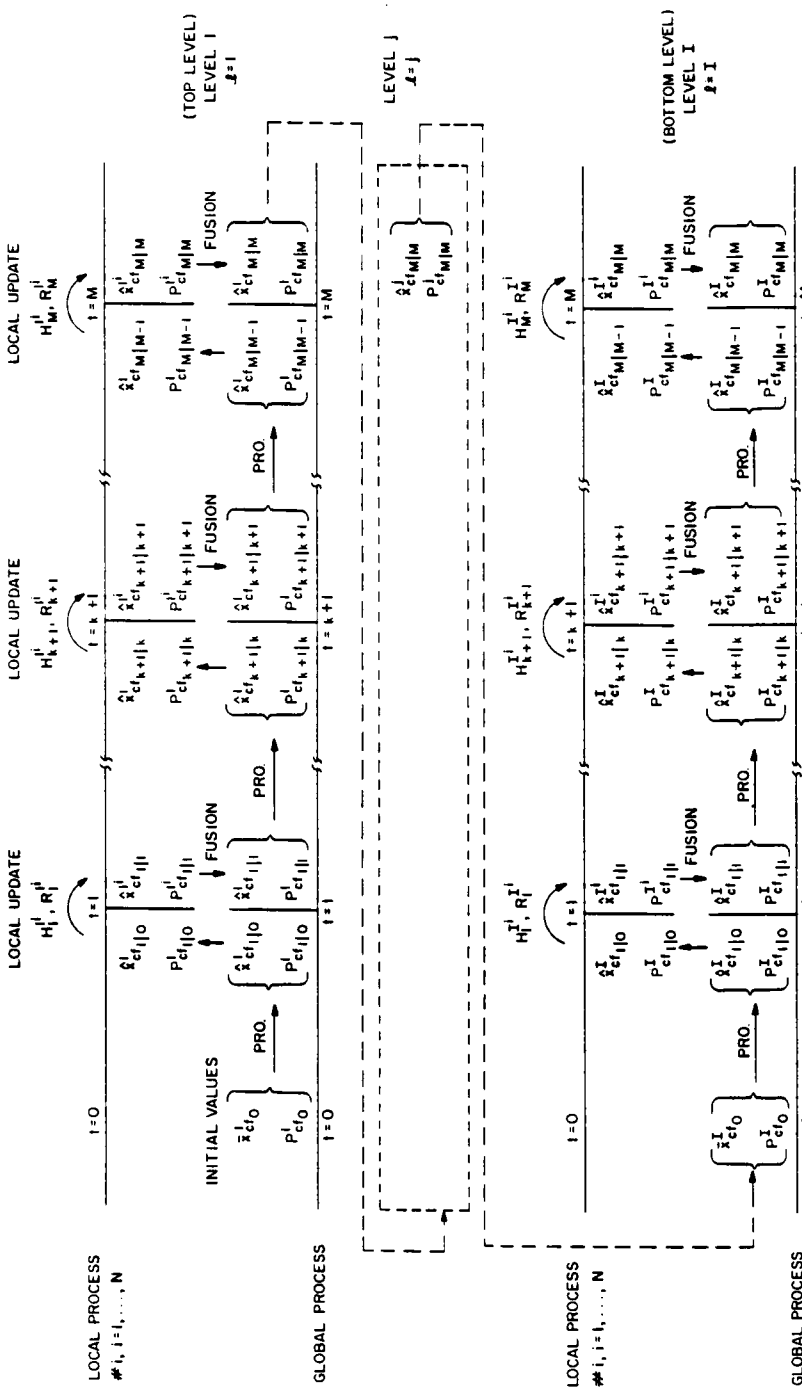


Figure 5.2: A signal flow chart for hierarchical fusion of confidence values.

- Optimal, since the algorithm minimizes local error covariances, Equation (5.25), and global error covariances, Equation (5.29)
- Computationally efficient, since N local updates, Equation (5.24), can be performed in parallel, and the dimension of the global fusion process, Equation (5.26), is low
- Capable of dynamic fusion, since the algorithm continuously updates the local and global confidence values when new measurements are available

5.3.2 An Algorithm for 3-D Information Fusion

Fusion of 3-D information is performed at the lowest level of the resolution hierarchy. Since at higher levels, the spheres derived at the current moment are mapped onto the next image pair to predict the areas for matching in the next image pair, no true 3-D points are needed. At the lowest level, true 3-D points in the spheres are desired. Therefore, fusing the 3-D information obtained at different positions and different times reduces the spatial uncertainties in calculating the true 3-D points.

The initial conditions of the algorithm are

$$\underline{X}_{0|-1} = \underline{X}_0, \quad \mathbf{P}_{0|-1} = \mathbf{P}_0, \tag{5.31}$$

where $\underline{X}_0$ is the center of sphere C_k, and the elements of the diagonal matrix $\mathbf{P}_0$ are proportional to the radius of C_k:

$$\mathbf{P}_0 = diag\{(r_k/q)^2, (r_k/q)^2, (r_k/q)^2\}. \tag{5.32}$$

The constant q, defined by Equation (4.11), depends on the specified probability that the true 3-D point lies in the sphere C_k. From kth moment to $(k+1)th$ moment, the distributed model-based algorithm for 3-D information fusion incorporates the following steps:

- Propagation of previously fused 3-D information, $\underline{\hat{X}}_{k|k}$, and the associated error covariance, $\mathbf{P}_{k|k}$, from t_k to t_{k+1}

$$\underline{\hat{X}}_{k+1|k} = \mathbf{J}_k \underline{\hat{X}}_{k|k} + \underline{T}_k, \tag{5.33}$$

$$\mathbf{P}_{k+1|k} = \mathbf{J}_{0k} \mathbf{P}_{k|k} \mathbf{J}_{0k}^T + \mathbf{J}_{a_k} \mathbf{Q}_{J_k} \mathbf{J}_{a_k}^T + \mathbf{Q}_{T_k}, \tag{5.34}$$

where

$$\mathbf{J}_k = \mathbf{J}_{0k} + \Delta \mathbf{J}_k, \quad \Delta \mathbf{J}_k \sim N(0, \mathbf{Q}_{J_k}), \tag{5.35}$$

$$\underline{T}_k = \underline{T}_{0k} + \Delta \underline{T}_k, \quad \Delta \underline{T}_k \sim N(0, \mathbf{Q}_{T_k}), \tag{5.36}$$

and $\mathbf{J}_{a_k}$ is a Jacobian given by Equation (4.33).

- Local update of the propagated global 3-D information. For the $ith, i = 1, ..., N$ local process, the local gain is calculated by

$$\mathbf{K}_{k+1}^i = \mathbf{P}_{k+1|k}^i [\mathbf{P}_{k+1|k}^i + \mathbf{R}_{k+1}^i]^{-1} \tag{5.37}$$

$$= \mathbf{P}_{k+1|k+1}^i (\mathbf{R}_{k+1}^i)^{-1}. \tag{5.38}$$

The kth moment's global 3-D information is locally updated in parallel in N local processes by local measurements $\underline{Z}_{k+1}^i$:

$$\underline{\hat{X}}_{k+1|k+1}^i = [\mathbf{I} - \mathbf{K}_{k+1}^i] \underline{\hat{X}}_{k+1|k} + \mathbf{K}_{k+1}^i \underline{Z}_{k+1}^i. \tag{5.39}$$

The associated local error covariance is minimized by

$$(\mathbf{P}^i_{k+1|k+1})^{-1} = (\mathbf{P}^i_{k+1|k})^{-1} + (\mathbf{R}^i_{k+1})^{-1}. \tag{5.40}$$

- Global fusion of 3-D information at the $(k+1)th$ moment is obtained by fusing N local 3-D information updates:

$$\begin{aligned} \underline{\hat{X}}_{k+1|k+1} = \; & \mathbf{P}_{k+1|k+1}[\mathbf{P}^{-1}_{k+1|k}\underline{\hat{x}}_{k+1|k} + \sum_{i=1}^{N}((\mathbf{P}^i_{k+1|k+1})^{-1}\underline{\hat{X}}^i_{k+1|k+1} - \\ & \mathbf{P}^{-1}_{k+1|k}\underline{\hat{X}}_{k+1|k})] \end{aligned} \tag{5.41}$$

$$\begin{aligned} = \; & \mathbf{P}_{k+1|k+1}[\sum_{i=1}^{N}(\mathbf{P}^i_{k+1|k+1})^{-1}\underline{\hat{X}}^i_{k+1|k+1} - \\ & (N-1)\mathbf{P}^{-1}_{k+1|k}\underline{\hat{X}}_{k+1|k}], \end{aligned} \tag{5.42}$$

and the update equation for the associated minimum global error covariance is

$$\mathbf{P}^{-1}_{k+1|k+1} = \mathbf{P}^{-1}_{k+1|k} + \sum_{i=1}^{N}((\mathbf{P}^i_{k+1|k})^{-1} - \mathbf{P}^{-1}_{k+1|k}) \tag{5.43}$$

$$= \sum_{i=1}^{N}(\mathbf{P}^i_{k+1|k+1})^{-1} - (N-1)\mathbf{P}^{-1}_{k+1|K}. \tag{5.44}$$

A signal flow chart for the 3-D information fusion is shown in Figure 5.3. The algorithm for 3-D information fusion shares the advantages of the algorithm for confidence value fusion, listed in the discussion on Page 87.

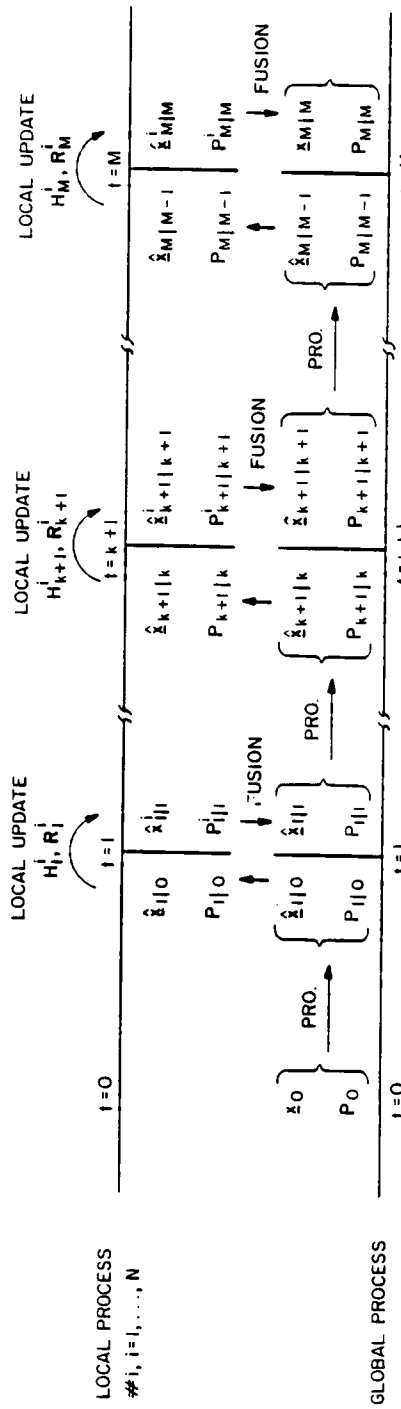


Figure 5.3: A signal flow chart for 3-D information fusion.

CHAPTER 6

EXPERIMENTAL RESULTS

The vision system developed in this work has been applied to several binocular noisy image sequences, and the results obtained are very satisfactory. The task of 3-D scene reconstruction consists of 3 stages: image acquisition, which is described in Section 6.1; preprocessing, presented in Section 6.2 in which features are extracted and the resolution hierarchy is formed; and matching and information fusion, which give the final 3-D information about the scene and are the subjects of Section 6.3.

6.1 Acquisition of Binocular Image Sequences

Images are acquired by using a Fairchild CCD-3000 camera with a lens focus length 1.3 *cm* mounted on the Cincinnati Milacron $T^3 - 726$ industrial robot with 6 degrees of freedom, as shown in Figure 6.1. The resolution of input images in this work is 256 $\times$ 256 pixels. The sequences of binocular images are obtained in such a way that the camera moves along its optical axis towards the objects and takes pictures sequentially in two parallel paths. The distance between the two paths (the length of baseline b) is 3 *inches*, and the incremental movement of the camera is 1.5 *inches*. Therefore, the motion parameters of the equivalent

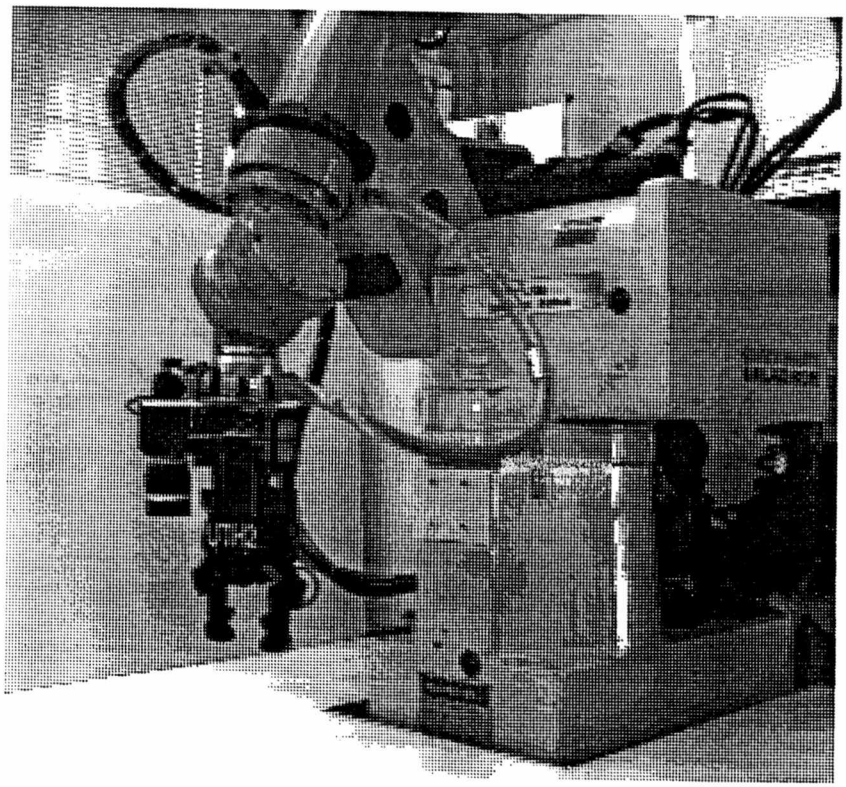


Figure 6.1: A Cincinnati Milacron $T^3 - 726$ industrial robot with 6 degrees of freedom and a CCD camera mounted on the robot.

binocular cameras are

$$\mathbf{J}_k = \mathbf{I} + \Delta\mathbf{J}_k,\ \Delta\mathbf{J}_k \sim N(0, \mathbf{Q}_{J_k}), \tag{6.1}$$

$$\underline{T}_k = [0, 0, 1.5]^T + \Delta\underline{T}_k,\ \Delta\underline{T}_k \sim N(0, \mathbf{Q}_{T_k}), \tag{6.2}$$

where $\mathbf{I}$ is an identity matrix, and $\Delta\mathbf{J}_k$ and $\Delta\underline{T}_k$ represent the uncertainties of the motion parameters characterized by $\mathbf{Q}_{J_k}$ and $\mathbf{Q}_{T_k}$. Based on experience and the specifications of the robot, the values of the error covariances $\mathbf{Q}_{J_k}$ and $\mathbf{Q}_{T_k}$ are chosen as

$$\mathbf{Q}_{J_k} = diag\{3, 3, 3, 3, 3, 3, 3, 3, 3\} \tag{6.3}$$

and

$$\mathbf{Q}_{T_k} = diag\{5, 5, 5\}, \tag{6.4}$$

assuming that the elements of $\mathbf{J}_k$ and $\underline{T}_k$ are independent of each other. The larger the values of these error covariances, the larger the uncertainties in the motion parameters to be incorporated. But the larger the values of these error covariances, the slower the convergence rate in the fusion process. The proposed algorithms have been tested on two binocular sequences, each of which has ten pairs of images. For convenience, hereafter, notation for the first set of binocular image sequences will be called $\mathcal{S}_1$, and the second set will be called $\mathcal{S}_2$. The left and right images in the *ith* pair of $\mathcal{S}_k$, $k = 1, 2$, are denoted as $L_{i\mathcal{S}_k}$ and $R_{i\mathcal{S}_k}$. The first two image pairs of $\mathcal{S}_1$ and $\mathcal{S}_2$ are shown in Figures 6.2 and Figure 6.3. The noisy image sequences are obtained by adding Gaussian white noise to the images, as shown in Figure 6.4, where $\sigma = 15$ and in Figure 6.5, where $\sigma = 20$.

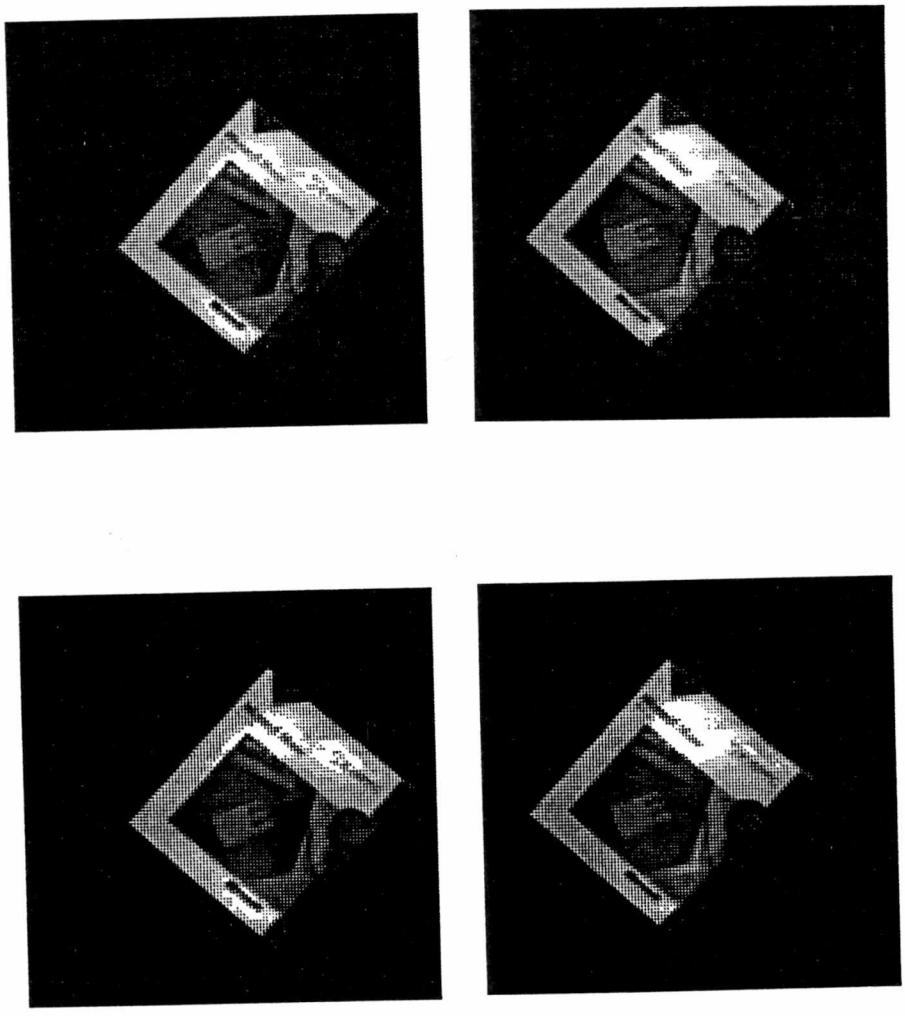


Figure 6.2: The first two noise free image pairs of $\mathcal{S}_1$. Upper: the first pair, and lower: the second pair.

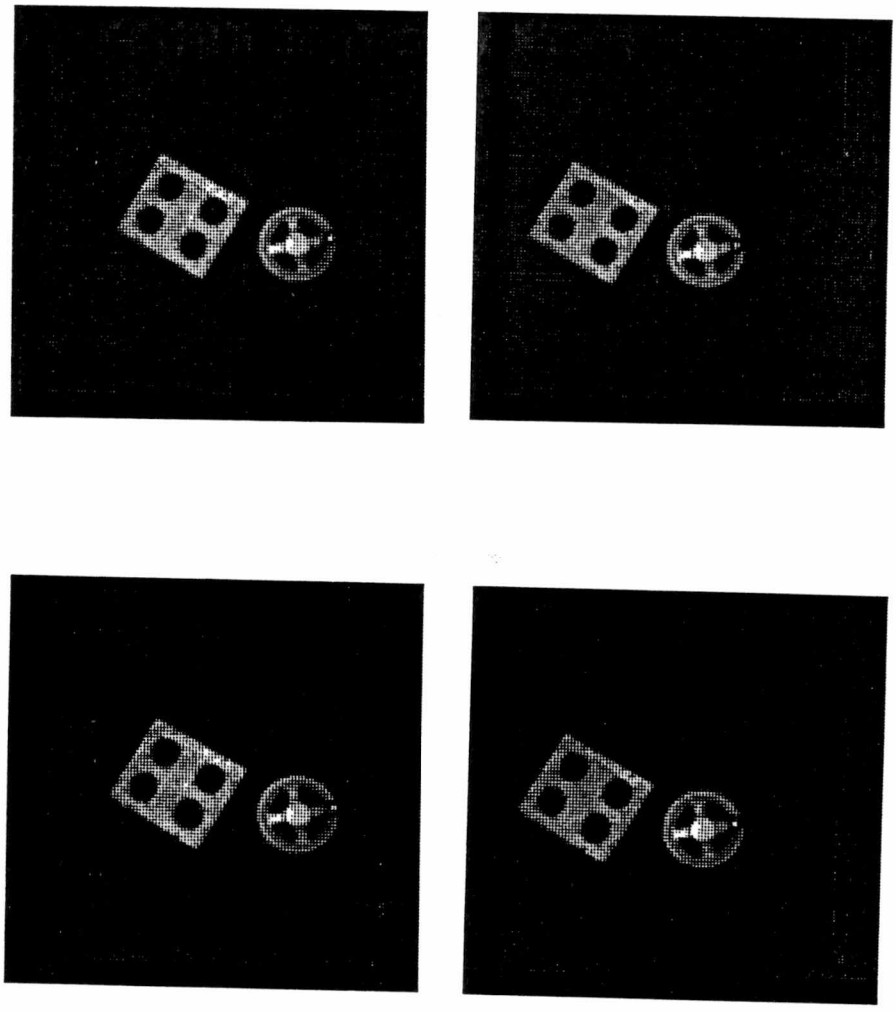


Figure 6.3: The first two noise free image pairs of $\mathcal{S}_2$. Upper: the first pair, and lower: the second pair.

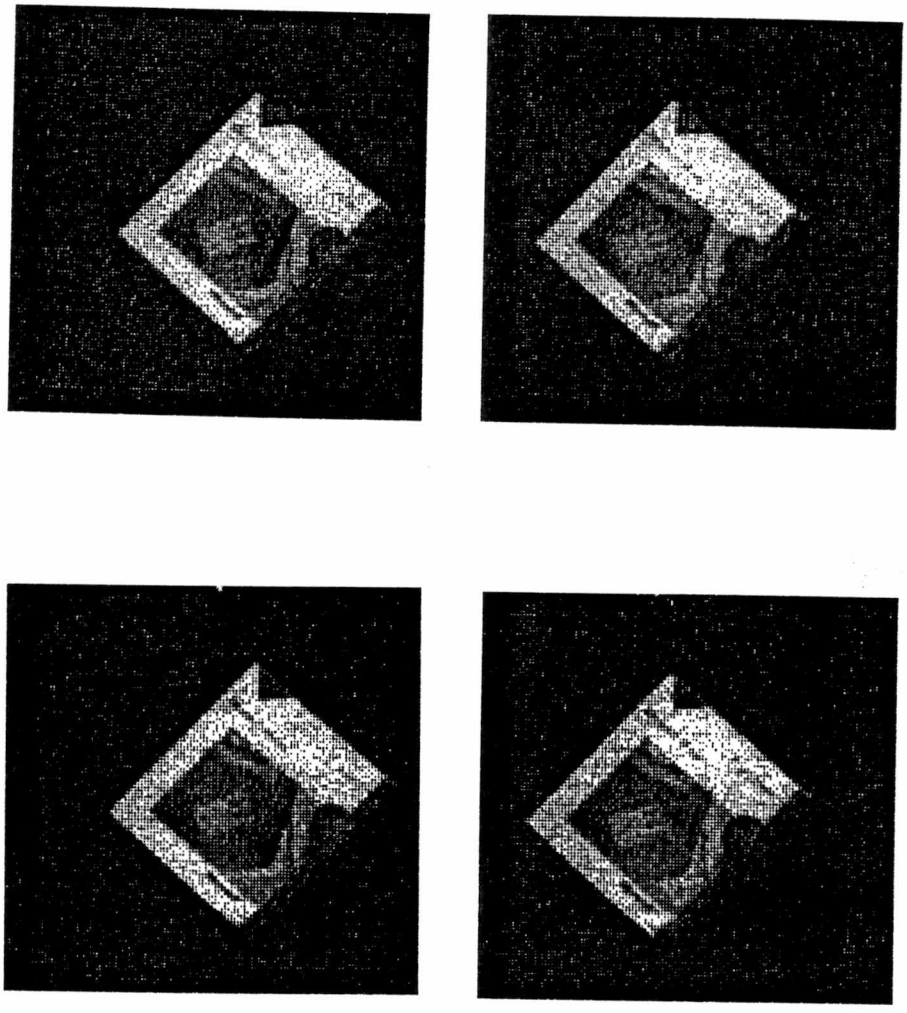


Figure 6.4: The first two noisy image pairs of $\mathcal{S}_1$ ($\sigma = 15$). Upper: the first pair, and lower: the second pair.

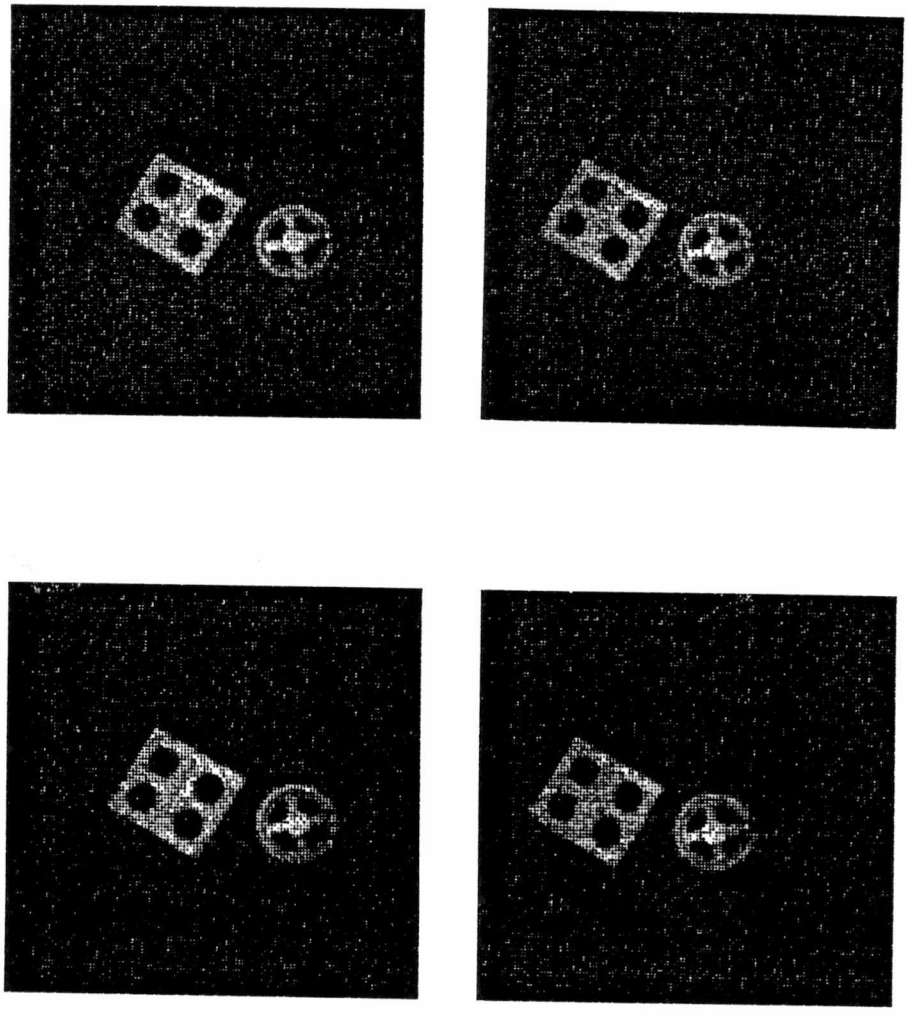


Figure 6.5: The first two noisy image pairs of $\mathcal{S}_2$ ($\sigma = 20$). Upper: the first pair, and lower: the second pair.

6.2 Preprocessing

The inputs of the vision system are sequences of noisy images. Preprocessing is first performed to extract the raw information from the images. Pyramid structures of the noisy images are constructed first; then the features of images at each level are extracted. Pyramid structures with 2 levels and 3 levels are both used and compared. Two 3-level pyramids of signs of zero-crossings of the noisy images $L_{1\mathcal{S}_1}$ and $L_{1\mathcal{S}_2}$ are shown in Figures 6.6 and 6.7, where the positive zero-crossings are in red and the negative zero-crossings are in green. Zero-crossings are obtained by convolving images with the Laplacian of Gaussian with $\sigma = 1.5$, for the purpose of compromising between retaining the details of the images and smoothing the noise in the images. The strengths and signs of the zero-crossings are then extracted. Figures 6.8 and 6.9 show the strength and sign maps with resolution 256×256 of the noisy image $L_{1\mathcal{S}_1}$, and the strength and sign maps with resolution 256×256 of the noisy image $L_{1\mathcal{S}_2}$ are shown in Figures 6.10 and 6.11. The positive zero-crossings in the sign maps, shown in Figures 6.9 and 6.11, are in white and the negative zero-crossings are in gray.

6.3 Matching, Assignment of Confidence Values and Information Fusion

Matching begins with the first image pair at the top level of the hierarchy. To reconstruct the 3-D scene, the vision system scans, row by row, the sign map

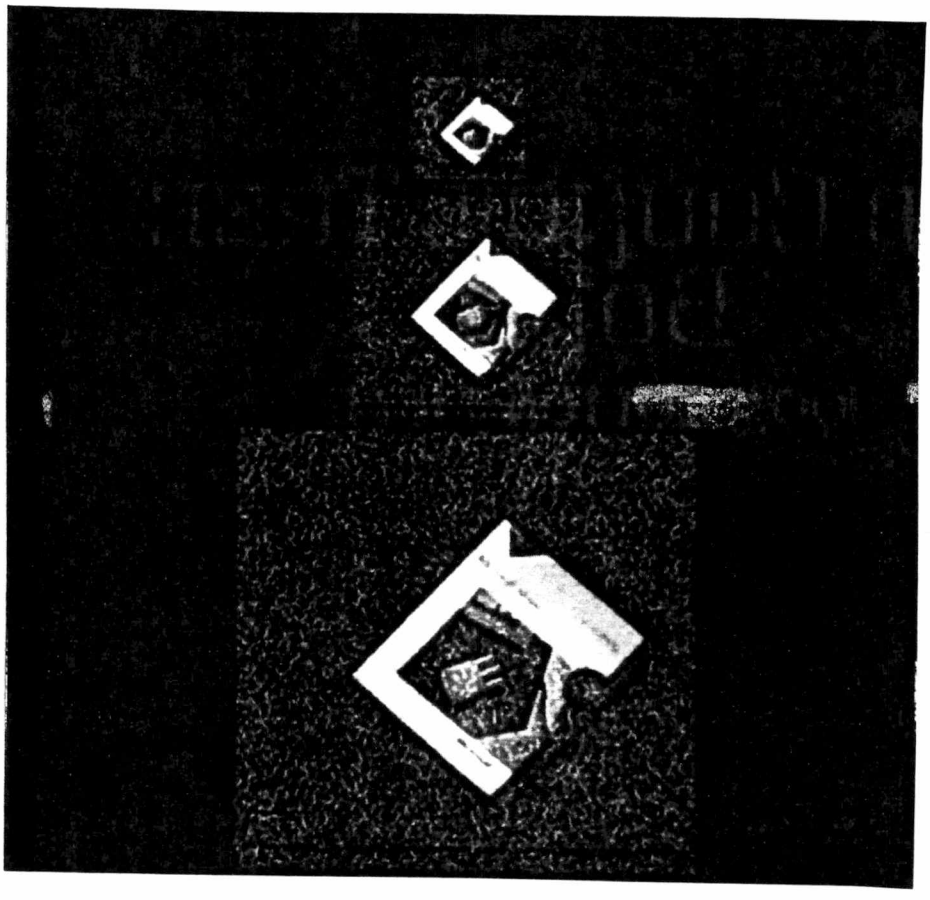


Figure 6.6: A 3-level pyramid of signs of zero-crossing of noisy image L_{1S_1}.

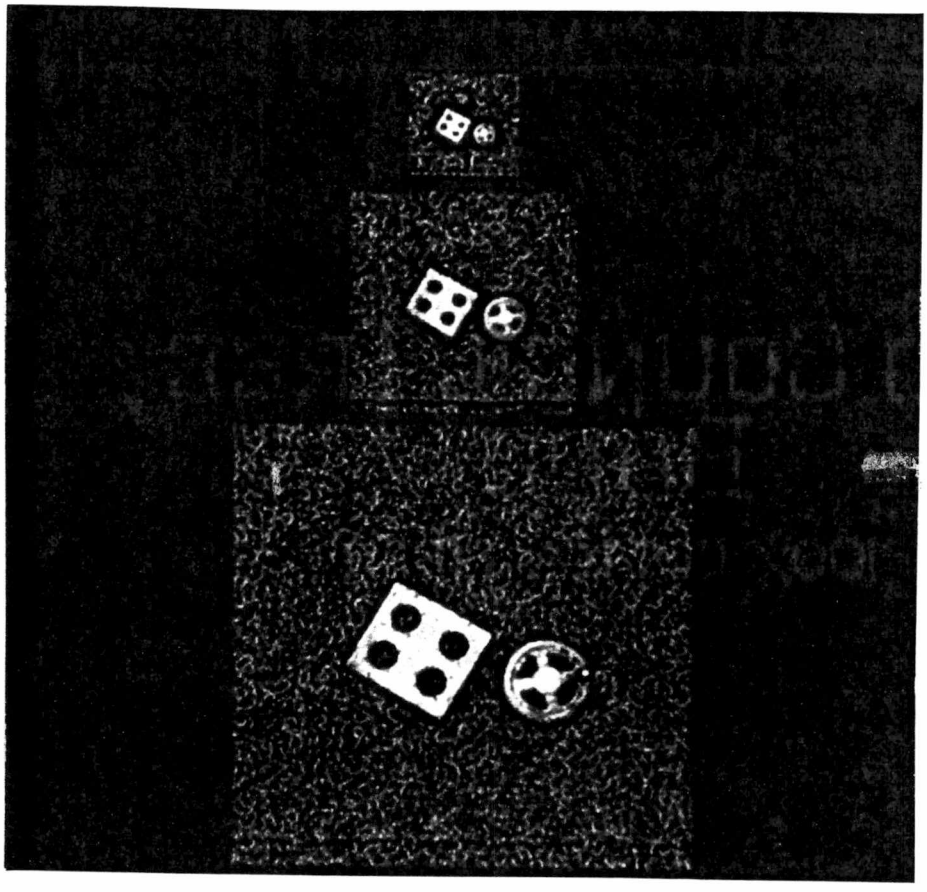


Figure 6.7: A 3-level pyramid of signs of zero-crossing of noisy image L_{1S_2}.

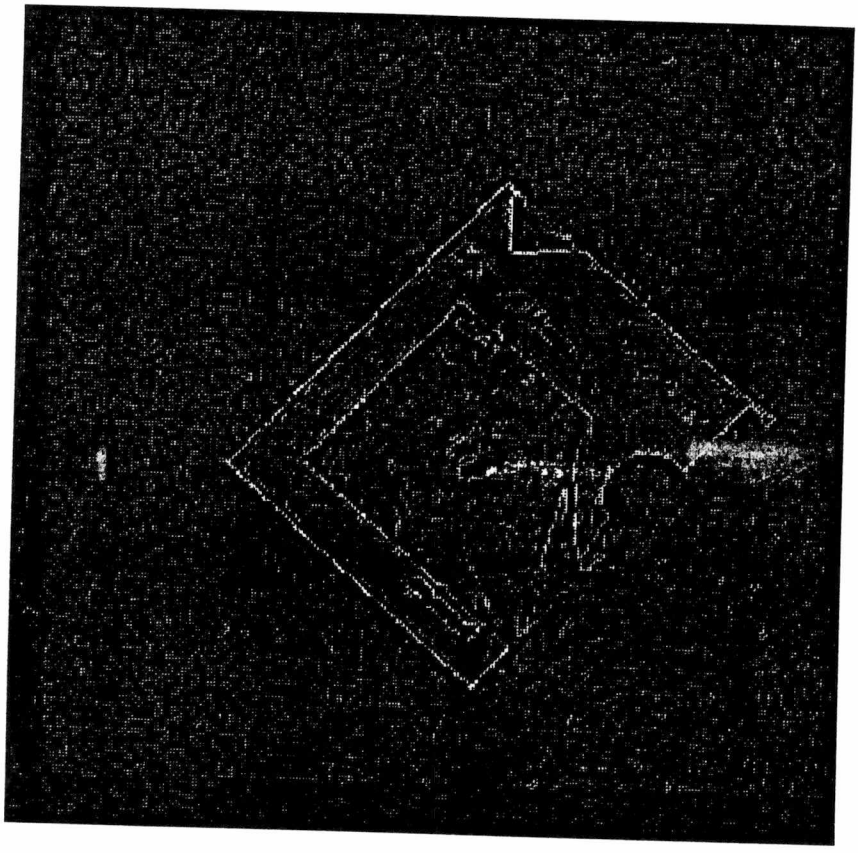


Figure 6.8: The strength map of the noisy image L_{1S_1}.

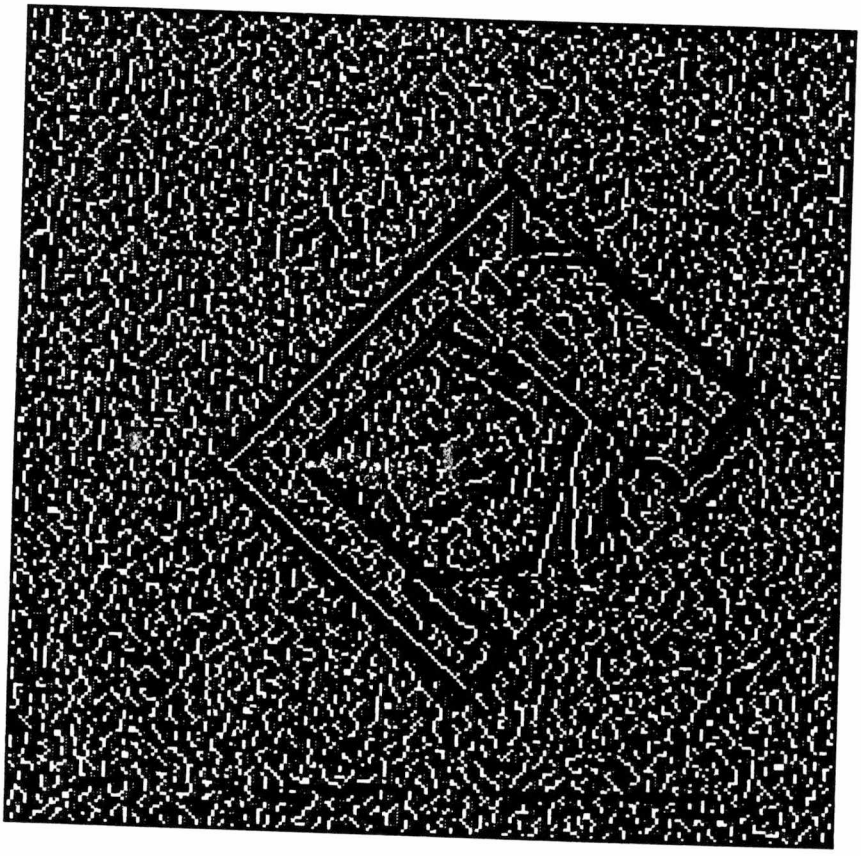


Figure 6.9: The sign map of the noisy image $L_{1\mathcal{S}_1}$.

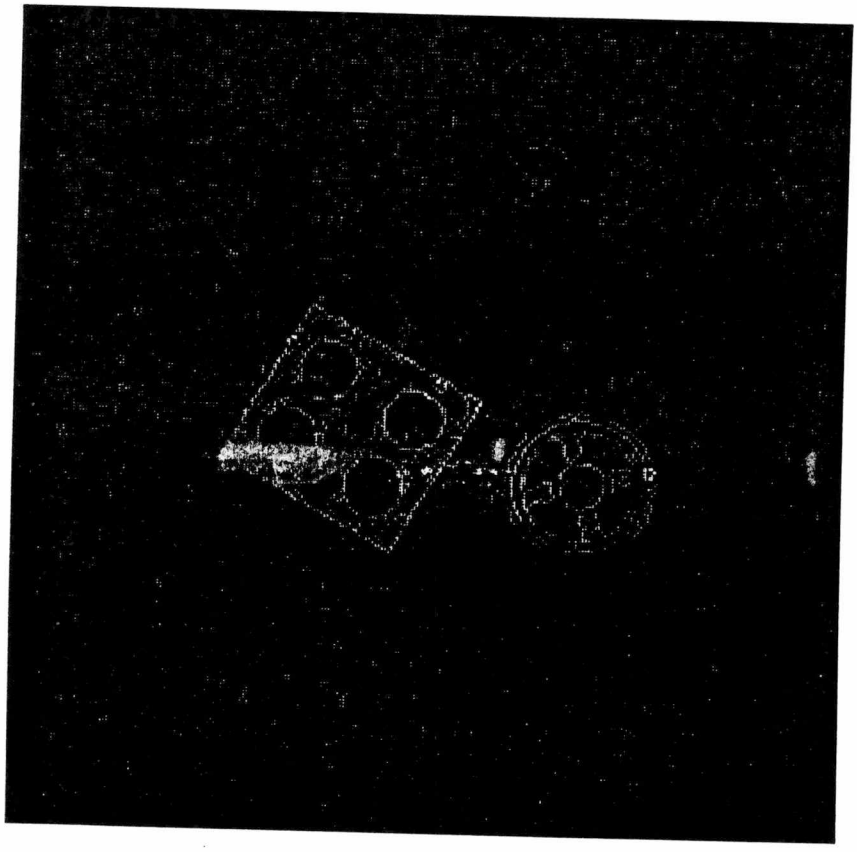


Figure 6.10: The strength map of the noisy image L_{1S_2}.

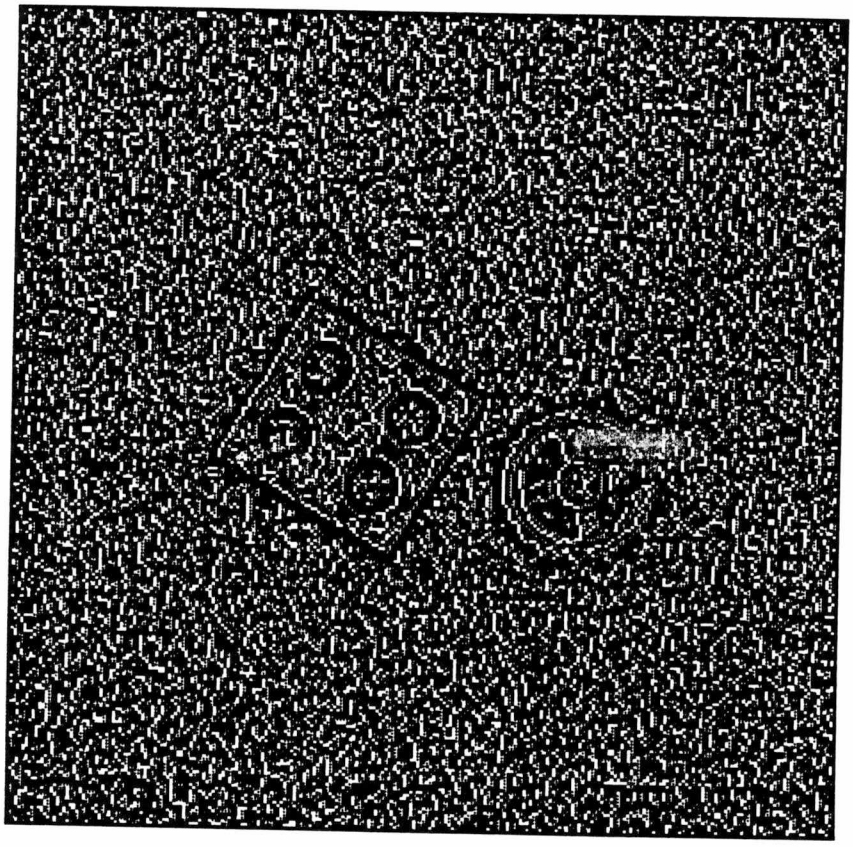


Figure 6.11: The sign map of the noisy image $L_{1\mathcal{S}_2}$.

of the left image of the first pair $L_{1\mathcal{S}_k}$, $k = 1, 2$, at the top level. The points on the zero-crossings with positive (or negative) signs are taken as candidate points for matching. For every candidate point in $L_{1\mathcal{S}_k}$, $k = 1, 2$, matching is performed within the window W in the $R_{1\mathcal{S}_k}$, $k = 1, 2$. Since the minimum distance between cameras and scenes in two sets of sequences, $\mathcal{S}_1$ and $\mathcal{S}_2$, is 40 cm, the window size is 0.061912 cm for 3-level pyramids and 0.123824 cm for 2-level pyramids, as calculated by Equation (3.4). In *pixels*, the window size is calculated by

$$W_{pixel} = W_{cm}/a, \tag{6.5}$$

where the factor a is derived by camera calibration. The window size is 14 *pixels* for the 3-level pyramids and 28 *pixels* for 2-level pyramids. Possible matches in the first pair of images at the top level are found based on the signs, the strengths, and the patterns of the zero-crossings, as described in Section 3.2. The thresholds used for these two sets of sequences are

$$T^P_{S_{zc}} = T^S_{S_{zc}} = 30, \quad T^P_{P_{zc}} = 3 \quad \text{and} \quad T^S_{P_{zc}} = 4, \tag{6.6}$$

where $T^P_{S_{zc}}$ and $T^S_{S_{zc}}$ are the thresholds for pairwise and sequencewise strength differences, and $T^P_{P_{zc}}$ and $T^S_{P_{zc}}$ are thresholds for pairwise and sequencewise pattern differences, as used in Equations (3.24), (3.25), (3.26), (3.27), (3.28) and (3.29). Confidence values are assigned to the possible matches found in the first pair at the top level by using Equation (3.20) with the coefficients chosen as

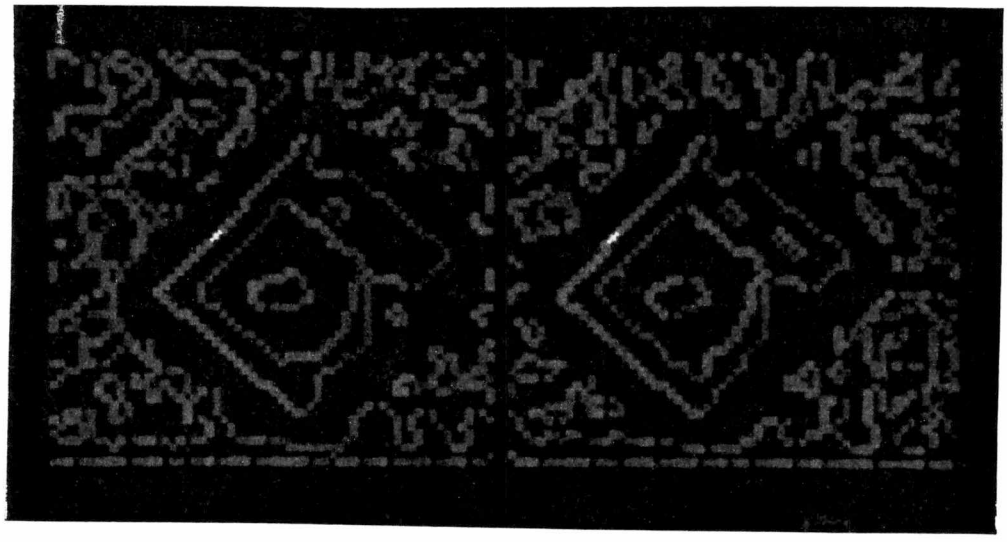
$$\alpha_1 = 0.3, \quad \alpha_2 = 0.4, \text{ and } \alpha_3 = 0.3.$$

The confidence values are then updated by the information from the sequences. Matching in the second pair at the top level is done in the regions predicted by the possible matches found in the first pair. Figure 6.12 (a) illustrates in green the possible matches found in the first pair and Figure 6.12 (b) demonstrates in green the regions predicted in the second pair at the top level. Regions predicted in the first image pairs at other levels by the possible matches found in the upper level are shown in Figure 6.13, where the predicted regions are in green. Confidence values of possible matches found in all image pairs, except the first pair at the top level, are calculated by Equation (3.30) using the coefficients

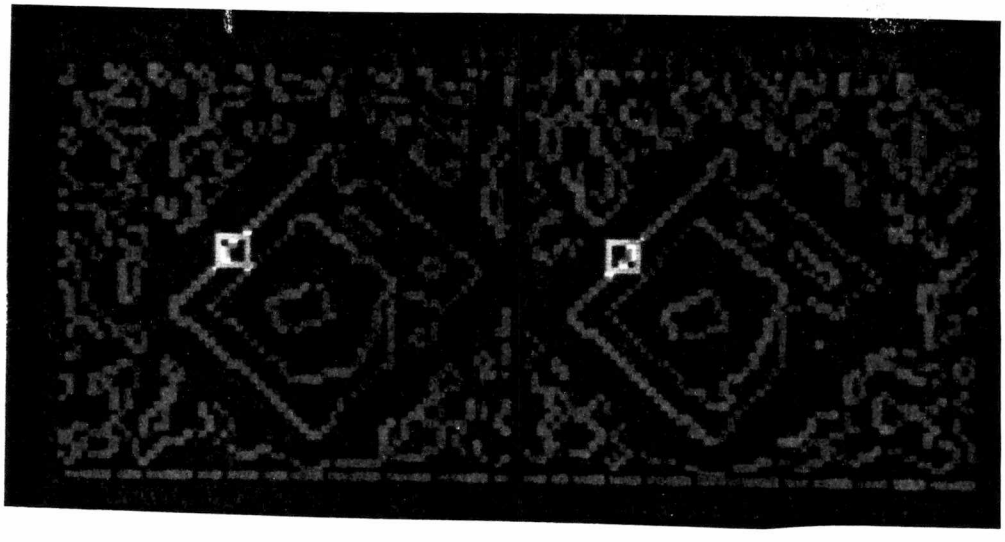
$$\gamma_1 = 0.5 \quad \text{and} \quad \gamma_2 = 0.5.$$

To make use of the information from its neighborhood, a confidence value of a possible match is refined by a weighted average of confidence values over its neighborhood of the zero-crossing. The range ΔL for averaging, adopted here, is 3 *pixels*, Figure 3.7.

The confidence values are updated by the dynamic, optimal information fusion algorithm, in which the initial values for the state variables are the confidence values of the first pair at the top level, and the initial value for the error covariance is $P_0 = 10$. The convergence of the fusion process is reflected by the convergence of the error covariance of the process. A few curves of error covariances of the possible matches, shown in Figure 6.14, demonstrate the fast convergence of the fusion process. The system converges after 3-4 iterations. In Figure 6.14, one can see clearly the behavior of the fusion process moving down from the top level



(a)



(b)

Figure 6.12: Possible matches found in the first pair of $\mathcal{S}_1$ and the regions predicted in the second pair at top level. (a) Possible matches found in the first image pair, and (b) regions predicted in the second image pair.

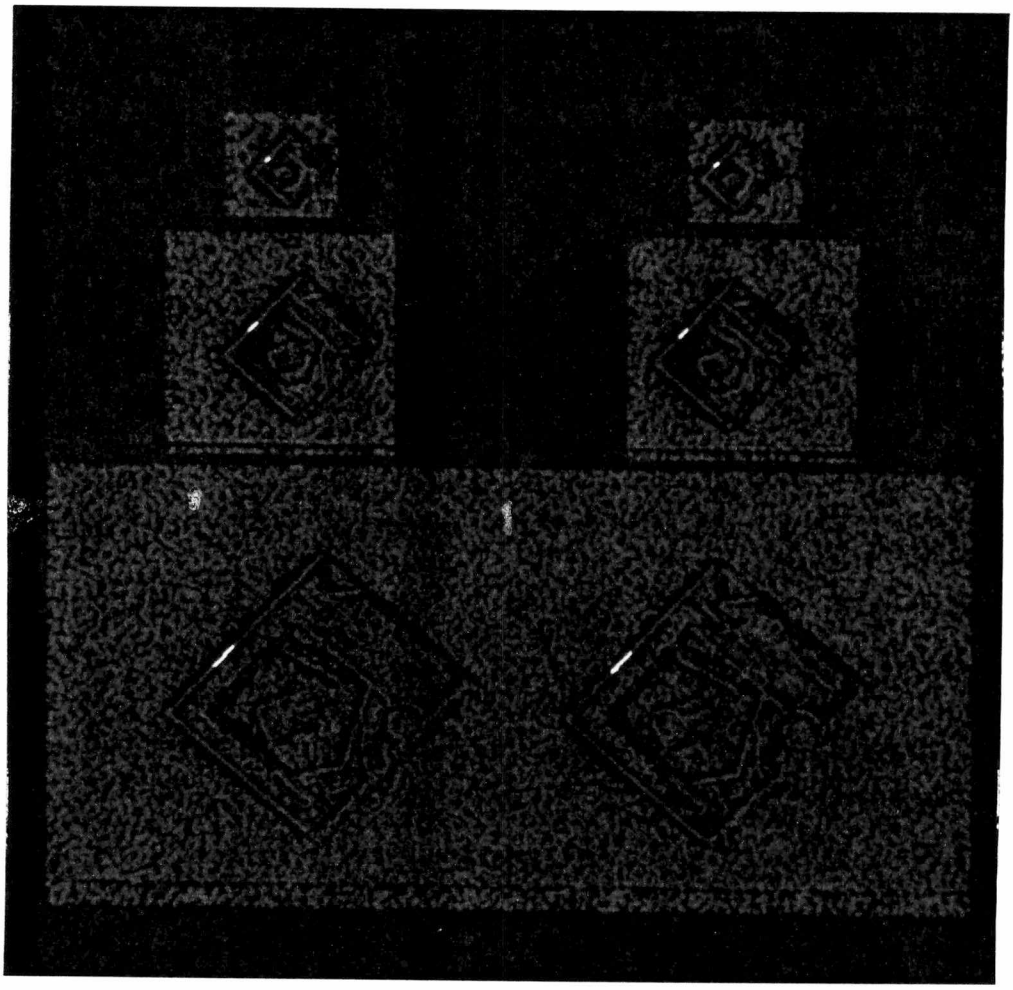


Figure 6.13: Regions in the first image pairs of $\mathcal{S}_1$ predicted by the possible matches found in the first image pair at the top level.

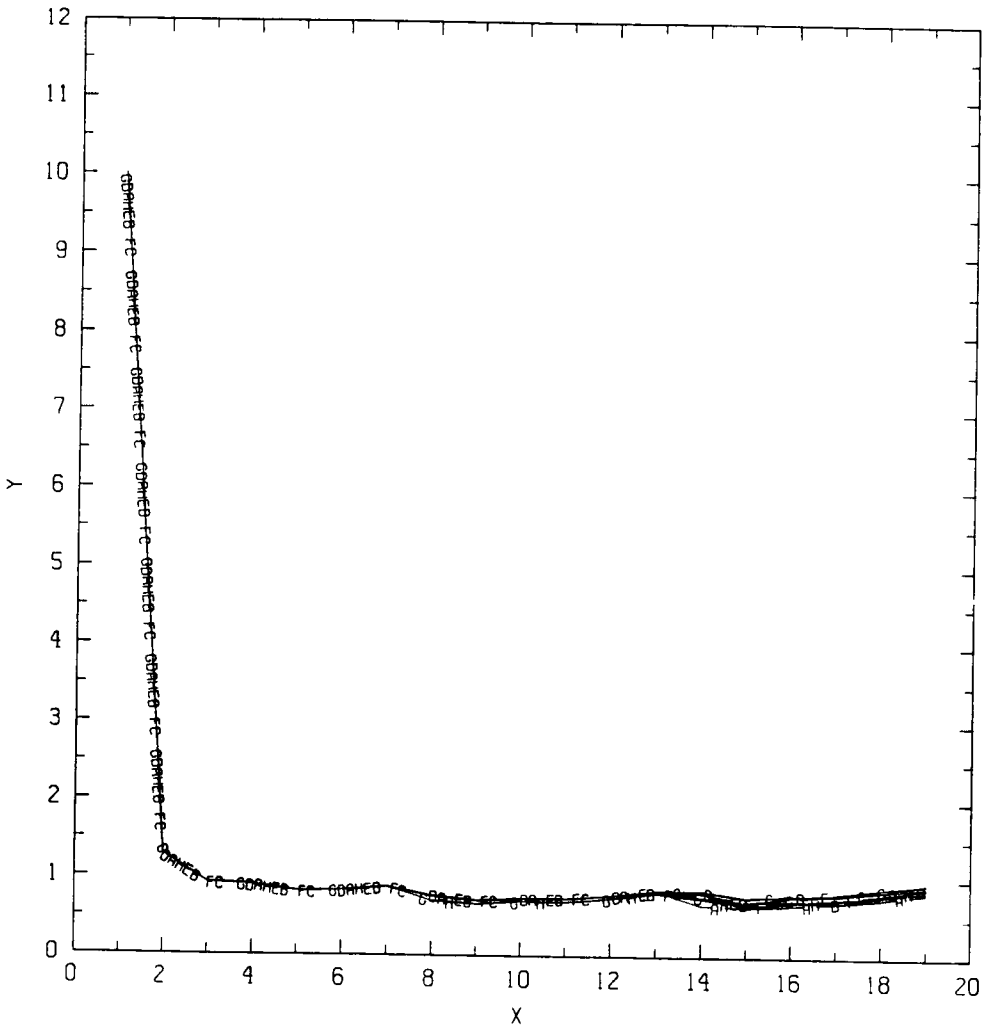


Figure 6.14: Error covariances of the fusion process.

to the bottom level of the hierarchy which has 3 levels and 7 image pairs for each level. Figure 6.15 shows the update process of the confidence values of a *correct* match found in the first pair at the top level and Figure 6.16 shows the update process of the confidence values of a *false* match found in the first pair at the top level. Figures 6.15 and 6.16 are typical examples of confidence value updates. As expected, the confidence values of the correct matches do not decrease; however the confidence values of the false matches decrease in the fusion process. The increasing confidence values at the ends of some curves, shown in Figure 6.16, are due to the stronger effect of the noise at the lower levels. But the increase is much lower than the threshold which tells if the possible matches are correct or false. Since the sizes of the images at the lower level are larger than those at the higher level, one possible match at the top level is expanded into several possible matches at the bottom level. This is shown in Figures 6.14, 6.15 and 6.16. After fusion of the confidence values, the points in the first pair at the bottom level, whose confidence values are greater than a threshold (0.6), are considered to be correct matches. Figures 6.17 and 6.18 present the correct matches found for the sequences $\mathcal{S}_1$ and $\mathcal{S}_2$ using a 3-level hierarchy and Figures 6.19 and 6.20 show the correct matches found for the sequences $\mathcal{S}_1$ and $\mathcal{S}_2$ using a 2-level hierarchy. As expected, one can see that the fewer levels in the pyramid, the more details are matched, but the more mismatches are also found due to the stringer effect of noise. Overall, the percentage of the true matches over the correct matches found is more than 95 %.

The 3-D information calculated from the correct matches in the first pair at

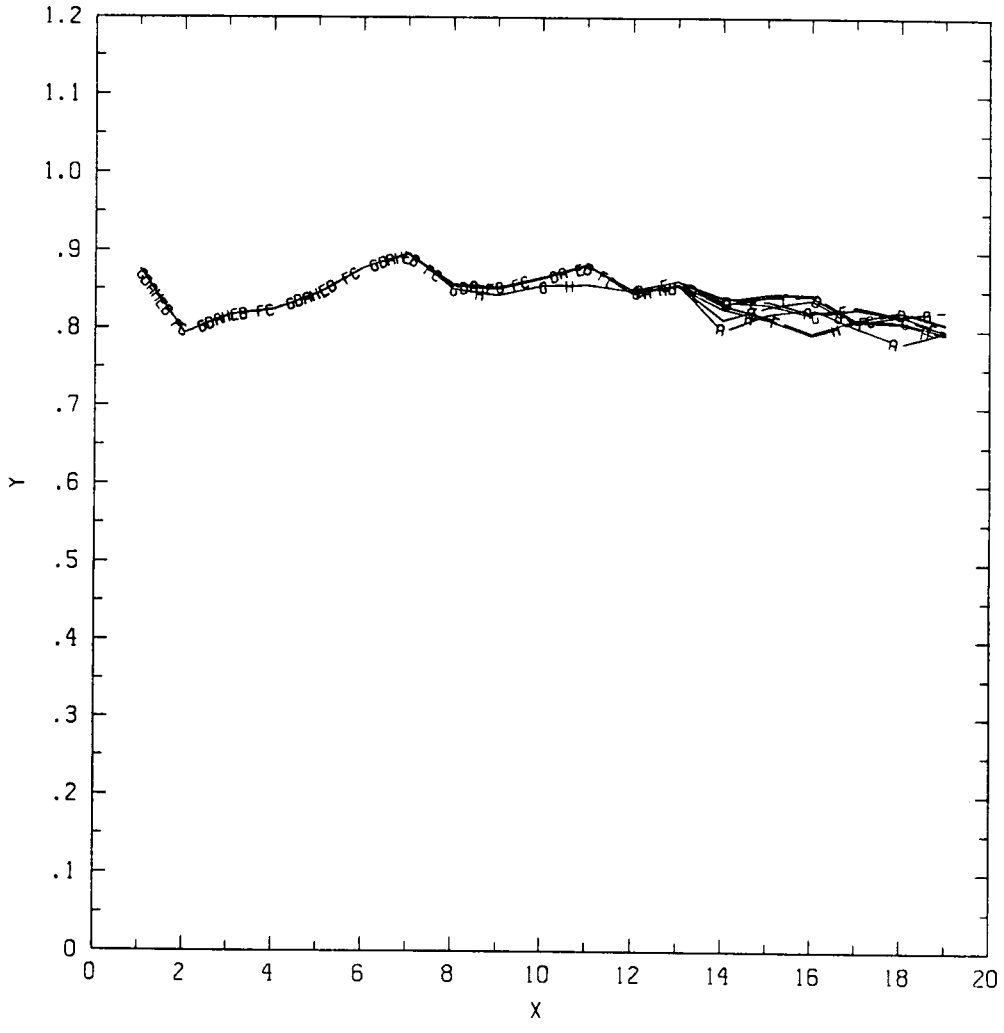


Figure 6.15: Confidence values as a function of time for a correct match in the first image pair at the top level.

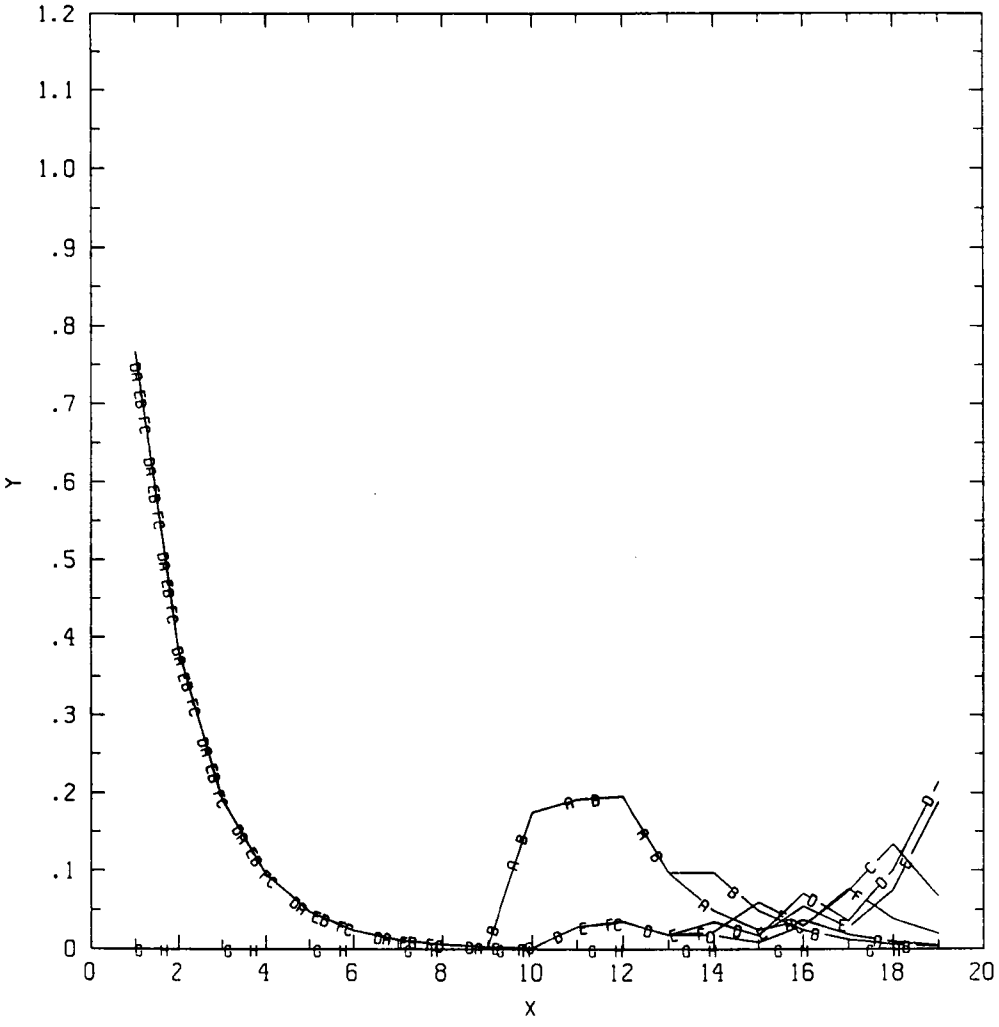


Figure 6.16: Confidence values as a function of time for a false match in the first image pair at the top level.

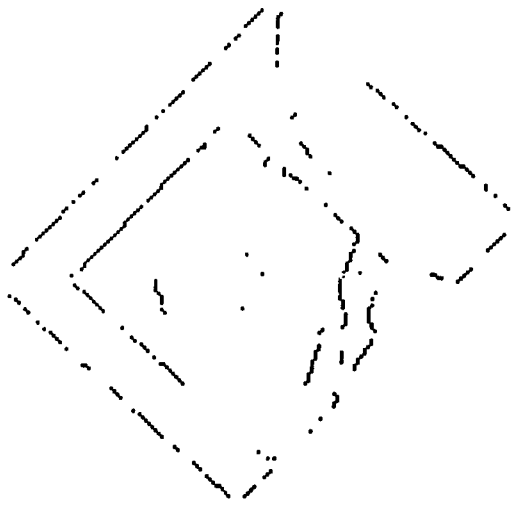


Figure 6.17: Correct matches found for $\mathcal{S}_1$ using a 3-level hierarchy.

Figure 6.18: Correct matches found for $\mathcal{S}_2$ using a 3-level hierarchy.

Figure 6.19: Correct matches found for $\mathcal{S}_1$ using a 2-level hierarchy.

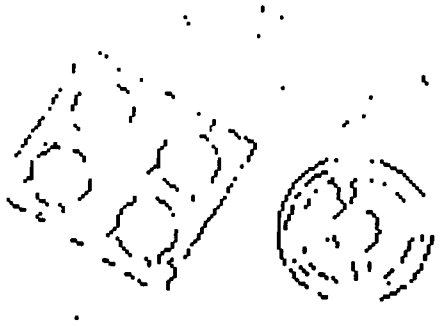


Figure 6.20: Correct matches found for $\mathcal{S}_2$ using a 2-level hierarchy.

the bottom level is fused with 3-D information from the second to Mth pairs at bottom level to reduce the uncertainties. A curve showing the reduction of the uncertainty is plotted in Figure 6.21. Table 1 lists fused 3-D distances of some selected points of sequences $\mathcal{S}_1$, shown in Figure 6.22, together with ground data of the selected points. The average of errors in computed Z distances for sequences $\mathcal{S}_1$ is about 2.55 %, which is very good for such noisy images. The fused 3-D distances of some selected points of sequences $\mathcal{S}_2$, shown in Figure 6.23, and the ground data are listed in Table 2. The average of errors in computed Z distances for sequences $\mathcal{S}_2$ is about 3.11 %. The reconstructed objects in sequences $\mathcal{S}_1$ and $\mathcal{S}_2$, shown in Figures 6.24 and 6.25, are obtained by using moving least square surface interpolation over sparse fused 3-D points.

The vision system is computationally efficient, although it is implemented on VAX 11/785 in a sequential manner. However, it can be parallelized, since the fusion process of one confidence value is independent of those of others. The CPU time for one fusion process at one level is less than 1 second of CPU time on VAX 11/785.

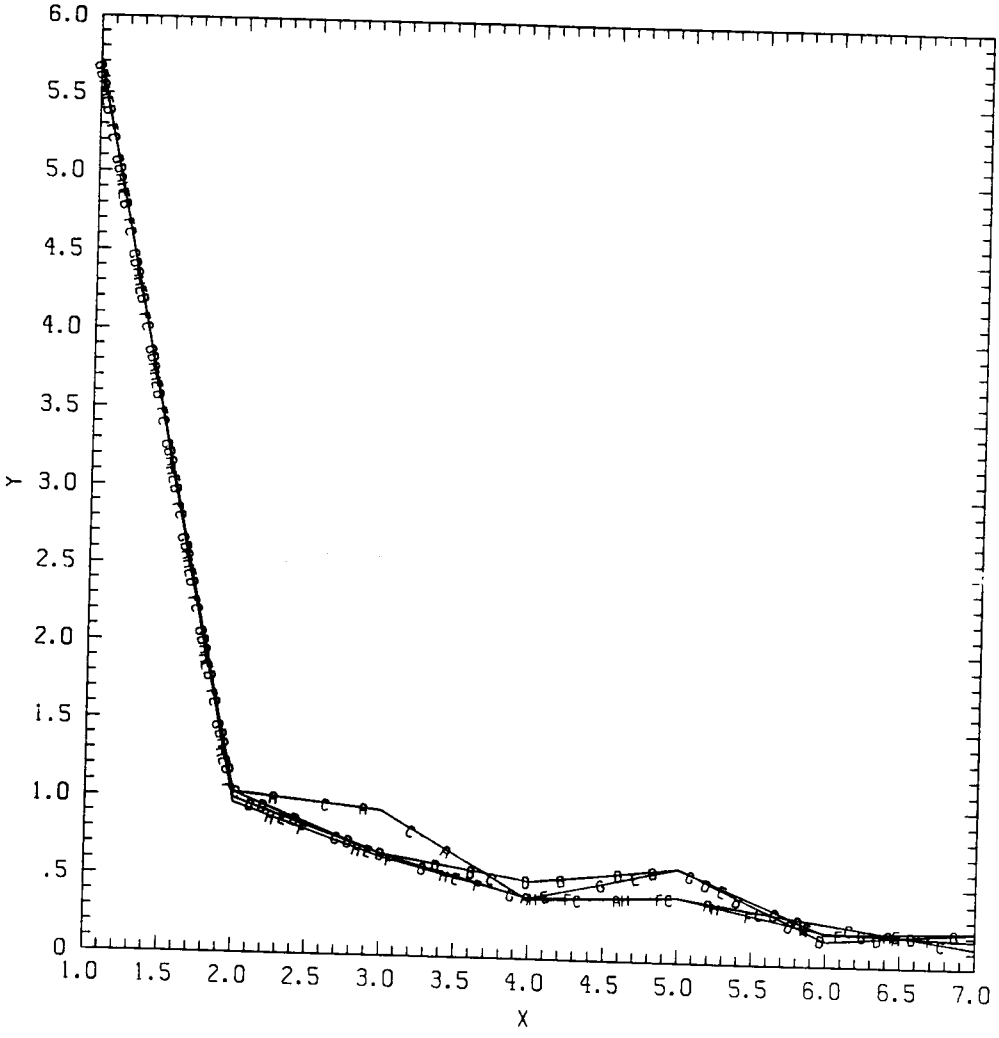


Figure 6.21: Reduction of 3-D distance uncertainty.

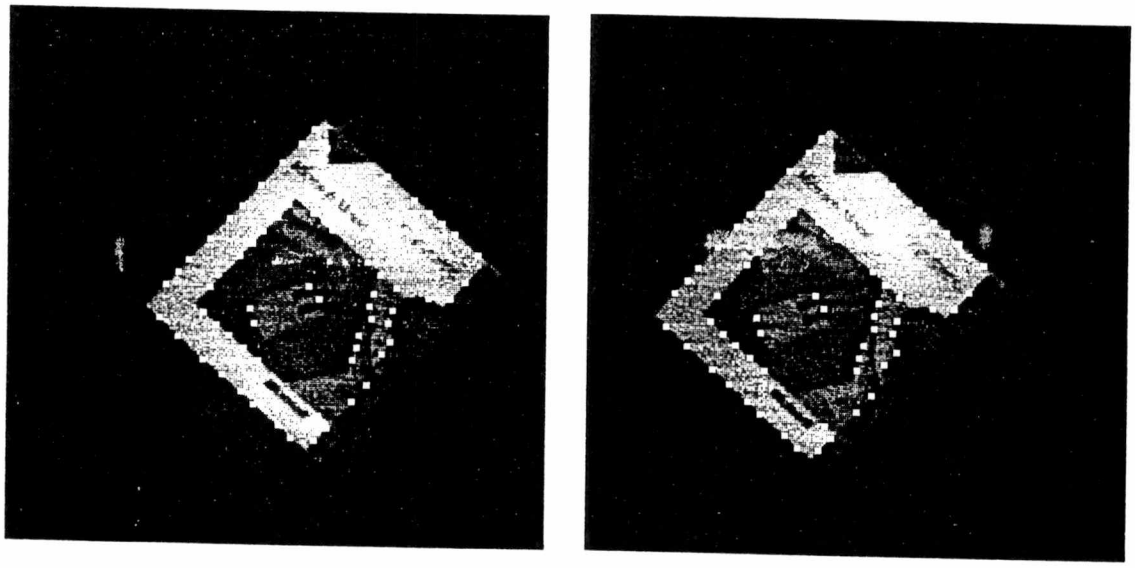


Figure 6.22: Selected matches in $\mathcal{S}_1$ whose 3-D distances are given by Table 1.

Table 6.1: Fused 3-D distances of selected points in $\mathcal{S}_1$.

point	computed Z (cm)	measured Z (cm)	ΔZ (cm)	error %	point	computed Z(cm)	measured Z (cm)	ΔZ (cm)	error %
1	53.88	54.02	0.14	0.25	34	52.94	54.02	1.08	2.01
2	53.80	54.02	0.22	0.41	35	50.82	54.02	3.20	5.92
3	52.93	54.02	1.09	2.01	36	51.77	54.02	2.25	4.16
4	54.51	54.02	-0.49	0.91	37	51.26	54.02	2.76	5.11
5	53.50	54.02	0.52	0.96	38	52.67	54.02	1.35	2.49
6	55.89	54.02	-1.87	3.46	39	50.84	54.02	3.18	5.89
7	52.27	54.02	1.75	3.23	40	53.20	54.02	0.82	1.51
8	53.62	54.02	0.40	0.75	41	51.34	54.02	2.68	4.97
9	52.94	54.02	1.08	2.00	42	53.29	54.02	0.73	1.36
10	55.63	54.02	-1.61	2.98	43	51.53	54.02	2.49	4.61
11	53.12	54.02	0.90	1.67	44	52.89	54.02	1.13	2.10
12	51.76	54.02	2.26	4.18	45	53.56	54.02	0.46	0.85
13	53.36	54.02	0.66	1.22	46	51.55	54.02	2.47	4.57
14	51.78	54.02	2.24	4.15	47	53.14	54.02	0.88	1.62
15	55.90	54.02	-1.88	3.48	48	54.80	54.02	-0.78	1.44
16	52.86	54.02	1.16	2.15	49	51.42	54.02	2.60	4.82
17	51.43	54.02	2.59	4.79	50	51.50	54.02	2.52	4.67
18	54.14	54.02	-0.12	0.23	51	52.72	54.02	1.30	2.40
19	52.91	54.02	1.11	2.06	52	50.47	54.02	3.55	6.58
20	56.23	54.02	-2.21	4.09	53	51.56	54.02	2.46	4.55
21	55.03	54.02	-1.01	1.86	54	53.59	54.02	0.43	0.79
22	54.29	54.02	-0.27	0.50	55	52.09	54.02	1.93	3.57
23	54.94	54.02	-0.92	1.71	56	53.21	54.02	0.81	1.50
24	56.72	54.02	-2.70	4.99	57	51.80	54.02	2.22	4.11
25	50.28	54.02	3.74	6.92	58	52.92	54.02	1.10	2.04
26	56.53	54.02	-2.51	4.65	59	51.44	54.02	2.58	4.77
27	51.41	54.02	2.61	4.82	60	52.53	54.02	1.49	2.76
28	51.39	54.02	2.63	4.86	61	51.64	54.02	2.38	4.41
29	58.50	54.02	-4.48	8.29	62	53.09	54.02	0.93	1.73
30	53.29	54.02	0.73	1.35	63	52.09	54.02	1.93	3.57
31	51.50	54.02	2.52	4.67	64	53.63	54.02	0.39	0.72
32	55.12	54.02	-1.10	2.04	65	51.63	54.02	2.39	4.42
33	51.58	54.02	2.44	4.53					
averaged error of ΔZ: 1.0051 (cm)					averaged error % : 3.1106 %				

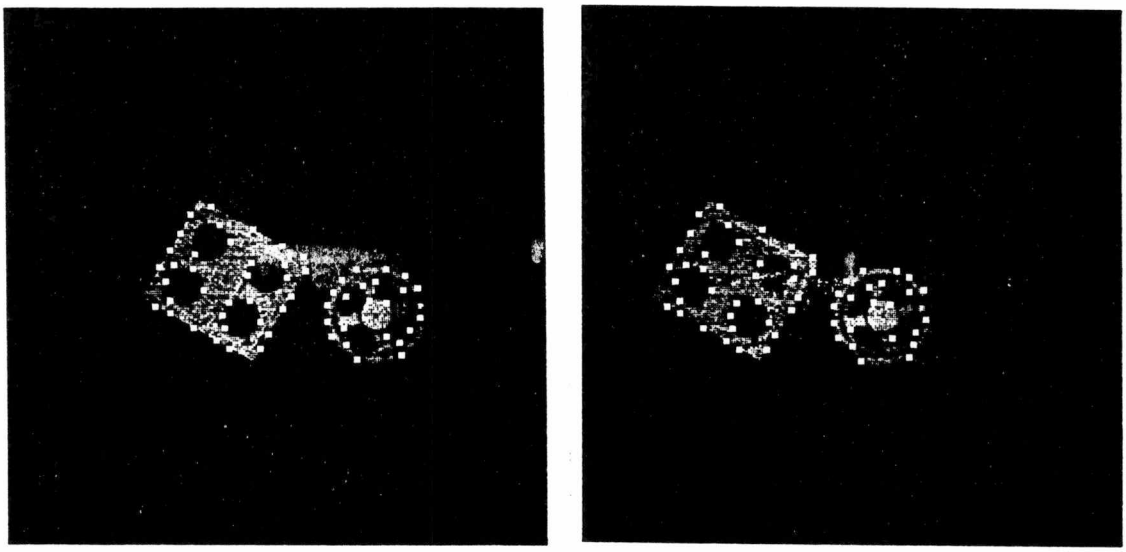


Figure 6.23: Selected matches in $\mathcal{S}_2$ whose 3-D distances are given by Table 2.

Table 6.2: Fused 3-D distances of selected points in $\mathcal{S}_2$.

point	computed Z (cm)	measured Z (cm)	ΔZ (cm)	error %	point	computed Z(cm)	measured Z (cm)	ΔZ (cm)	error %
1	59.30	60.72	1.42	2.33	30	59.53	60.72	1.19	1.97
2	62.73	60.72	-2.01	3.30	31	57.89	59.72	1.83	3.06
3	60.44	60.72	0.28	0.47	32	62.05	60.72	-1.33	2.19
4	58.01	60.72	2.71	4.47	33	58.19	60.72	2.53	4.16
5	63.43	60.72	-2.71	4.46	34	56.24	59.72	3.48	5.82
6	61.00	60.72	-0.28	0.46	35	61.98	60.72	-1.26	2.07
7	59.88	60.72	0.84	1.38	36	58.15	59.72	1.57	2.62
8	63.86	60.72	-3.14	5.17	37	57.94	59.72	1.78	2.98
9	61.91	60.72	-1.19	1.96	38	59.80	60.72	0.92	1.52
10	59.30	60.72	1.42	2.35	39	59.33	60.72	1.39	2.29
11	62.24	60.72	-1.52	2.50	40	61.36	60.72	-0.64	1.06
12	62.04	60.72	-1.32	2.18	41	61.28	60.72	-0.56	0.92
13	59.52	60.72	1.20	1.97	42	56.78	59.72	2.94	4.92
14	64.75	60.72	-4.03	6.64	43	60.32	60.72	0.40	0.66
15	60.86	60.72	-0.14	0.23	44	57.85	59.72	1.87	3.13
16	58.02	59.72	1.70	2.85	45	61.64	60.72	-0.92	1.51
17	60.23	59.72	-0.51	0.86	46	60.82	60.72	-0.10	0.17
18	59.20	60.72	1.52	2.51	47	57.97	59.72	1.75	2.94
19	60.39	60.72	0.33	0.54	48	60.04	59.72	-0.32	0.53
20	58.92	59.72	0.80	1.34	49	60.80	60.72	-0.08	0.14
21	60.06	60.72	0.66	1.09	50	57.19	59.72	2.53	4.23
22	57.61	59.72	2.11	3.53	51	57.25	59.72	2.47	4.13
23	63.70	60.72	-2.98	4.90	52	61.91	60.72	-1.19	1.96
24	57.00	59.72	2.72	4.56	53	59.03	59.72	0.69	1.16
25	58.37	59.72	1.35	2.26	54	56.64	59.72	3.08	5.16
26	58.00	59.72	1.72	2.88	55	61.33	60.72	-0.61	1.00
27	61.82	60.72	-1.10	1.80	56	60.33	60.72	0.39	0.64
28	58.29	59.72	1.43	2.40	57	55.48	59.72	4.24	7.10
29	57.59	59.72	2.13	3.57	58	58.12	59.72	1.60	2.68
averaged error of ΔZ: 0.5698 (cm)					averaged error % : 2.5462 %				

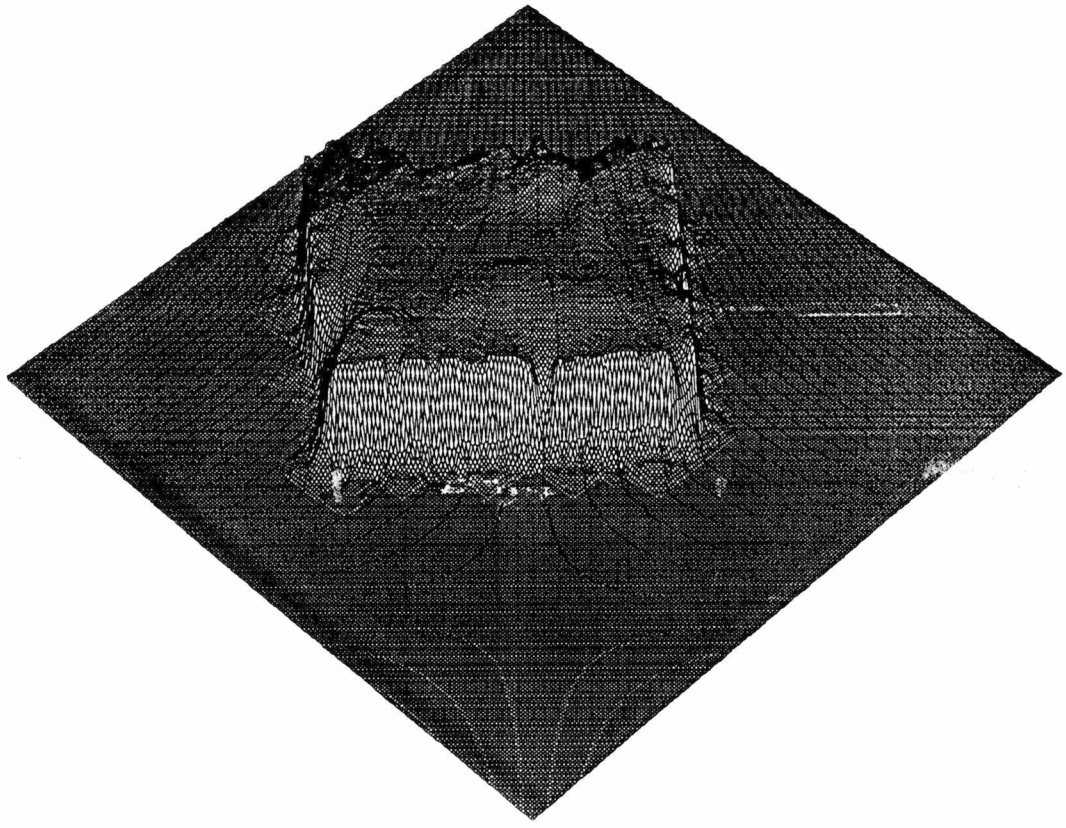


Figure 6.24: Reconstructed object in sequence $\mathcal{S}_1$ by moving least square surface interpolation.

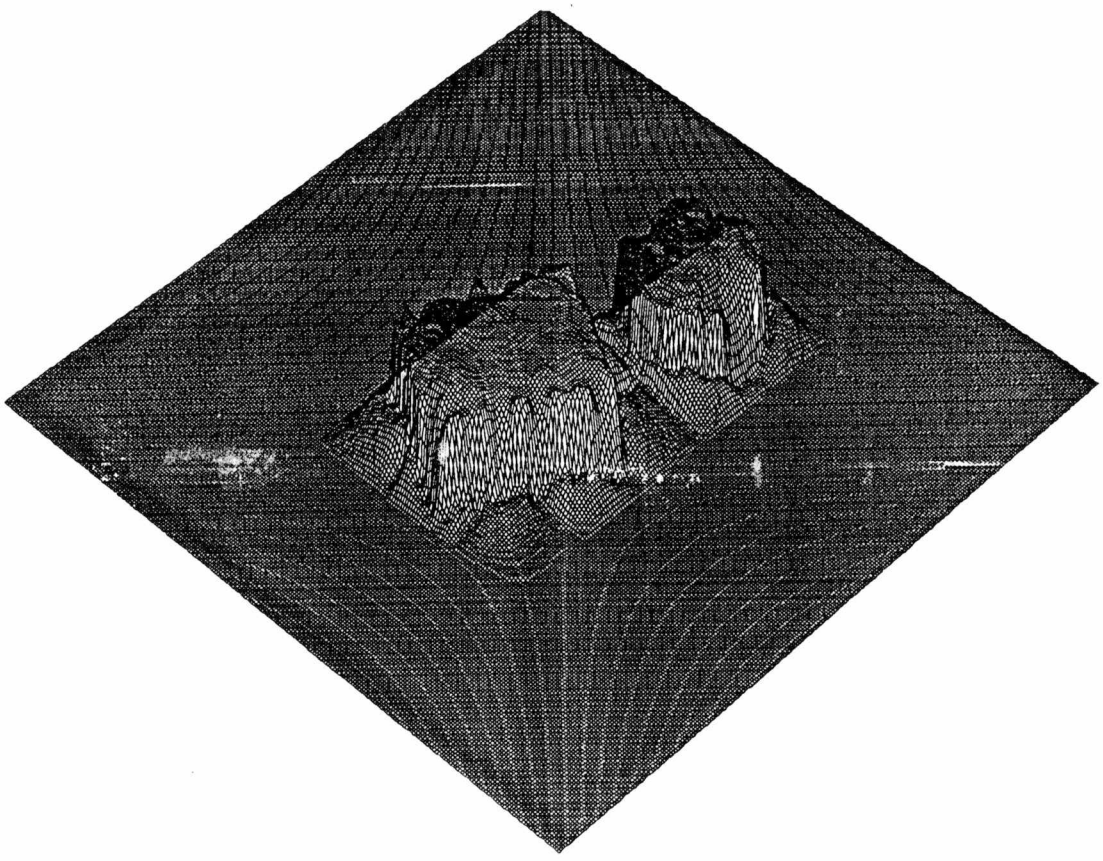


Figure 6.25: Reconstructed object in sequence $\mathcal{S}_2$ by moving least square surface interpolation.

CHAPTER 7

CONCLUSIONS

This work presents a vision system which reconstructs a 3-D scene from sequences of binocular noisy images using optimal, dynamic information fusion algorithms. The vision system is robust to the presence of noise in images, since: (1) it fuses a large amount of information provided by the image pairs and image sequences, and (2) it assigns confidence values to the possible matches and updates the confidence values in the fusion process instead of solving the correspondence problem. The vision system is also robust to the uncertainties of the motion parameters since it takes the uncertainties into account in the geometrical modeling. The vision system is computationally efficient, since (1) it employs a resolution hierarchy, and (2) it utilizes the distributed fusion algorithms which can be implemented in parallel. The spatial uncertainties inherent in calculating 3-D information about the scene are reduced by fusing 3-D information provided by the images that were taken at different locations and different times.

The possible matches in the first image pair at the top level are found by a set of criteria and the corresponding confidence values are assigned based on the signs, the differences of strengths and the differences of the patterns of zero-crossings. These confidence values are then updated by information, provided by the sequences, in the fusion process, propagating from the top level to the

bottom level of the resolution hierarchy. The possible matches found at the higher level predict regions in the first image pair at the lower level, wherein the matching is performed. At the end of the bottom level, the points with confidence values higher than a threshold are considered correct matches. A 3-D scene is reconstructed from the correct matches.

The distributed algorithms for information fusion–confidence value fusion and 3-D information fusion–are developed. The algorithms are optimal and dynamic since they minimize the local and global error covariances and continuously update the fused results when new information is available. The models required for the information fusion are derived by geometrical modeling in which the effect of finite camera resolution is considered and compensated.

Two sets of binocular sequences of noisy images are tested. The results obtained are very satisfactory. The percentage of correct matches is over 95% and the average error of fused 3-D distances of the selected points of the scene is less than 5%. The objects in the scene are reconstructed by using moving least square surface interpolation.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Aisbett, J., "Optical Flow With an Intensity-Weighted Smoothing," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 11, No. 5, pp. 512-522, 1989.

[2] Aggarwal, J.K. and M. J. Magee, "Determining Motion Parameters Using Intensity Guided Range Sensing," *Pattern Recognition*, Vol. 19, No. 2, pp. 169-180, 1986.

[3] Aggarwal, J.K. and N. Nandhakumar, "On the Computation of Motion From Sequences of Images – A Review," *Proc. of IEEE*, Vol. 76, pp. 917-935, 1988.

[4] Allen, P., "Object recognition Using Vision and Touch," *Ph.D. Dissertation*, University of Penn., 1985.

[5] Aloimonos, J., "Robust Computation of Intrinsic Images From Multiple Cues," *Adances in Computer Vision*, Ed. C. Brown, pp. 115-163, Lawrence Erlbaum Assoc., Hillsdale, 1988.

[6] Ayache, N. and O. D. Faugeras, "Building, Registrating, and Fusing Noisy Visual Maps," *Proc. of 1st. International Conference on Computer Vision*, pp. 248-293, London, England, June, 1987.

[7] Ayache, N. and F. Lustman, "Fast and Reliable Passive Trinocular Stereovision," *Proc. of 1st. International Conference on Computer Vision*, pp. 422-427, London, England, June, 1987.

[8] Ayache, N. and B. Faverjon, "A Fast Stereovision Matcher upon Prediction and Verification of Hypothesis," *Proc. of 3rd. Workshop on Computer Vision: representation and Control*, pp. 27-37, Oct., 1985.

[9] Ayache, N. and B. Faverjon, "Efficient Registration of Stereo Images by Matching Graph Descriptions of Edge Segments," *Int. Journ. Computer Vision*, pp. 107-131, 1987.

[10] Barnard, S. T. and W. B. Thompson, "Disparity Analysis of Images," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. PAMI-2, No.4, pp. 333-340, 1980.

[11] Barnard, S. T. and M. A. Fischler, "Computational Stereo," *ACM Computing Surveys,* Vol. 14, No. 4, pp. 553-572, 1982.

[12] Bolle, R. M. and D. B. Cooper, "On Optimally Combining Pieces of Information, With Application to Estimating 3-D Complex-Object Position From Range Data," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. PAMI-8, No.5, pp. 619-638, 1986.

[13] Bruss, A. R. and B. K. P. Horn, "Passive Navigation," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 21, pp. 3-20, 1983.

[14] Burt, P.J., "The Pyramid as a Structure for Efficient Computation," in *Multiresolution Image Processing and Analysis* Ed. A. Rosenfeld, Springer-Verlag, Berlin, pp. 6-35, 1984.

[15] Buxton, B.F., D.W. Murray, H. Buxton, and N.S. Williams, "Structure-From-Motion Algorithms for Computer Vision on an SIMD Architecture," *Comp. Phys. Comm.* Vol. 37, pp. 273-280, 1985.

[16] Canny, J.F., "Finding Edges and Lines," *MIT AI Lab Rep.* No. 720, Cambridge, MA, 1983.

[17] Chang, K. C., and Y. Bar-Shalom, "Distributed Multiple Model Estimation," *Proc. of 1987 American Control Conference*, pp. 797-802, Minneapolis, MN, June, 1987.

[18] Chatterjee, P.S. and T. L. Huntsberger, "Comparison of Techniques for Sensor Fusion under Uncertain Conditions," *Proc. of Sensor Fusion: Spatial Reasoning and Scene Interpretation*, pp. 194-199, Cambridge, MA, Nov., 1988.

[19] Chen, L. H. and T. E. Boult, "An Integrated Approach to Stereo Matching, Surface Reconstruction and Depth Segmentation Using Consistent Smoothness Assumptions," *Proc. of DARPA Image Understanding Workshop*, pp. 166-174, Cambridge, MA, April, 1988.

[20] Chen, S., "A Geometric Approach to Multisensor Fusion and Spatial Reasoning," *Spatial Reasoning and Multi-sensor Fusion: Proc. of the 1987 Workshop*, pp. 201-210, St. Charles, IL, Oct., 1987.

[21] Chen, S., "Fusion of Shading and Structured Lighting Images for Obtaining 3-D Scene Information," *Proc. of Sensor Fusion: Spatial Reasoning and Scene Interpretation*, pp. 38-42, Cambridge, MA, Nov., 1988.

[22] Chiu, S. L., D. J. Morley, and J. F. Martin, "Sensor Data Fusion on a Parallel Processor," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1629-1633, San Francisco, CA, April, 1986.

[23] Chong, C. Y., S. Mori, and K. C. Chang, "Information Fusion in Distributed Sensor Networks," *Proc. of 1985 American Control Conference*, pp. 830-835, Boston, MA, June, 1985.

[24] Chong, C. Y., K.C. Chang, and S. Mori, "Distributed Tracking in Distributed Sensor Networks," *Proc. of 1986 American Control Conference*, pp. 1863-1868, Seattle, WA, June, 1986.

[25] Clark, J.J., "Authenticating Edges Produced by Zero-Crossing Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No.1, pp. 43-57, 1989.

[26] Cohen, L., L. Vinet, and P. T. Sander, "Hierarchical Region Based Stereo Matching," *Proc. Computer Vision and Pattern Recognition*, pp. 416-421, San Diego, CA, June, 1989.

[27] Crowley, J.L., "Navigation for an Intelligent Mobile Robot," *IEEE J. Robotics Automation*, Vol. 1, No. 1, pp. 31-41, 1985.

[28] Delcroix, C. J. and M.A. Abidi, "Fusion of Range and Intensity Edge Maps," *Proc. of Sensor Fusion: Spatial Reasoning and Scene Interpretation*, pp. 145-152, Cambridge, MA, Nov., 1988.

[29] Duda, R.O., D. Nitzan, and P. Barrett, "Use of Range and Reflectance Data to Find Plannar Surface Regions," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-1, No.3, pp. 259-271, 1979.

[30] Durrant-Whyte, H. F., "Consistent Integration and Propagation of Disparate Sensor Observations," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1464-1469, San Francisco, CA, April, 1986.

[31] Durrant-Whyte, H.F., *Integration, Coordination and Control of Multi-Sensor Robot Systems*, Kluwer Academic Publ., Boston, MA, 1988.

[32] Eason, R. O., "Determining the Pose of Three-Dimensional Objects by Multisensor Fusion," *Ph.D. Dissertation*, University of Tenn., Knoxville, 1988.

[33] Eastman, R.D. and A.M. Waxman, "Disparity Functionals and Stereo Vision," *Proc. of DARPA Image Understanding Workshop*, pp. 245-254, 1985.

[34] Fang, J.Q. and T.S. Huang, "Solving Three-Dimensional Small-rotation Motion Equations: Uniqueness, Algorithms, and Numerical Results," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 26, pp. 183-206, 1984.

[35] Faugeras, O. D., N. Ayache, and B. Faverjon, "Building Visual Maps by Combining Noisy Stereo Measurements," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1433-1438, San Francisco, CA, April, 1986.

[36] Faugeras, O. D. and G. Toscanl, "Calibration Problem for Stereo," *proc. Computer Vision and Pattern Recognition*, pp. 15-20, Miami Beach, FL, June, 1986.

[37] Faugeras, O. D., F. Lustman, and G. Toscani, "Motion and Structure From Point and Line Matches," *Proc. of 1st. International Conference on Computer Vision*, pp. 25-34, London, England, June, 1987.

[38] Flynn, A.M., "Combining Sonar and Infrared Sensors for Mobile Robot Navigation," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 5-14, 1988.

[39] Fu, K.S., R.C. Gonzalez, and C.S.G. Lee, *Robotics: Control, Sensing, Vision, and Intelligence*, McGraw-Hill, New York, NY, 1987.

[40] Fuh, C.S. and P. Maragos, "Region-Based Optical Flow Estimation," *Proc. Computer Vision and Pattern Recognition,* pp. 130-134, San Diego, CA, June, 1989.

[41] Gamble, E. and T. Poggio, "Visual Integration and Detection of Discontinuities: the Key Role of Intensity Edges," *MIT AI Memo*, No. 970, Oct. 1987.

[42] Garvey, T.D., "A Survey of AI Approaches to the Integration of Information," *Proc. SPIE*, Cambridge, MA, Nov., 1987.

[43] Garvey, T.D., J.D. Lowrance, and M. Fischler, "An Inference Technique for Integrating Knowledge From Disparate Sources," *Reading in Knowledge Representation,* Eds. R. J. Brachman and H.J. Levesque, pp. 457-464, Mogan Kaufman, Inc., Los Altos, CA, 1985.

[44] Gazit, S.L. and G. Medioni, "Contour Correspondences in Dynamic Imagery," *Proc. of DARPA Image Understanding Workshop*, pp. 423-432, Cambridge, MA, April, 1988.

[45] Gelb, A., *Applied Optimal Estimation.* MIT Press, Cambridge, MA, 1974.

[46] Geman, S. and G. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. PAMI-6, No. 6, pp. 721-741, 1984.

[47] Gennery, D. B., "Stereo Vision for Acquisition and Tracking of Moving Three-Dimensional Objects," in *Techniques for 3-D Machine Perception,* Ed. A. Rosenfeld, North-Holland, Amsterdam, 1986.

[48] Gennery, D. B., "Object Detection and Measurement Using Stereo Vision," *Proc. of DARPA Image Understanding Workshop*, pp. 161-167, College Park, MD, April, 1980.

[49] Gil, B., A. Mitiche, and J.K. Aggarwal, "Experiments in Combining Intensity and Range Edge Maps," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 21, pp. 395-411, 1983.

[50] Glazer, F., "Hierarchical Gradient-Based Motion Detection," *Proc. of DARPA Image Understanding Workshop*, pp. 733-748, Los Angeles, CA, Feb., 1987.

[51] Grandjean P. and A.R.D.S. Vincent, "3-D Modeling of Indoor Scenes by Fusion of Noisy Range and Stereo Data," *Proc. of IEEE Conference on Robotics and Automation*, pp. 681-687, Scottsdale, AZ, May, 1989.

[52] Grimson, W.E.L., *From Images to Surfaces*, MIT Press, Cambridge, MA., 1981.

[53] Grimson, W.E.L., "Computational Experiments With a Feature Based Stereo Algorithm," *IEEE Trans. PAMI*, vol. PAMI-7, No. 1, pp. 17-33, 1985.

[54] Griswold, N.C. and C. P. Yeh, "A New Stereo Vision Model Based upon the Binocular Fusion Concept," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 41, pp. 153-171, 1988.

[55] Hagel, H.H., "Image Sequences - Ten (Octal) Years- From Phenomenology towards a Theoretical Foundation," *Proc. Int. Conf. Pattern Recognition*, pp. 1174-1185, Paris, France, Nov., 1986.

[56] Hager, G. and M. Mintz, "Task-Directed Multi-Sensor Fusion," *Proc. of IEEE Conference on Robotics and Automation*, pp. 662-667, Scottsdale, AZ, May, 1989.

[57] Harmon, S.Y., G.L. Bianchini, and B.E. Pinz, "Sensor Data Fusion Through A Distributed Blackboard," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1449-1454, San Francisco, CA, April, 1986.

[58] Hashemipour, H. R., S. roy, and A. J. Laub, "Decentralized Structure for Parallel Kalman Filtering," *IEEE Trans. Automat. Contr.* Vol. 33, No. 1, pp. 88-94, 1988.

[59] Hassan, M.F., G. Salut, M. G. Singh, and A. Titli, "A Decentralized Computational Algorithm for the Global Kalman Filter," *IEEE Tran. Auto. Contr.*, Vol. AC-23, No. 2, pp. 262-267, 1978.

[60] Henderson, T., C. Hansen, and B. Bhanu, "The Specification of Distributed Sensing and Control," *Journ. of Robot. Systems,*, Vol. 4, No. 4, pp. 387-396, 1985.

[61] Henderson, T., E. Weitz, C. Hansen, and A. Mitiche, "Multisensor Knowledge Systems: Interpreting 3-D Structure," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 114-137, 1988.

[62] Hoff, W. and N. Ahuja, "Extracting Surfaces From Stereo Images: An Integrated Approach," *Proc. of 1st. International Conference on Computer Vision*, pp. 248-293, London, England, June, 1987.

[63] Hoff, W. and N. Ahuja, "Surfaces From Stereo: Integrating Features Matching, Disparity Estimation, and Contour Detection," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No.2, pp. 121-136, 1989.

[64] Hoffman, D.D., "Inferring Local Surface Orientation From Motion Fields," *J. Opt. Soc. Amer.* Vol. 72, No. 7, pp. 888-892, 1982.

[65] Horn, B.K.P. and B.G. Schunck, "Determining Optical Flow," *Artif. Intell.*, Vol. 17, pp. 185-203, 1981.

[66] Horn, B. K. P. and E. J. Weldon, Jr., "Computationally-Efficient Methods for recovering Translational Motion," *Proc. of 1st. International Conference on Computer Vision*, pp. 2-11, London, England, June, 1987.

[67] Hu, G. and G. Stockman, "3-D Scene Analysis via Fusion of Light Striped Image and Intensity Image," *Spatial Reasoning and Multi-sensor Fusion: Proc. of the 1987 Workshop*, pp. 138-147, St. Charles, IL, Oct., 1987.

[68] Huang, T.S. and R.Y. Tsai, "Image Sequence Analysis: Motion Estimation," *Image Sequence Analysis,* Ed. T.S. Huang, Springer-Verlag, Berlin, 1981.

[69] Kanatani, K., "Structure and Motion From Optical Flow under Perspective Projection," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 38, pp. 122-146, 1987.

[70] Kent, E. W., M.O. Shneier, and T. Hong, "Building Representaions From Fusion of Multiple Views," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1634-1637, San Francisco, CA, April, 1986.

[71] Kim, N.K. and A.C. Bovik, "A Solution to the Stereo Correspondence Problem using Disparity Smoothness Constraint," *Proc. Int. Conf. Sys. Man Cyber.*, pp. 987-991, Atlanta, GA, Oct., 1986.

[72] Kim, Y.C. and L.K. Aggarwal, "Positioning Three-Dimensional Objects Using Stereo Images," *IEEE Journ. of Robotics and Automation*, Vol. RA-3, No. 4, pp. 361-373, 1987.

[73] Koenderink, J.J. and A. J. Van Doorn, "Geometry of Binocular Vision and A model for Stereopsis," *Biological Cybernetics,* Vol. 21, pp. 29-35, 1976.

[74] Lajeunesse, T. J., "Sensor Fusion," *Defense Science and Electronics,* pp. 21-31, Sept. 1986.

[75] Liu, Y.C. and T.S. Huang, "Estimation of Rigid Body Motion Using Straight Line Correspondences," *Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis*, pp. 47-52, May, 1986.

[76] Lowrance, J.D., T. D. Garvey, and T. M. Strat, "A Framework for Evidential-Reasoning Systems," *Proc. Nati. Conf. Artifi. Intell.* pp. 896-903, Menlo Park, CA, 1986.

[77] Luo, R.C., M.H. Lin, and R.S. Scherp, "The Issues and Approaches of Robot Multi-Sensor Integration," *Porc. IEEE Int. Conf. on robot. and Auto.*, pp. 1941-1946, Raleigh, NC, March, 1987.

[78] Luo, R.C. and M.H. Lin, "Robot Multi-Sensor Fusion and Integration: Optimum Estimation of Fused Sensor Data," *Porc. IEEE Int. Conf. on Robot. and Auto.*, pp. 1076-1081, Philadelphia, Penn., April, 1988.

[79] Longuet-Higgins, H.C., F. R. S., and K. Prazdny, "The Interpretation of a Moving Retinal Image," *Proc. R. Soc. Lond. B*, Vol. 208, pp. 385-397, 1980.

[80] Longuet-Higgins, H.C., F. R. S., "A Computer Algorithm for Reconstructing a Scene From Two Projections," *Nature,* Vol. 239, pp. 133- 135, 1981.

[81] Magee, M. J. and J. K. Aggarwal, "Using Multisensory Images to Derive the Structure of Three-Dimensional Objects – A Review," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 32, pp. 145-157, 1985.

[82] Mann, R.C., "Multi-Sensor Integration Using Concurrent Computing," *Proc. of SPIE Symposiumon Infrared Sensors and Sensor Fusion*, Orlando, FL, May, 1987.

[83] Marr, D. and T. Poggio, "Cooperative Computation of Stereo Disparity," *Science,* Vol. 194, pp. 283-287, Oct., 1976.

[84] Marr, D. and T. Poggio, "A Computational Theory of Human Stereo Vision," *Proc. R. Soc. Lond. B*, Vol. 204, pp. 301-328, 1979.

[85] Marr, D. and E.C. Hildreth, "Theory of Edge Detection," *Proc. R. Soc. Lond. B*, Vol. 207, pp. 187-217, 1980.

[86] Matthies, L. and S. A. Shafer, "Error Modeling in Stereo Navigation," *IEEE Journ. of Robotics and Automation*, Vol. RA-3, No. 3, pp. 239-248, 1987.

[87] Matthies, L., R. Szeliski, and T. de, "Incremental Estimation of Dense Depth Maps From Images Sequences," *Proc. Computer Vision and Pattern Recognition,* pp. 2-9, Ann Arbor, MI, June, 1988.

[88] Mayhew, J.E.W. and J. P. Frisby, "The Computation of Binocular Edges," *Perception*, Vol. 9, pp. 69-86, 1980.

[89] Mayhew, J.E.W. and J. P. Frisby, "Psychophysical and Computational Studies towards a Theory of Human Stereopsis," *Artif. Intell.* Vol. 17, pp. 349-385, 1981.

[90] Mayhew, J.E.W., "The Interpretation of Stereo-Disparity Information: The Computation of Surface Orientation and Depth," *Perception*, Vol. 11, pp. 387-403, 1982.

[91] Medioni, G.G. and R. Nevatia, "Segment-Based Stereo Matching," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 31, pp. 2-18, 1985.

[92] Mitiche, A. and J.K. Aggarwal, "Image Segmentation by Conventional and Information-Integrating Techniques: A Synopsis," *Image and Vision Computing*, Vol. 3, No. 2, pp. 50-62, 1985.

[93] Mitiche, A. and J.K. Aggarwal, "Multiple Sensor Integration/Fusion Through Image Processing: A Review," *Optical Engineering*, Vol. 25, No. 3, pp. 380-386, 1986.

[94] Mohan, R., G. Medioni, and R. Nevatia, "Stereo Error Detection, Correction, and Evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No.2, pp. 113-120, 1989.

[95] Moravec, H. P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," *Ph.D. Dissertation*, Stanford Univ., Stanford, CA, Sept., 1980.

[96] Nagel, H.H., "Representation of Moving Rigid Objects Based on Visual Observations," *Computer*, pp. 29-39, 1981.

[97] Nakamura, Y. and Y. Xu, "Geometrical Fusion Method for Multi- Sensor Robotic Systems," *Proc. of IEEE Conference on Robotics and Automation*, pp. 668-673, Scottsdale, AZ, May, 1989.

[98] Ohta, Y. and T. Kanade, "Stereo by Intra- and Inter-Scanline Search," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-7, No.2, pp. 139-154, 1985.

[99] Pau, L.F., "Sensor Data Fusion," *Journ. of Intell. and Robot. Sys.*, No. 1, pp. 103-116, 1988.

[100] Pietikainen, M. and D. Harwood, "Depth From Three Camera Stereo," *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2–8, Miami, FL, June, 1986.

[101] Poggio, T. and Staff, "MIT Progress in Understanding Images," *Proc. of DARPA Image Understanding Workshop*, pp. 41-54, Los Angeles, CA, Feb., 1987.

[102] Poggio, T., J. Little, E. Gamble, W. Gillett, D. Geiger, D. Weinshall, M. Villalba, N. Larson, T. Cass, H. Bulthoff, M. Drumheller, P. Oppenheimer, W. Yang, and A. Hurlbert, "The MIT Vision Machine," *Proc. of DARPA Image Understanding Workshop*, pp. 177-198, Cambridge, MA, April, 1988.

[103] Porrill, J., S.B. Pollard, S. B., and J. E. W. Mayhew, "Optimal Combination of Multiple Sensors Including Stereo Vision," *Image and Vision Computing*, Vol. 5, No. 2, pp. 174-180, 1987.

[104] Porrill, J., "Optimal Combination and Constraints for Geometrical Sensor Data," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 66-77, 1988.

[105] Prazdny, K., "Determining the Instantaneous Direction of Motion From Optical Flow Generated by a Curvilinearly Moving Observer," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 17, pp. 238-248, 1981.

[106] Richardson, J.M. and K. A. Marsh, "Fusion of Multisensor Data," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 78-96, 1988.

[107] Roach, J. W. and J. K. Aggarwal, "Determining the Movement of Objects From a Sequence of Images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-2, No. 6, pp. 554-562, 1980.

[108] Rodriguez, J.J. and J.K. Aggarwal, "Quantization Error in Stereo Imaging," *Proc. Computer Vision and Pattern Recognition*, pp. 153-158, Ann Arbor, MI, June, 1988.

[109] Roecker and J.A., C.D. McGillem, "Comparison of Two-sensor Tracking Methods Based on State Vector Fusion and Measurement Fusion," *IEEE Tran. Aerosp. and Electron. Sys.*, Vol. 24, No. 4, pp. 447-449, 1988.

[110] Rosenfeld, A., "Survey - Image Analysis and Computer Vision: 1988," *Comp. Vision, Graphics and Imag. Proc.*, Vol. 46, pp. 196-264, 1989.

[111] Schweppe, F. C., *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[112] Shekhar, S., O. Khatib, and M. Shimojo, "Sensor Fusion and Object Localization," *Proc. of IEEE Conference on Robotics and Automation*, pp. 1623-1628, San Francisco, CA, April, 1986.

[113] Shekhar, S., O. Khatib, and M. Shimojo, "Object Localization With Multiple Sensors," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 34-43, 1988.

[114] Sher, D., "Evidence Combination Using Likelihood Generators," *Proc. of DARPA Image Understanding Workshop*, pp. 655-662, Los Angeles, CA, Feb., 1987.

[115] Slama, C.C., *Manual of Photogrammetry*, Falls Church, VA: American Society of Photogrammetry, 1980.

[116] Smith, R.C. and P. Cheeseman, "On the Representation and Estimation of Spatial uncertainty," *Int. Journ. Robotics Research*, Vol. 5, No. 4, pp. 56-68, 1985.

[117] Solina, F., "Errors in Stereo due to Quantization," *Tech. Rep. MS-CIS-85-34*, Univ. Penn., Sept., 1985.

[118] Stansfield, S.A., "A Robotic Perceptual System Utilizing Passive Vision and Active Touch," *Int. Journ. Robotics Research*, Vol. 7, No. 6, pp. 138-161, 1988.

[119] Stansfield, S.A., "Integrating Multiple Views into a Single Representation of a Range Imaged Object," *Proc. of Sensor Fusion: Spatial Reasoning and Scene Interpretation*, pp. 52-67, Cambridge, MA, Nov., 1988.

[120] Sufi, N., "Fuzzy Pyramid Linking," *M.S. Thesis*, University of Tenn., Knoxville, Dec., 1989.

[121] Thomopoulos, S.C.A., R. Viswanathan, D. K. Bougoulias, and L. Zhang, "Optimal and Suboptimal Distributed Decision Fusion," *Proc. of 1988 American Control Conference*, pp. 414-418, Atlanta, GA, June, 1988.

[122] Tsai, R.T. and T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-6, No.1, pp. 13-27, 1984.

[123] Ullman, S., *The Interpretation of Vision Motion*, MIT Press, Cambridge, MA, 1979.

[124] Ullman, S., "The Interpretation of Structure From Motion," *Proc. of Royal Society of London*, Series B, Vol. 203, pp. 405-426, 1979.

[125] Verri, A. and T. Poggio, "Against Quantitative Optical Flow," *Proc. of 1st. International Conference on Computer Vision*, pp. 171-180, London, England, June, 1987.

[126] Wang, Y.F. and Aggarwal, J.K., "On Modeling 3-D Objects Using Multiple Sensory Data," *Proc. IEEE Int. Conf. on Robot. and Auto.*, pp. 1098-1102, Raleigh, NC, March, 1987.

[127] Waxman, A. M. and J. H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-8, No.6, pp. 715-729, 1986.

[128] Weng, J., T.S.Huang, and N. Ahuja, "3-D Motion Estimation, Understanding, and Prediction From Noisy Image Sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-9, No.3, pp. 370-389, 1987.

[129] Weng, J., T.S.Huang, and N. Ahuja, "Motion and Structure From Two Perspective Views: Algorithms, Error Analysis, and Estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, No. 5, pp. 451-476, 1989.

[130] Weng, J., N. Ahuja, and T.S. Huang, "Optimal Motion and Structure Estimation," *Proc. Computer Vision and Pattern Recognition*, pp. 144-152, San Diego, CA, June, 1989.

[131] Willner, D., C. B. Chang, and K. P. Dunn, "Kalman Filter Algorithms for A Multi-Sensor System," *Proc. of IEEE Conf. Decision Contr.*, pp. 570-574, Clearwater, FL, Dec., 1976.

[132] Wolff, L. B., "Accurate Measurement of Orientantion From Stereo Using Line Correspondence," *Proc. Computer Vision and Pattern Recognition*, pp. 410-415, San Diego, CA, June, 1989.

[133] Wu, J., R. Brockett, and K. Wohn, "A Contour-Based Recovery of Image Flow: Iterative Method," *Proc. Computer Vision and Pattern Recognition*, pp. 124-129, San Diego, CA, June, 1989.

[134] Yachida, M., "Determining Velocity Map by 3-D Iterative Estimator," *Proc. Int. Joint Conf. On Artif. Intell.*, pp. 716-718, 1981.

[135] Yachida, M., Y. Kitamura, and M. Kimachi, "Trinocular Vision: New Approach for Correspondence Problem," *Proc. Int. Conf. Pattern Recognition*, pp. 1041-1044, Paris, France, Nov., 1986.

[136] Zhuang, T.S. Huang, and X. and R. M. Haralick, "Two-View Motion Analysis: A Unified Algorithm," *J. Opt. Soc. Amer.* Vol. 3, No. 9, pp. 1492-1500, 1986.

APPENDIX

MULTI-SENSOR/INFORMATION FUSION ALGORITHMS

The general algorithms for model-based multi-sensor/information fusion are derived in this appendix. A theory for multi-sensor/information fusion for static and dynamic sensor systems is presented in the following. The concept of sensor in this work is general and refers to both physical and logical sensors. Cameras and touch sensors are examples of physical sensors, and 3-D information provided by a binocular matching algorithm is an example of a logical sensor. For physical sensors, the term *multi-sensor fusion* is applied, while the term *information fusion* is associated with logical sensors. In the following, the distributed and centralized algorithms are derived first for linear dynamic multi-sensor systems. Then the distributed and centralized algorithms for linear static multi-sensor systems are evolved from the algorithms for dynamic systems by considering the static systems as special cases of dynamic ones. The multi-sensor systems studied here are such that each sensor is associated with its own coordinate system and the communication networks between sensors are with uncertainties.

The algorithms for multi-sensor/information fusion developed here are:

- General and applicable to different multi-sensor systems

- Optimal in the sense of minimizing a specific cost function
- Computationally efficient
- Easy to implement
- Flexible in manipulating uncertainties and noise

A.1 Multi-Sensor System Descriptions

A multi-sensor system can be dynamic or static depending on the motion of sensors or objects. A static sensor system can be treated as a special case of a dynamic sensor system. When sensors or objects move, Figure A.1, they are modeled by a dynamic system,

$$\begin{aligned} \underline{x}_{k+1} &= \mathbf{\Phi}_k(t_{k+1}, t_k)\underline{x}_k + \mathbf{\Gamma}_k(\underline{x}_k, t_k)\underline{w}_k, \quad E\{\underline{w}_k\} = 0, \; E\{\underline{w}_k\underline{w}_k^T\} = \mathbf{Q}_k, \\ \underline{z}_k^i &= \mathbf{H}_k^i(t_k)\underline{x}_k^i + \underline{v}_k^i, \qquad i = 1, ..., N, \quad E\{\underline{v}_k^i\} = 0, \quad E\{\underline{v}_k^i(\underline{v}_k^i)^T\} = \mathbf{R}_k^i, \end{aligned} \tag{1.1}$$

where $\underline{x}_k$, which is defined in an l-dimensional real vector space, i.e., $\underline{x}_k \in \mathcal{R}^l$, is the quantity being measured and estimated at the kth moment. The system matrix $\mathbf{\Phi}_k(t_{k+1}, t_k) \in \mathcal{R}^{l \times l}$ is determined by the relative motion between sensors and objects, and the modeling error $\underline{w}_k \in \mathcal{R}^l$ represents the uncertainties in the motion parameters characterized by the error covariance matrix $\mathbf{Q}_k$. There are N measurement models in the sensor system (1), each of which describes a sensor. The symbol $\underline{z}_k^i \in \mathcal{R}^m$ describes the measurement acquired by sensor i with an error $\underline{v}_k^i$ specified by $\mathbf{R}_k^i$, and $\mathbf{H}_k^i(t_k) \in \mathcal{R}^{m \times l}$ represents the ideal measurement operation of the sensor i. The communication networks between the sensors and

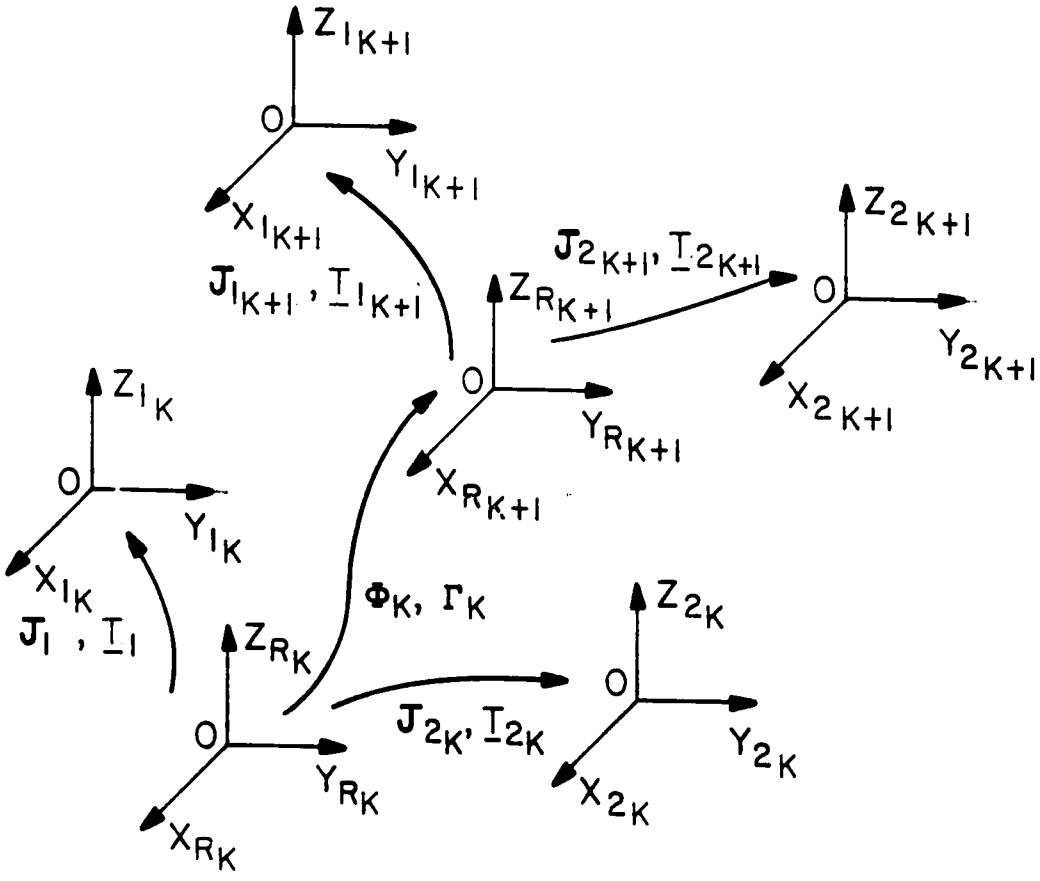


Figure A.1: A dynamic multi-sensor system.

the central processor are

$$\underline{x}_k^i = \mathbf{J}_{i_k}\underline{x}_k + \underline{T}_{i_k}, \quad i = 1, ..., N, \tag{1.2}$$

where $\mathbf{J}_{i_k}$ is a rotation matrix and $\underline{T}_{i_k}$ is a translation vector [39,116], which have the uncertainties

$$\mathbf{J}_{i_k} = \mathbf{J}_{i0_k} + \Delta\mathbf{J}_{i_k}, \quad \Delta\mathbf{J}_{i_k} \sim N(0, \mathbf{Q}^i_{J_k}), \tag{1.3}$$

$$\underline{T}_{i_k} = \underline{T}_{i0_k} + \Delta\underline{T}_{i_k}, \quad \Delta\underline{T}_{i_k} \sim N(0, \mathbf{Q}^i_{T_k}). \tag{1.4}$$

A Special Case – Static Sensor System

When the sensors and the objects are steady, Figure A.2, the sensor system becomes static and can be described using the dynamic sensor system, Eq. (1), by setting $\mathbf{\Phi}_k(t_{k+1}, t_k) = \mathbf{I}$ and $\mathbf{Q}_k = 0$ and dropping the subscript k, i.e.,

$$\underline{z}^i = \mathbf{H}^i(\underline{x}) + \underline{v}^i, \quad i = 1, ..., N, \quad E\{\underline{v}^i\} = 0, \quad E\{\underline{v}^i(\underline{v}^i)^T\} = \mathbf{R}^i. \tag{1.5}$$

The communication networks between the sensors and the central processor become

$$\underline{x}^i = \mathbf{J}_i\underline{x} + \underline{T}_i, \quad i = 1, ..., N, \tag{1.6}$$

where

$$\mathbf{J}_i = \mathbf{J}_{i0} + \Delta\mathbf{J}_i, \quad E\{\Delta\mathbf{J}_i\} = 0, \quad E\{\Delta\mathbf{J}_i\Delta\mathbf{J}_i^T\} = \mathbf{Q}_{J_i}, \tag{1.7}$$

and

$$\underline{T}_i = \underline{T}_{i0} + \Delta\underline{T}_i, \quad E\{\Delta\underline{T}_i\} = 0, \quad E\{\Delta\underline{T}_i\Delta\underline{T}_i^T\} = \mathbf{Q}_{T_i}. \tag{1.8}$$

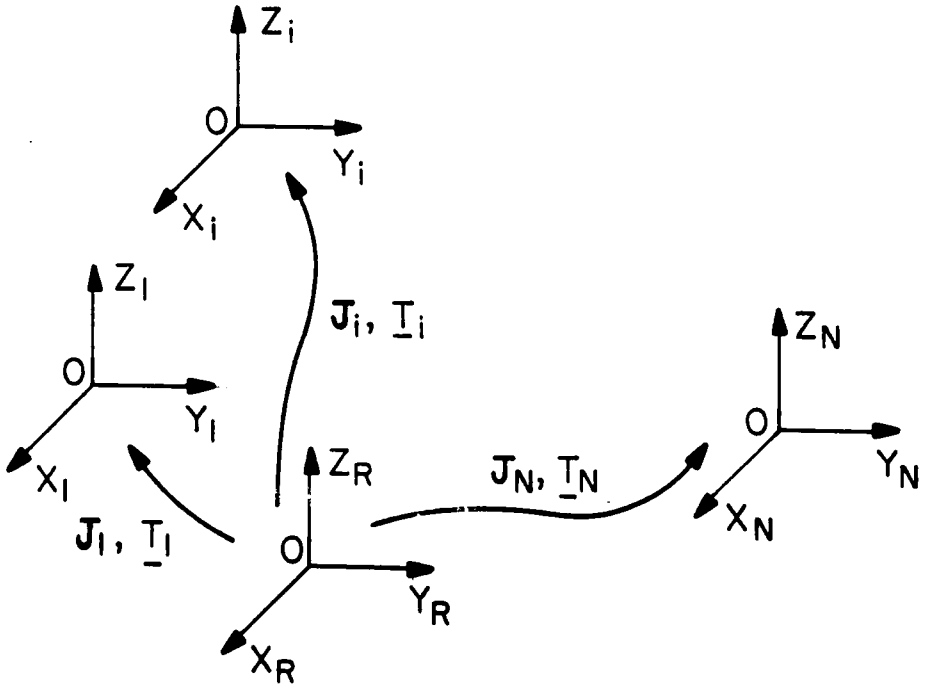


Figure A.2: A static multi-sensor system.

A.2 A Transformation of Uncertainties

A transformation of uncertainties which is utilized extensively in this work is presented first before the derivation of the algorithms. A general transformation between the coordinate systems of sensors is given by

$$\underline{x}^i = \mathbf{J}_i \underline{x} + \underline{T}_i \quad i = 1, \ldots, N, \tag{1.9}$$

where $E\{\underline{x}\} = \underline{m}_x$ and $E\{(\underline{x} - \underline{m}_x)(\underline{x} - \underline{m}_x)^T\} = \Psi$ and the mapping functions $\mathbf{J}_i$ and $\underline{T}_i$ are with uncertainties

$$\mathbf{J}_i = \mathbf{J}_{i0} + \Delta\mathbf{J}_i, \quad E\{\Delta\mathbf{J}_i\} = 0, \quad E\{\Delta\mathbf{J}_i \Delta\mathbf{J}_i^T\} = \mathbf{Q}_{J_i}, \tag{1.10}$$

and

$$\underline{T}_i = \underline{T}_{i0} + \Delta\underline{T}_i, \quad E\{\Delta\underline{T}_i\} = 0, \quad E\{\Delta\underline{T}_i \Delta\underline{T}_i^T\} = \mathbf{Q}_{T_i}. \tag{1.11}$$

The mean vector and the covariance matrix of $\underline{x}^i$ are calculated as follows.

i) Mean vector

Taking expectation in both sides of Equation (9) gives

$$\begin{aligned} \underline{m}_{x^i} = E\{\underline{x}^i\} &= E\{\mathbf{J}_i \underline{x} + \underline{T}_i\} && (1.12) \\ &= E\{(\mathbf{J}_{i0} + \Delta\mathbf{J}_i)\underline{x} + (\underline{T}_{i0} + \Delta\underline{T}_i)\} && (1.13) \\ &= \mathbf{J}_{i0}\underline{m}_x + \underline{T}_{i0}, && (1.14) \end{aligned}$$

where the property of $E\{\Delta\mathbf{J}_i \underline{x}\} = 0$ was used.

ii) Covariance matrix

Since

$$\underbrace{\begin{bmatrix} x^i \\ y^i \\ z^i \end{bmatrix}}_{\underline{x}^i} = \underbrace{\begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix}_i}_{\mathbf{J}_i} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\underline{x}} + \underbrace{\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}_i}_{\underline{T}_i} \tag{1.15}$$

or

$$x^i = J_{i11}x + J_{i12}y + J_{i13}z + T_{i1}, \tag{1.16}$$

$$y^i = J_{i21}x + J_{i22}y + J_{i23}z + T_{i2}, \tag{1.17}$$

$$z^i = J_{i31}x + J_{i32}y + J_{i33}z + T_{i3}, \tag{1.18}$$

expanding x^i, y^i, z^i into Taylor series at nominal point: $\mathbf{J}_{i0}$, $\underline{T}_{i0}$ and $\underline{m}_x$ and retaining only the first order terms, we have the following

$$\underline{m}_{x^i} = \begin{bmatrix} \overline{x}^i \\ \overline{y}^i \\ \overline{z}^i \end{bmatrix} = \mathbf{J}_{i0}\underline{m}_x + \underline{T}_{i0},$$

$$\begin{aligned} \Delta x^i = x^i - \overline{x}^i = &\left[\frac{\partial x^i}{\partial J_{i11}}\right]_* \Delta J_{i11} + \left[\frac{\partial x^i}{\partial x}\right]_* \Delta x + \left[\frac{\partial x^i}{\partial J_{i12}}\right]_* \Delta J_{i12} \\ &+ \left[\frac{\partial x^i}{\partial y}\right]_* \Delta y + \left[\frac{\partial x^i}{\partial J_{i13}}\right]_* \Delta J_{i13} + \left[\frac{\partial x^i}{\partial z}\right]_* \Delta z \\ &+ \left[\frac{\partial x^i}{\partial T_{i1}}\right]_* \Delta T_{i1} \end{aligned} \tag{1.19}$$

where

$$[\ ]_* = [\ ]_{J_{i0}, T_{i0}, \underline{m}_x = [\overline{x}, \overline{y}, \overline{z}]^T}. \tag{1.20}$$

Written in another form, Equation (19) becomes

$$\Delta x^i = \mathbf{J}_{11} \begin{bmatrix} \Delta J_{i11} \\ \Delta J_{i12} \\ \Delta J_{i13} \end{bmatrix} + \mathbf{J}_{12} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \mathbf{J}_{13} \begin{bmatrix} \Delta T_{i1} \\ \Delta T_{i2} \\ \Delta T_{i3} \end{bmatrix}, \tag{1.21}$$

where

$$\mathbf{J}_{11} = \left[\frac{\partial x^i}{\partial J_{i11}}, \frac{\partial x^i}{\partial J_{i12}}, \frac{\partial x^i}{\partial J_{i13}} \right]_* \tag{1.22}$$

$$= [\overline{x}, \overline{y}, \overline{z}] \tag{1.23}$$

$$= \underline{m}_x^T, \tag{1.24}$$

$$\mathbf{J}_{12} = \left[\frac{\partial x^i}{\partial x}, \frac{\partial x^i}{\partial y}, \frac{\partial x^i}{\partial z} \right]_* \tag{1.25}$$

$$= [J_{i0_{11}}, J_{i0_{12}}, J_{i0_{13}}], \tag{1.26}$$

and

$$\mathbf{J}_{13} = \left[\frac{\partial x^i}{\partial T_{i1}}, \frac{\partial x^i}{\partial T_{i2}}, \frac{\partial x^i}{\partial T_{i3}} \right]_* \tag{1.27}$$

$$= [1, 0, 0]. \tag{1.28}$$

Similarly, y^i can be expanded as

$$\Delta y^i = \mathbf{J}_{21} \begin{bmatrix} \Delta J_{i21} \\ \Delta J_{i22} \\ \Delta J_{i23} \end{bmatrix} + \mathbf{J}_{22} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \mathbf{J}_{23} \begin{bmatrix} \Delta T_{i1} \\ \Delta T_{i2} \\ \Delta T_{i3} \end{bmatrix}, \tag{1.29}$$

where

$$\mathbf{J}_{21} = \left[\frac{\partial y^i}{\partial J_{i21}}, \frac{\partial y^i}{\partial J_{i22}}, \frac{\partial y^i}{\partial J_{i23}} \right]_* \tag{1.30}$$

$$= \underline{m}_x^T, \tag{1.31}$$

$$\mathbf{J}_{22} = \left[\frac{\partial y^i}{\partial x}, \frac{\partial y^i}{\partial y}, \frac{\partial y^i}{\partial z}\right]_* \tag{1.32}$$

$$= [J_{i0_{21}}, J_{i0_{22}}, J_{i0_{23}}], \tag{1.33}$$

and

$$\mathbf{J}_{23} = \left[\frac{\partial y^i}{\partial T_{i1}}, \frac{\partial y^i}{\partial T_{i2}}, \frac{\partial y^i}{\partial T_{i3}}\right]_* \tag{1.34}$$

$$= [0, 1, 0]. \tag{1.35}$$

For z^i we also have

$$\Delta z^i = \mathbf{J}_{31} \begin{bmatrix} \Delta J_{i31} \\ \Delta J_{i32} \\ \Delta J_{i33} \end{bmatrix} + \mathbf{J}_{32} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \mathbf{J}_{33} \begin{bmatrix} \Delta T_{i1} \\ \Delta T_{i2} \\ \Delta T_{i3} \end{bmatrix}, \tag{1.36}$$

where

$$\mathbf{J}_{31} = \left[\frac{\partial z^i}{\partial J_{i31}}, \frac{\partial z^i}{\partial J_{i32}}, \frac{\partial Z^i}{\partial J_{i33}}\right]_* \tag{1.37}$$

$$= \underline{m}_x^T, \tag{1.38}$$

$$\mathbf{J}_{32} = \left[\frac{\partial z^i}{\partial x}, \frac{\partial z^i}{\partial y}, \frac{\partial z^i}{\partial z}\right]_* \tag{1.39}$$

$$= [J_{i0_{31}}, J_{i0_{32}}, J_{i0_{33}}], \tag{1.40}$$

and

$$\mathbf{J}_{33} = \left[\frac{\partial z^i}{\partial T_{i1}}, \frac{\partial z^i}{\partial T_{i2}}, \frac{\partial z^i}{\partial T_{i3}}\right]_* \tag{1.41}$$

$$= [0, 0, 1]. \tag{1.42}$$

Written in the form of matrix equation, the increment of $\underline{x}$ at the nominal point

can be expressed as

$$\begin{bmatrix} \Delta x^i \\ \Delta y^i \\ \Delta z^i \end{bmatrix} = \mathbf{J}_a \cdot \Delta \underline{J}_i + \mathbf{J}_{i0} \Delta \underline{x} + \mathbf{I} \Delta \underline{T}_i, \tag{1.43}$$

where

$$\mathbf{J}_a = \begin{bmatrix} \mathbf{J}_{11} & 0 & 0 \\ 0 & \mathbf{J}_{21} & 0 \\ 0 & 0 & \mathbf{J}_{31} \end{bmatrix} \tag{1.44}$$

is called Jacobian,

$$\Delta J_i = [\Delta J_{i11} \Delta J_{i12} \Delta J_{i13} \Delta J_{i21} \Delta J_{i22} \Delta J_{i23} \Delta J_{i31} \Delta J_{i32} \Delta J_{i33}]^T, \tag{1.45}$$

$$\Delta \underline{x} = [\Delta x \Delta y \Delta z]^T, \tag{1.46}$$

and

$$\Delta \underline{T}_i = [\Delta T_{i1}, \Delta T_{i2}, \Delta T_{i3}]^T. \tag{1.47}$$

The covariance matrix of $\underline{x}^i$ then is determined by

$$\mathbf{\Psi}^i = E\{[\Delta x^i \Delta y^i \Delta z^i][\Delta x^i \Delta y^i \Delta z^i]^T\} \tag{1.48}$$

$$= \mathbf{J}_{i0} \mathbf{\Psi} \mathbf{J}_{i0}^T + \mathbf{J}_a \mathbf{Q}_{Ji} \mathbf{J}_a^T + \mathbf{Q}_{T_i}. \tag{1.49}$$

Equations (14) and (49) are the basic formulas for uncertainty transformation. Other transformations of uncertainties are evolved from these two basic equations.

A.3 Centralized Multi-Sensor/Information Fusion

In the following, a centralized algorithm for the dynamic sensor systems is derived first and the centralized algorithms for the static sensor systems are derived from the dynamic algorithm.

A.3.1 Centralized Multi-sensor/Information Fusion for Dynamic Sensor Systems

For simplicity, $\underline{w}_k$ and $\underline{v}_k^i$ are assumed to be Gaussian white noise,

$$\underline{w}_k \sim N(0, \mathbf{Q}_k), \;\; \underline{v}_k^i \sim N(0, \mathbf{R}_k^i). \tag{1.50}$$

In the centralized estimation, all the measurements are transmitted to the central processor to update the global estimate. The transformed measurement model of sensor i represented in the reference coordinate system is

$$\begin{aligned}
\underline{z}_{R_{k+1}}^i &= \mathbf{J}_{i_{k+1}}^{-1}(\underline{z}_{k+1}^i - \underline{T}_{i_{k+1}}) && (1.51)\\
&= \mathbf{J}_{i_{k+1}}^{-1}(\mathbf{H}_{k+1}^i(\mathbf{J}_{i_{k+1}}\underline{x}_{k+1} + \underline{T}_{i_{k+1}}) + \underline{v}_{k+1}^i - \underline{T}_{i_{k+1}}) && (1.52)\\
&= \mathbf{J}_{i_{k+1}}^{-1}\mathbf{H}_{k+1}^i\mathbf{J}_{i_{k+1}}\underline{x}_{k+1} + \mathbf{J}_{i_{k+1}}^{-1}\mathbf{H}_{k+1}^i\underline{T}_{i_{k+1}} + \mathbf{J}_{i_{k+1}}^{-1}\underline{v}_{k+1}^i - \mathbf{J}_{i_{k+1}}^{-1}\underline{T}_{i_{k+1}} && (1.53)
\end{aligned}$$

where the subscript R denotes that the quantity is represented in the reference coordinate system. Equation (53) can be written in the form

$$\underline{z}_{R_{k+1}}^i = \mathbf{H}_{R_{k+1}}^i\underline{x}_{k+1} + \underline{v}_{R_{k+1}}^i \tag{1.54}$$

where

$$\mathbf{H}_{R_{k+1}}^i = \mathbf{J}_{i_{k+1}}^{-1}\mathbf{H}_{k+1}^i\mathbf{J}_{i_{k+1}}, \tag{1.55}$$

and

$$\underline{v}_{R_{k+1}}^i = \mathbf{J}_{i_{k+1}}^{-1}\underline{v}_{k+1}^i + \mathbf{J}_{i_{k+1}}^{-1}(\mathbf{H}_{k+1}^i\underline{T}_{i_{k+1}} - \underline{T}_{i_{k+1}}). \tag{1.56}$$

Because of the translation, the equivalent noise $\underline{v}_{R_{k+1}}^i$ does not have zero mean anymore,

$$\underline{m}_{v_{R_{k+1}}}^i = E\{\underline{v}_{R_{k+1}}^i\} = \mathbf{J}_{i0_{k+1}}^{-1}(\mathbf{H}_{k+1}^i\underline{T}_{i0_{k+1}} - \underline{T}_{i0_{k+1}}). \tag{1.57}$$

To acquire zero mean transformed noise, the means are extracted from the transformed noise

$$\tilde{\underline{v}}^i_{R_{k+1}} = \underline{v}^i_{R_{k+1}} - \underline{m}^i_{v_{R_{k+1}}}, \quad i = 1, ..., N. \tag{1.58}$$

The modified transformed measurement model of sensor i becomes

$$\tilde{\underline{z}}^i_{R_{k+1}} = \mathbf{H}^i_{R_{k+1}} \underline{x}_{k+1} + \tilde{\underline{v}}^i_{R_{k+1}}, \quad i = 1, ..., N \tag{1.59}$$

where

$$\tilde{\underline{z}}^i_{R_{k+1}} = \underline{z}^i_{R_{k+1}} - \underline{m}^i_{v_{R_{k+1}}}. \tag{1.60}$$

Combining Equations (56) and (58) yields

$$\underline{v}^i_{k+1} = \mathbf{J}_{i_{k+1}} \underline{v}^i_{R_{k+1}} + \underline{T}_{i_{k+1}} - \mathbf{H}^i_{k+1} \underline{T}_{i_{k+1}} + \mathbf{J}_{i_{k+1}} \underline{m}^i_{v_{R_{k+1}}}. \tag{1.61}$$

The variance of the equivalent noise $\tilde{\underline{v}}^i_{R_{k+1}}$ is calculated by expanding $\underline{v}^i_{k+1}$ into Taylor series and keeping first order approximation

$$\Delta \underline{v}^i_{k+1} = \mathbf{J}_{a_{k+1}} \Delta \mathbf{J}_{i_{k+1}} + \mathbf{J}_{i0_{k+1}} \Delta \underline{v}^i_{R_{k+1}} + \Delta \underline{T}_{i_{k+1}} - \mathbf{H}^i_{k+1} \Delta \underline{T}_{i_{k+1}}, \tag{1.62}$$

or

$$\Delta \underline{v}^i_{R_{k+1}} = \mathbf{J}^{-1}_{i0_{k+1}} \Delta \underline{v}^i_{k+1} - \mathbf{J}^{-1}_{i0_{k+1}} \mathbf{J}_{a_{k+1}} \Delta \mathbf{J}_{i_{k+1}} - \mathbf{J}^{-1}_{i0_{k+1}} (\mathbf{I} - \mathbf{H}^i_{k+1}) \Delta \underline{T}_{i_{k+1}} \tag{1.63}$$

where $\mathbf{J}_{a_{k+1}}$ is a Jacobian. The covariance is then

$$\mathbf{R}^i_{R_{k+1}} = E\{\Delta \underline{v}^i_{R_{k+1}} (\Delta \underline{v}^i_{R_{k+1}})^T\} \tag{1.64}$$

$$= \begin{array}{l} \mathbf{J}^{-1}_{i0_{k+1}} \mathbf{R}^i_{k+1} (\mathbf{J}^{-1}_{i0_{k+1}})^T + \mathbf{J}^{-1}_{i0_{k+1}} \mathbf{J}_{a_{k+1}} \mathbf{Q}_{J_{i_{k+1}}} \mathbf{J}^T_{a_{k+1}} (\mathbf{J}^{-1}_{i0_{k+1}})^T + \\ \mathbf{J}^{-1}_{i0_{k+1}} (\mathbf{I} - \mathbf{H}^i_{k+1}) \mathbf{Q}_{T_{i_{k+1}}} (\mathbf{I} - \mathbf{H}^i_{k+1})^T (\mathbf{J}^{-1}_{i0_{k+1}})^T \end{array}. \tag{1.65}$$

Because of the uncertainties in the mapping functions, the information transmitted from each sensor to the central processor involves an additional error. The portion, in Equation (65),

$$\mathbf{J}_{i0_{k+1}}^{-1}\mathbf{J}_{a_{k+1}}\mathbf{Q}_{J_{i_{k+1}}}\mathbf{J}_{a_{k+1}}^{T}(\mathbf{J}_{i0_{k+1}}^{-1})^{T}+\mathbf{J}_{i0_{k+1}}^{-1}(\mathbf{I}-\mathbf{H}_{k+1}^{i})\mathbf{Q}_{T_{i_{k+1}}}(\mathbf{I}-\mathbf{H}_{k+1}^{i})^{T}(\mathbf{J}_{i0_{k+1}}^{-1})^{T}$$

is due the mapping function's uncertainties. If the mapping function is known exactly, i.e., $\mathbf{Q}_{J_{i_{k+1}}}=0,\ \mathbf{Q}_{T_{i_{k+1}}}=0$, the covariance becomes

$$\mathbf{R}_{R_{k+1}}^{i}=\mathbf{J}_{i0_{k+1}}^{-1}\mathbf{R}_{k+1}^{i}(\mathbf{J}_{i0_{k+1}}^{-1})^{T}. \tag{1.66}$$

The centralized measurement model in the reference coordinate system is obtained by combining N transformed sensor models, Equation (59),

$$\underline{z}_{R_{k+1}}=\mathbf{H}_{R_{k+1}}\underline{x}_{k+1}+\underline{v}_{R_{k+1}},\ \ E\{\underline{v}_{R_{k+1}}\}=0,\ \ E\{\underline{v}_{R_{k+1}}\underline{v}_{R_{k+1}}^{T}\}=\mathbf{R}_{R_{k+1}}, \tag{1.67}$$

where

$$\underline{z}_{R_{k+1}} = [(\underline{\tilde{z}}_{R_{k+1}}^{1})^{T},(\underline{\tilde{z}}_{R_{k+1}}^{2})^{T},\cdots,(\underline{\tilde{z}}_{R_{k+1}}^{N})^{T}]^{T}\in\mathcal{R}^{3N}, \tag{1.68}$$

$$\mathbf{H}_{R_{k+1}} = [(\mathbf{H}_{R_{k+1}}^{1})^{T},(\mathbf{H}_{R_{k+1}}^{2})^{T},\cdots,(\mathbf{H}_{R_{k+1}}^{N})^{T}]^{T}\in\mathcal{R}^{3N\times 3N}, \tag{1.69}$$

$$\underline{v}_{R_{k+1}} = [(\underline{\tilde{v}}_{R_{k+1}}^{1})^{T},(\underline{\tilde{v}}_{R_{k+1}}^{2})^{T},\cdots,(\underline{\tilde{v}}_{R_{k+1}}^{N})^{T}]^{T}\in\mathcal{R}^{3N}, \tag{1.70}$$

and

$$\mathbf{R}_{R_{k+1}}=diag\{\mathbf{R}_{R_{k+1}}^{1},\mathbf{R}_{R_{k+1}}^{2},\cdots,\mathbf{R}_{R_{k+1}}^{N}\}\in\mathcal{R}^{3N\times 3N}. \tag{1.71}$$

After forming the centralized measurement model, the Kalman filter algorithm can be applied to the centralized dynamic sensor system

$$\begin{aligned}\underline{x}_{k+1} &= \mathbf{\Phi}_{k}(t_{k+1},t_{k})\underline{x}_{k}+\mathbf{\Gamma}_{k}(t_{k})\underline{w}_{k},\ \ E\{w_{k}\}=0,\ \ E\{\underline{w}_{k}\underline{w}_{k}^{T}\}=\mathbf{Q}_{k},\\ \underline{z}_{R_{k}} &= \mathbf{H}_{R_{k}}\underline{x}_{k}+\underline{v}_{R_{k}},\ \ E\{\underline{v}_{R_{k}}\}=0,\ \ E\{\underline{v}_{R_{k}}\underline{v}_{R_{k}}^{T}\}=\mathbf{R}_{R_{k}}.\end{aligned} \tag{1.72}$$

The algorithm involves:

- Initial conditions:

$$\hat{\underline{x}}_{0|-1} = \overline{\underline{x}}_0, \tag{1.73}$$

and

$$\underline{\mathbf{P}}_{0|-1} = E\{(\underline{x} - \overline{\underline{x}}_0)(\underline{x} - \overline{\underline{x}}_0)^T = \mathbf{P}_0. \tag{1.74}$$

- State updated equation

$$\begin{aligned} \hat{\underline{x}}_{k+1|k+1} &= \hat{\underline{x}}_{k+1|k} + \mathbf{K}_{k+1}(\underline{z}_{R_{k+1}} - \mathbf{H}_{R_{k+1}}\hat{\underline{x}}_{k+1|k}) && (1.75) \\ &= \hat{\underline{x}}_{k+1|k} + \sum_{i=1}^{N} \mathbf{K}^i_{k+1}(\underline{z}^i_{R_{k+1}} - \mathbf{H}^i_{R_{k+1}}\hat{\underline{x}}_{k+1|k}), && (1.76) \end{aligned}$$

where

$$\mathbf{K}_{k+1} = diag\{\mathbf{K}^1_{k+1}, \mathbf{K}^2_{k+1}, ..., \mathbf{K}^N_{k+1}\}. \tag{1.77}$$

- Gain equation

$$\begin{aligned} \mathbf{K}^i_{k+1} &= \mathbf{P}^i_{k+1|k}(\mathbf{H}^i_{R_{k+1}})^T[\mathbf{H}^i_{R_{k+1}}\mathbf{P}_{k+1|k}(\mathbf{H}^i_{R_{k+1}})^T + \mathbf{R}^i_{R_{k+1}}]^{-1} && (1.78) \\ &= \mathbf{P}_{k+1|k+1}(\mathbf{H}^i_{R_{k+1}})^T(\mathbf{R}^i_{R_{k+1}})^{-1}. && (1.79) \end{aligned}$$

- Error covariance updated equation

$$\mathbf{P}^{-1}_{k+1|k+1} = \mathbf{P}^{-1}_{k+1|k} + \sum_{i=1}^{N}((\mathbf{H}^i_{R_{k+1}})^T(\mathbf{R}^i_{R_{k+1}})^{-1}\mathbf{H}^i_{R_{k+1}}). \tag{1.80}$$

A dynamic centralized estimator is shown in Figure A.3. Note that the dimension of the centralized measurement model, Equation (67), increases with

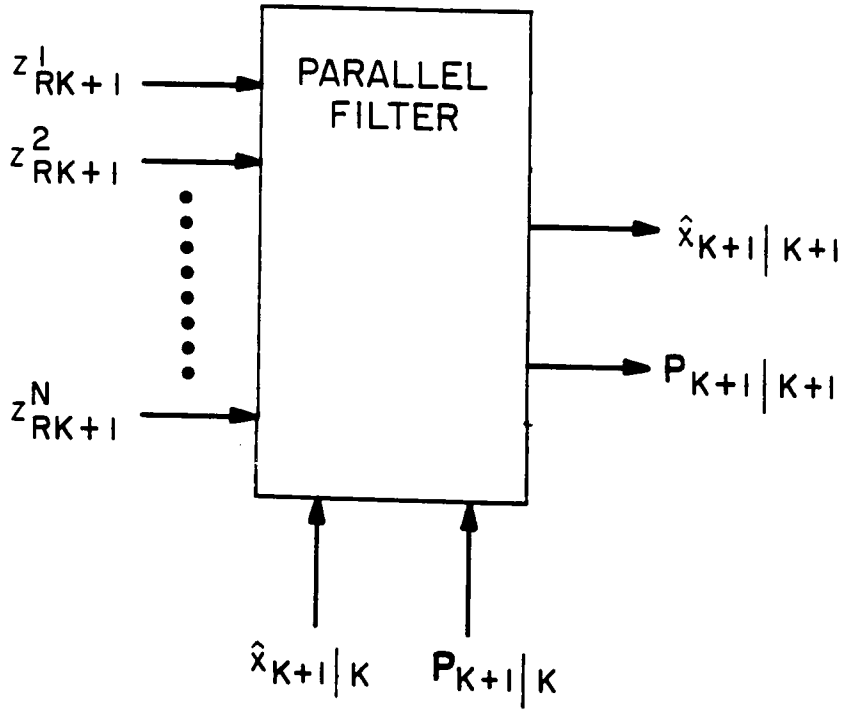


Figure A.3: A dynamic centralized estimator.

the number of sensors. Therefore, when the number of sensors is large, the dimension of the centralized model is very high. Since the computational burden is approximately square-proportional to the dimension of the model, the distributed estimation is preferred for the cases where the number of sensors is large, in order to achieve the computational efficiency.

A.3.2 Centralized Multi-sensor/Information Fusion for Static Sensor Systems

A static multi-sensor system can be considered as a special case of a dynamic sensor system. A static sensor system is given in Equation (5) and the corresponding communication networks are specified by Equation (6). The fusion algorithm for the dynamic sensor systems derived in the preceding Section is applicable to the static sensor systems. The only difference is that the dynamic algorithm has subscript k and the static algorithm does not. The following equiv-

alences are used in the derivation:

Table A.1

Static algorithm		Dynamic algorithm
$\underline{x}$	$\longleftrightarrow$	$\underline{x}_k,\ \underline{x}_{k+1}$
$\underline{m}_x$	$\longleftrightarrow$	$\underline{\hat{x}}_{0\|-1},\ \underline{\hat{x}}_{k\|k},\ \underline{\hat{x}}_{k+1\|k}$
$\underline{m}_{x^i}$	$\longleftrightarrow$	$\underline{\hat{x}}^i_{k+1\|k}$
$\mathbf{\Psi}$	$\longleftrightarrow$	$\mathbf{P}_0,\ \mathbf{P}_{k\|k},\ \mathbf{P}_{k+1\|k}$
$\mathbf{\Psi}^i$	$\longleftrightarrow$	$\mathbf{P}^i_{k+1\|k}$
$\underline{\hat{x}}^i$	$\longleftrightarrow$	$\underline{\hat{x}}_{k+1\|k+1}$
$\underline{\hat{x}}^i_R$	$\longleftrightarrow$	$\underline{\hat{x}}^i_{R_{k+1\|k+1}}$
$\sum$	$\longleftrightarrow$	$\mathbf{P}_{k+1\|k+1}$
$\sum^i$	$\longleftrightarrow$	$\mathbf{P}^i_{k+1\|k+1}$

Similar to the algorithm for the dynamic sensor systems, all the measurements are transmitted to the center processor in the reference coordinate system. Then fusion is done at the center processor, Figure A.4. The transmitted measurements in the reference coordinate system are, Equation (53),

$$\underline{z}^i_R = \mathbf{J}_i^{-1}(\underline{z}^i - \underline{T}_i) \tag{1.81}$$

$$= \mathbf{J}_i^{-1}\mathbf{H}^i\mathbf{J}_i\underline{x} + \mathbf{J}_i^{-1}(\mathbf{H}^i\underline{T}_i - \underline{T}_i) + \mathbf{J}_i^{-1}\underline{v}^i,\ \ i = 1,\ldots,N. \tag{1.82}$$

After the transmission, each sensor is modeled with respect to the reference coordinate system as, Equation (54),

$$\underline{z}^i_R = \mathbf{H}^i_R\underline{x} + \underline{v}^i_R\ \ (i = 1,\ldots,N), \tag{1.83}$$

where

$$\mathbf{H}_R^i = \mathbf{J}_i^{-1}\mathbf{H}^i\mathbf{J}_i \tag{1.84}$$

and

$$\underline{v}_R^i = \mathbf{J}_i^{-1}(\mathbf{H}^i\underline{T}_i - \underline{T}_i) + \mathbf{J}_i^{-1}\underline{v}^i. \tag{1.85}$$

The centralized modeling error $\underline{v}_R^i$ is described by the mean vector, Equation (57),

$$\underline{m}_{v_R}^i = E\{\underline{v}_R^i\} = \mathbf{J}_{i0}^{-1}(\mathbf{H}^i - \mathbf{I})\underline{T}_{i0} \neq 0, \tag{1.86}$$

and by the covariance matrix, Equation (65),

$$\mathbf{R}_{v_R}^i = \mathbf{J}_{i0}^{-1}\mathbf{R}^i(\mathbf{J}_{i0}^{-1})^T + \mathbf{J}_{i0}^{-1}\mathbf{J}_{a_{v^i}}\mathbf{Q}_{J_i}\mathbf{J}_{a_{v^i}}^T(\mathbf{J}_{i0}^{-1})^T + \mathbf{J}_{i0}^{-1}(\mathbf{H}^i - \mathbf{I})\mathbf{Q}_{T_i}(\mathbf{H}^i - \mathbf{I})^T(\mathbf{J}_{i0}^{-1})^T \tag{1.87}$$

where $\mathbf{J}_{a_{v^i}}$ is a Jacobian.

A centralized measurement model is formed in the reference coordinate system, Equation (67),

$$\underline{z}_R = \mathbf{H}_R\underline{x} + \underline{v}_R, \tag{1.88}$$

where

$$\underline{z}_R = [(\underline{\tilde{z}}_R^1)^T,\ (\underline{\tilde{z}}_R^2)^T,\ \ldots,\ (\underline{\tilde{z}}_R^N)^T]^T \quad \in \mathcal{R}^{3N}, \tag{1.89}$$

$$\mathbf{H}_R = [(\mathbf{H}_R^1)^T,\ (\mathbf{H}_R^2)^T, \ldots, (\mathbf{H}_R^N)^T]^T \quad \in \mathcal{R}^{3N\times 3N}, \tag{1.90}$$

$$\underline{v}_R = [(\underline{\tilde{v}}_R^1)^T,\ (\underline{\tilde{v}}_R^2)^T,\ \ldots,\ (\underline{\tilde{v}}_R^N)^T]^T \quad \in \mathcal{R}^{3N},\ \ \underline{v}_R \sim (0, \mathbf{R}_{v_R}). \tag{1.91}$$

The elements $\underline{\tilde{z}}_R^i$ and $\underline{\tilde{v}}_R^i$, $i = 1, ..., N$ in Equations (89) and (91) are given by

$$\underline{\tilde{z}}_R^i = \underline{z}_R^i - \underline{m}_{v_R}^i,\ \ i = 1, ..., N, \tag{1.92}$$

and

$$\tilde{\underline{v}}_R^i = \underline{v}_R^i - \underline{m}_{v_R}^i, \quad i = 1, ..., N, \tag{1.93}$$

to yield a centralized modeling error $\underline{v}_R$ with zero mean. The covariance matrix $\mathbf{R}_{v_R}$ is a block diagonal matrix

$$\mathbf{R}_{v_R} = diag\{\mathbf{R}_{v_R}^1, \mathbf{R}_{v_R}^2, \ldots, \mathbf{R}_{v_R}^N\} \in \mathcal{R}^{3N \times 3N}. \tag{1.94}$$

Either the Bayesian estimator or the Fisher estimator can be used to fuse the information provided by N sensors, depending on the prior information about the quantity $\underline{x}$.

A.3.2.1 Static Centralized Bayesian Estimators:

When the prior information about the quantity $\underline{x}$, given by

$$E\{\underline{x}\} = \underline{m}_x, \quad E\{(\underline{x} - \underline{m}_x)(\underline{x} - \underline{m}_x)^T\} = \mathbf{\Psi}, \tag{1.95}$$

is available, a Bayesian estimator is used. The Bayesian estimator for the centralized estimation is, Equations (75) and (78),

$$\hat{\underline{x}} = \underline{m}_x + \mathbf{K}(\underline{z}_R - \mathbf{H}_R \underline{m}_x), \tag{1.96}$$

where

$$\mathbf{K} = \mathbf{\Psi}\mathbf{H}_R^T(\mathbf{H}_R\mathbf{\Psi}\mathbf{H}_R^T + \mathbf{R}_{v_R})^{-1} \tag{1.97}$$

with a minimum error covariance, Equation (80),

$$\sum = \mathbf{\Psi} - \mathbf{\Psi}\mathbf{H}_R^T(\mathbf{H}_R\mathbf{\Psi}\mathbf{H}_R^T + \mathbf{R}_{v_R})^{-1}\mathbf{H}_R\mathbf{\Psi}. \tag{1.98}$$

A.3.2.2 Static Centralized Fisher Estimators:

When the prior information for $\underline{x}$ is unknown, Fisher estimators are used

$$\hat{\underline{x}} = (\mathbf{H}_R^T \mathbf{R}_{v_R}^{-1} \mathbf{H}_R)^{-1} \mathbf{H}_R^T \mathbf{R}_{v_R}^{-1} \underline{z}_R \tag{1.99}$$

which gives a minimum error covariance

$$\sum = (\mathbf{H}_R^T \mathbf{R}_{v_R}^{-1} \mathbf{H}_R)^{-1}. \tag{1.100}$$

The computational requirement for the centralized estimator, described above, is high, especially for the system with many sensors.

A.4 Distributed Multi-Sensor/Information Fusion

To obtain the computational efficiency, the distributed estimation for sensor/in-

formation fusion is studied for both dynamic and static sensor systems.

A.4.1 Distributed Multi-Sensor/Information Fusion for Dynamic Sensor Systems

In the distributed algorithm for the dynamic sensor systems, the information gathered by sensor i, $i = 1, ..., N$, at t_{k+1} moment, together with the result of central estimation at t_k transmitted from the communication network, is locally processed to generate the local estimates. Then, the statistics of the locally estimated results are transmitted to the central processor to generate the globally estimated result at t_{k+1}. The algorithm involves the following steps, beginning

with initial conditions:

$$\hat{\underline{x}}_{0|-1} = \overline{x}_0, \quad \mathbf{P}_{0|-1} = \mathbf{P}_0. \tag{1.101}$$

- Propagation of centrally estimated result from t_k to t_{k+1}

$$\hat{\underline{x}}_{k+1|k} = \mathbf{\Phi}_k \hat{\underline{x}}_{k|k}, \tag{1.102}$$

$$\mathbf{P}_{k+1|k} = \mathbf{\Phi}_k \mathbf{P}_{k|k} \mathbf{\Phi}_k^T + \mathbf{\Gamma}_k \mathbf{Q}_k \mathbf{\Gamma}_k^T, \tag{1.103}$$

- Transmission of the centrally estimated result to the local processors

  The transmitted state is

$$\hat{\underline{x}}^i_{k+1|k} = \mathbf{J}_{i_{k+1}} \hat{\underline{x}}_{k+1|k} + \underline{T}_{i_{k+1}} \tag{1.104}$$

  with a transmitted error covariance

$$\mathbf{P}^i_{k+1|k} = \mathbf{J}_{i0_{k+1}} \mathbf{P}_{k+1|k} \mathbf{J}^T_{i0_{k+1}} + \mathbf{J}_{a_{k+1}} \mathbf{Q}_{J_{i_{k+1}}} \mathbf{J}^T_{a_{k+1}} + \mathbf{Q}_{T_{i_{k+1}}}, \tag{1.105}$$

  where $\mathbf{J}_{a_{k+1}}$ is a Jacobian. Note that $\hat{\underline{x}}^i_{k+1|k}$ was estimated at central processor based on the information gathered by the entire sensor system. Therefore, $\hat{\underline{x}}^i_{k+1|k}$ reflects the information passed from other sensors.

- Local estimation

  Local gain

$$\mathbf{K}^i_{k+1} = \mathbf{P}^i_{k+1|k} (\mathbf{H}^i_{k+1})^T [\mathbf{H}^i_{k+1} \mathbf{P}^i_{k+1|k} (\mathbf{H}^i_{k+1})^T + \mathbf{R}^i_{k+1}]^{-1} \tag{1.106}$$

$$= \mathbf{P}^i_{k+1|k+1} (\mathbf{H}^i_{k+1})^T (\mathbf{R}^i_{k+1})^{-1}. \tag{1.107}$$

Local state update

$$\underline{\hat{x}}^i_{k+1|k+1} = [\mathbf{I} - \mathbf{K}^i_{k+1}\mathbf{H}^i_{k+1}]\underline{\hat{x}}^i_{k+1|k} + \mathbf{K}^i_{k+1}\underline{z}^i_{k+1}. \tag{1.108}$$

Local error covariance update

$$(\mathbf{P}^i_{k+1|k+1})^{-1} = (\mathbf{P}^i_{k+1|k})^{-1} + (\mathbf{H}^i_{k+1})^T(\mathbf{R}^i_{k+1})^{-1}\mathbf{H}^i_{k+1} \tag{1.109}$$

- Transmission of the local estimates to the central processor at reference coordinate system

$$\underline{\hat{x}}^i_{R_{k+1|k+1}} = \mathbf{J}_{i_{k+1}}(\underline{\hat{x}}^i_{k+1|k+1} - \underline{T}_{i_{k+1}}) \tag{1.110}$$

and

$$\begin{aligned}\mathbf{P}^i_{R_{k+1|k+1}} = \; & \mathbf{J}^{-1}_{i0_{k+1}}\mathbf{P}^i_{k+1|k+1}(\mathbf{J}^{-1}_{i0_{k+1}})^T + \\ & + \mathbf{J}^{-1}_{i0_{k+1}}(\mathbf{J}_{a1_{k+1}}\mathbf{Q}_{J_{i_{k+1}}}\mathbf{J}^T_{a1_{k+1}} + \mathbf{J}_{a2_{k+1}}\mathbf{Q}_{J_{i_{k+1}}}\mathbf{J}^T_{a2_{k+1}})(\mathbf{J}^{-1}_{i0_{k+1}})^T \\ & + 2\mathbf{J}^{-1}_{i0_{k+1}}\mathbf{Q}_{T_{i_{k+1}}}(\mathbf{J}^{-1}_{i0_{k+1}})^T,\end{aligned} \tag{1.111}$$

where $\mathbf{J}_{a1_{k+1}}$ contains the derivatives of $\underline{\hat{x}}^i_{k+1|k+1}$, and $\mathbf{J}_{a2_{k+1}}$ contains the derivatives of $\underline{x}^i_{k+1}$, with respect to the elements of $\mathbf{J}_{i_k}$.

- Global estimation:

To derive the equations for fusing transmitted local estimates, we first write out the centralized update equations:

Centralized gain

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1|k}\mathbf{H}^T_{R_{k+1}}[\mathbf{H}_{R_{k+1}}\mathbf{P}_{k+1|k}\mathbf{H}^T_{R_{k+1}} + \mathbf{R}_{R_{k+1}}]^T \tag{1.112}$$

$$= \mathbf{P}_{k+1|k+1}\mathbf{H}^T_{R_{k+1}}\mathbf{R}^{-1}_{R_{k+1}} \tag{1.113}$$

where

$$\mathbf{H}_{R_{k+1}} = [(\mathbf{H}^1_{R_{k+1}})^T, \cdots, (\mathbf{H}^N_{R_{k+1}})^T]^T, \tag{1.114}$$

and

$$\mathbf{R}_{R_{k+1}} = diag\{\mathbf{R}^1_{R_{k+1}}, \cdots, \mathbf{R}^N_{R_{k+1}}\}. \tag{1.115}$$

Centralized covariance update

$$\mathbf{P}_{k+1|k+1} = [\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{R_{k+1}}]\mathbf{P}_{k+1|k}. \tag{1.116}$$

Centralized state update

$$\underline{\hat{x}}_{k+1|k} = \underline{\hat{x}}_{k+1|k} + \mathbf{K}_{k+1}[\underline{\tilde{z}}_{R_{k+1}} - \mathbf{H}_{R_{k+1}}\underline{\hat{x}}_{k+1|k}] \tag{1.117}$$

$$= [\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{R_{k+1}}]\underline{\hat{x}}_{k+1|k} + \mathbf{K}_{k+1}\underline{\tilde{z}}_{R_{k+1}}. \tag{1.118}$$

Since

$$\mathbf{K}_{k+1}\underline{\tilde{z}}_{R_{k+1}} = \mathbf{P}_{k+1|k+1}[(\mathbf{H}^1_{R_{k+1}})^T, \cdots, (\mathbf{H}^N_{R_{k+1}})^T]diag\{(\mathbf{R}^1_{R_{k+1}})^{-1}, \cdots, (\mathbf{R}^N_{R_{k+1}})^{-1}\}\underline{\tilde{z}}_{R_k} \tag{1.119}$$

$$= \mathbf{P}_{k+1|k+1}\sum_{i=1}^{N}(\mathbf{H}^i_{R_{k+1}})^T(\mathbf{R}^i_{R_{k+1}})^{-1}\underline{\tilde{z}}^i_{R_{k+1}}, \tag{1.120}$$

to eliminate $\underline{\tilde{z}}^i_{R_{k+1}}$, it follows from Equation (118) that

$$\underline{\hat{x}}^i_{R_{k+1|k+1}} = [\mathbf{I} - \mathbf{K}^i_{R_{k+1}}\mathbf{H}^i_{R_{k+1}}]\underline{\hat{x}}_{k+1|k} + \mathbf{K}^i_{R_{k+1}}\underline{\tilde{z}}^i_{R_{k+1}}, \tag{1.121}$$

where

$$\mathbf{K}^i_{R_{k+1}} = \mathbf{P}^i_{R_{k+1|k+1}}(\mathbf{H}^i_{R_{k+1}})^T(\mathbf{R}^i_{R_{k+1}})^{-1}. \tag{1.122}$$

Combining Equations (121) and (122), we have

$$(\mathbf{H}^i_{R_{k+1}})^T(\mathbf{R}^i_{R_{k+1}})^{-1}\underline{\tilde{z}}^i_{R_{k+1}} = (\mathbf{P}^i_{R_{k+1|k+1}})^{-1}\underline{\hat{x}}^i_{R_{k+1}} - (\mathbf{P}^i_{R_{k+1|k+1}})^{-1}[\mathbf{I} - k^i_{R_{k+1}}\mathbf{H}^i_{R_{k+1}}]\underline{\hat{x}}_{k+1|k}. \tag{1.123}$$

Now, since

$$\mathbf{I} - \mathbf{K}^i_{R_{k+1}} \mathbf{H}^i_{R_{k+1}} = \mathbf{P}^i_{R_{k+1|k+1}} \mathbf{P}^{-1}_{k+1|k}, \tag{1.124}$$

the following holds

$$(\mathbf{H}^i_{R_{k+1}})^T (\mathbf{R}^i_{R_{k+1}})^{-1} \tilde{z}^i_{R_{k+1}} = (\mathbf{P}^i_{R_{k+1|k+1}})^{-1} \hat{\underline{x}}^i_{R_{k+1}} - \mathbf{P}^{-1}_{k+1|k} \hat{\underline{x}}_{k+1|k}. \tag{1.125}$$

Finally, the global estimate update equation is

$$\hat{\underline{x}}_{k+1|k+1} = \mathbf{P}_{k+1|k+1} [\mathbf{P}^{-1}_{k+1|k} \hat{\underline{x}}_{k+1|k} + \sum_{i=1}^{N} ((\mathbf{P}^i_{R_{k+1|k+1}})^{-1} \hat{\underline{x}}^i_{R_{k+1|k+1}} - \mathbf{P}^{-1}_{k+1|k} \hat{\underline{x}}_{k+1|k})] \tag{1.126}$$

$$= \mathbf{P}_{k+1|k+1} [\sum_{i=1}^{N} (\mathbf{P}^i_{R_{k+1|k+1}})^{-1} \hat{\underline{x}}^i_{R_{k+1|k+1}} - (N-1) \mathbf{P}^{-1}_{k+1|k} \hat{\underline{x}}_{k+1|k}], \tag{1.127}$$

and the update equation for the associated global error covariance is

$$\mathbf{P}^{-1}_{k+1|k+1} = \mathbf{P}^{-1}_{k+1|k} + \sum_{i=1}^{N} ((\mathbf{P}^i_{R_{k+1|k}})^{-1} - \mathbf{P}^{-1}_{k+1|k}) \tag{1.128}$$

$$= \sum_{i=1}^{N} (\mathbf{P}^i_{R_{k+1|k+1}})^{-1} - (N-1) \mathbf{P}^{-1}_{k+1|K} \tag{1.129}$$

A signal flow chart for the dynamic distributed estimation is shown in Figure A.4.

A.4.2 Distributed Multi-Sensor/Information Fusion for Static Sensor Systems

By considering a static sensor system as a special case of a dynamic sensor system, the following algorithm for the static sensor systems is derived with the equivalences given in Table A.1.

A.4.2.1 Static Distributed Bayesian Estimators

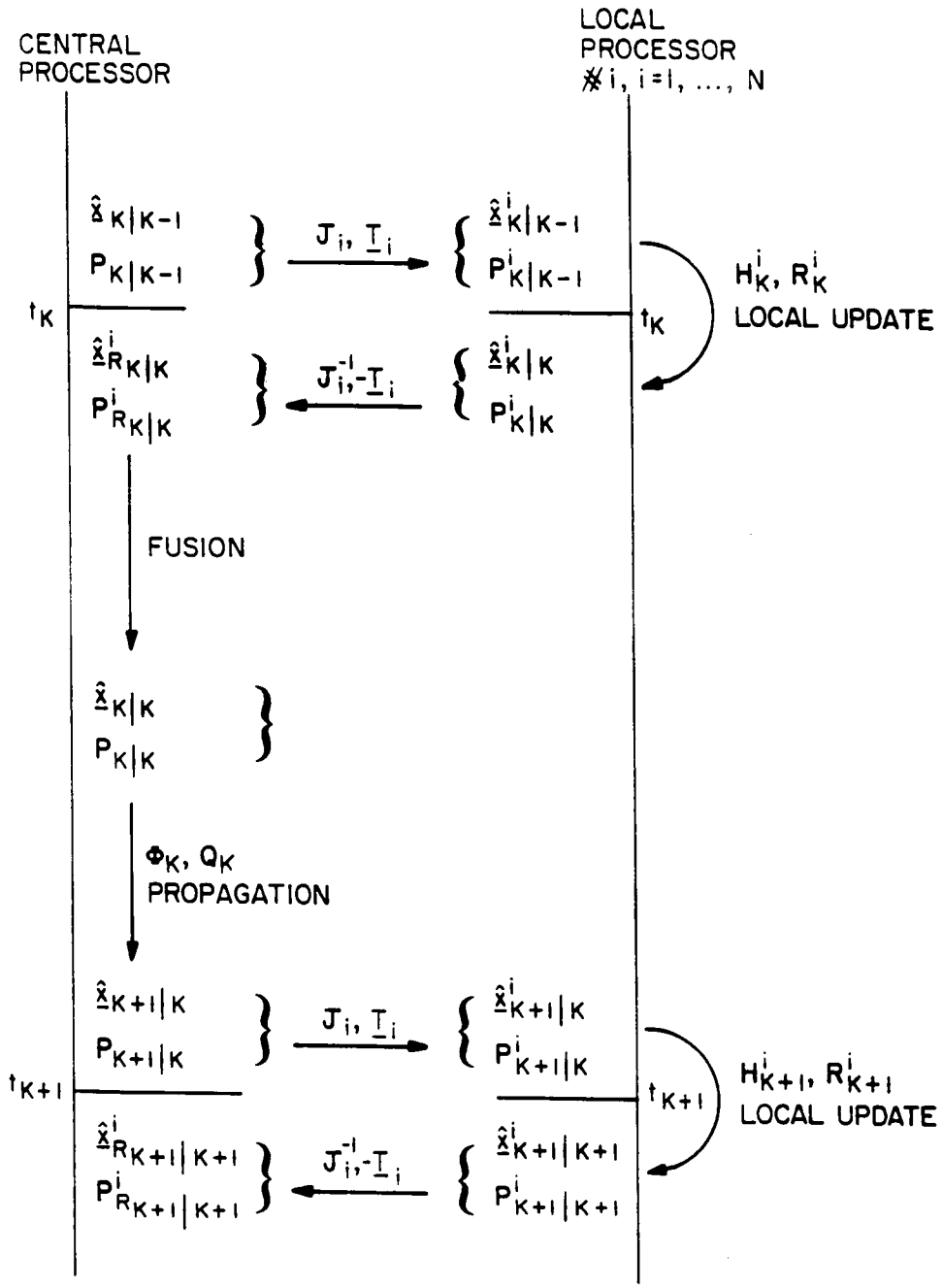


Figure A.4: A signal flow chart for dynamic multi-sensor/information fusion.

The algorithm of multi-sensor/information fusion for static sensor system involves the following steps:

- Local estimation that incorporates:

  1. Transmission of the prior information to local processors, Equations (104) and (105),

$$\underline{m}_{x^i} = \mathbf{J}_{i0}\underline{m}_x + \underline{T}_{i0}, \tag{1.130}$$

  and

$$\begin{aligned} \mathbf{\Psi}^i &= E\{(\underline{x}^i - \underline{m}_{x^i})(\underline{x}^i - \underline{m}_{x^i})^T\} && (1.131)\\ &= \mathbf{J}_{i0}\mathbf{\Psi}\mathbf{J}_{i0}^T + \mathbf{J}_a\mathbf{Q}_{Ji}\mathbf{J}_a^T + \mathbf{Q}_{T_i}, && (1.132) \end{aligned}$$

  where $\mathbf{J}_a$ is a Jacobian which contains the derivatives of $\underline{x}^i$ with respect to the elements of $\mathbf{J}_i$.

  2. Calculation of local gain, Equation (106),

$$\mathbf{K}^i = \mathbf{\Psi}^i(\mathbf{H}^i)^T[\mathbf{H}^i\mathbf{\Psi}^i(\mathbf{H}^i)^T + \mathbf{R}_i]^{-1}. \tag{1.133}$$

  3. Calculation of local Bayesian estimates, Equation (108),

$$\begin{aligned} \hat{\underline{x}}^i &= \underline{m}_{x^i} + \mathbf{K}^i[\underline{z}^i - \mathbf{H}^i\underline{m}_{x^i}] && (1.134)\\ &= \underline{m}_{x^i} + \mathbf{\Psi}^i(\mathbf{H}^i)^T[\mathbf{H}^i\mathbf{\Psi}^i(\mathbf{H}^i)^T + \mathbf{R}_i]^{-1}(\underline{z}^i - \mathbf{H}^i\underline{m}_{x^i}) && (1.135)\\ &= \underline{x}^i + \underline{\delta}_{x^i}. && (1.136) \end{aligned}$$

  4. Calculation of local minimum covariances, Equation (109),

$$\sum^i = (\mathbf{I} - \mathbf{K}^i\mathbf{H}^i)\mathbf{\Psi}^i \tag{1.137}$$

$$= \Psi^i - \Psi^i(\mathbf{H}^i)^T[\mathbf{H}^i\Psi^i(\mathbf{H}^i)^T + \mathbf{R}_i]^{-1}\mathbf{H}^i\Psi^i \quad (1.138)$$

$$= E\{\underline{\delta}_{x^i}\underline{\delta}_{x^i}^T\}. \quad (1.139)$$

– Global estimation that incorporates:

1. Transmission of local estimates to the central processor, Equations (110) and (111),

$$\underline{\hat{x}}_R^i = \mathbf{J}_i^{-1}(\underline{\hat{x}}^i - \underline{T}_i), \quad (1.140)$$

and

$$\sum_R^i = E\{(\underline{x} - \underline{\hat{x}}_R^i)(\underline{x} - \underline{\hat{x}}_R^i)^T\} \quad (1.141)$$

$$= \begin{array}{l} \mathbf{J}_{i0}^{-1}\sum^i(\mathbf{J}_{i0}^{-1})^T + \mathbf{J}_{i0}^{-1}(\mathbf{J}_{a1}\mathbf{Q}_{J_i}\mathbf{J}_{a1}^T + \mathbf{J}_{a2}\mathbf{Q}_{J_i}\mathbf{J}_{a2}^T)(\mathbf{J}_{i0}^{-1})^T \\ +2\mathbf{J}_{i0}^{-1}\mathbf{Q}_{T_i}(\mathbf{J}_{i0}^{-1})^T \end{array} \quad (1.142)$$

2. Calculation of global estimate, Equation (127),

$$\underline{\hat{x}} = \sum[(\sum_R^1)^{-1}\underline{\hat{x}}_R^1 + \ldots + (\sum_R^N)^{-1}\underline{\hat{x}}_R^N - (N-1)\Psi^{-1}\underline{m}_x] \quad (1.143)$$

3. Calculation of global minimum error covariance, Equation (129),

$$\sum^{-1} = (\sum_R^1)^{-1} + \ldots + (\sum_R^N)^{-1} - (N-1)\Psi^{-1}. \quad (1.144)$$

A diagram of signal transmission is shown in Figure A.5. When the prior information is not available, the Fisher estimator is used to give optimal sensor fusion.

A.4.2.2 Static Distributed Fisher Estimators:

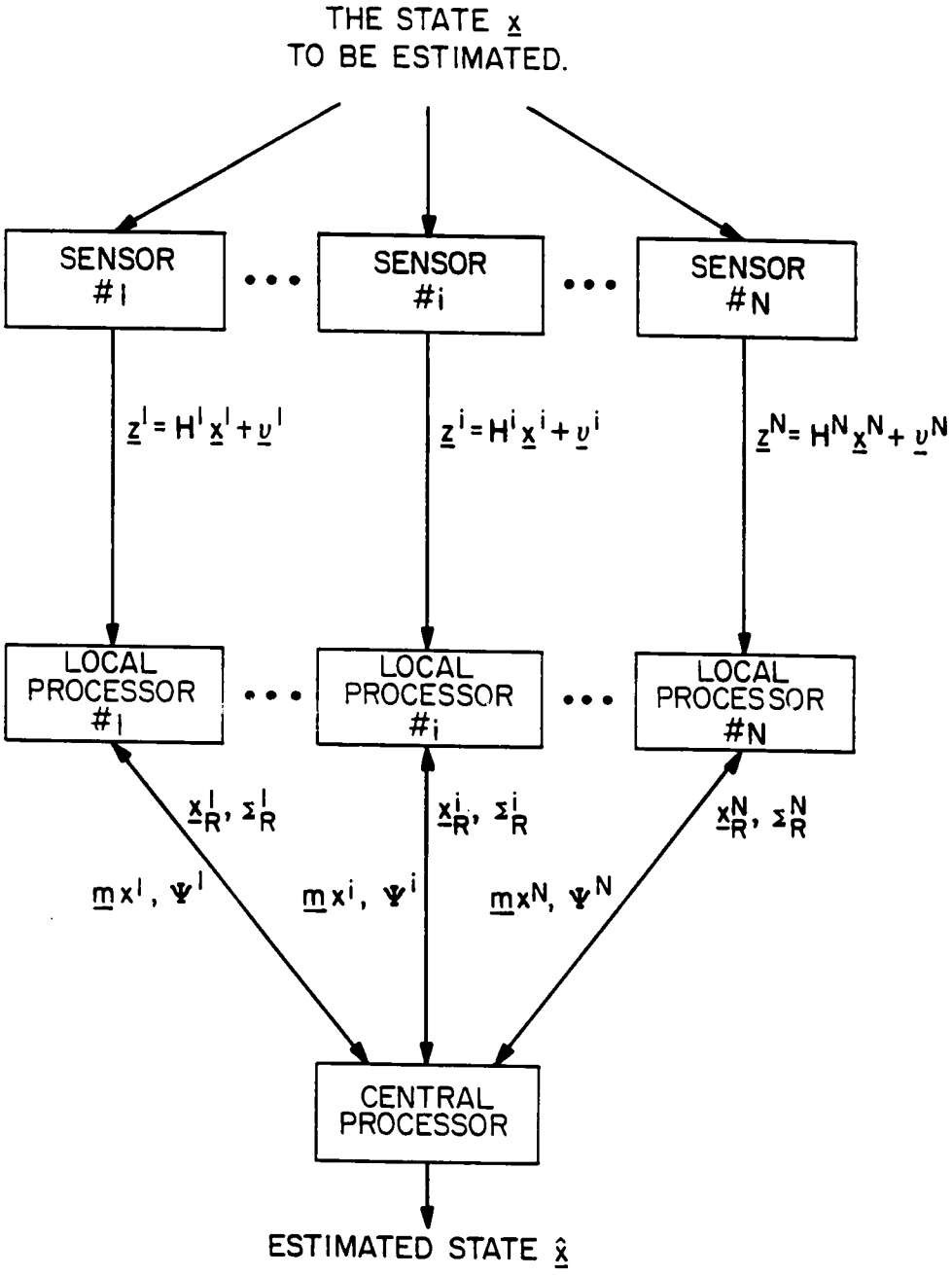


Figure A.5: The static distributed estimation.

When no prior information of $\underline{x}$ is available, Fisher estimators are used. In this case, the steps are:

- Local estimation

  Local processors give the local estimates based on the local information,

$$\underline{\hat{x}}^i = [(\mathbf{H}^i)^T(\mathbf{R}^i)^{-1}\mathbf{H}^i]^{-1}(\mathbf{H}^i)^T(\mathbf{R}^i)^{-1}\underline{z}^i, \quad (1.145)$$

$$= \underline{x}^i + \underline{\delta}_{x^i}, \quad (1.146)$$

  which gives a minimum error covariance

$$\sum^i = [(\mathbf{H}^i)^T(\mathbf{R}^i)^{-1}\mathbf{H}^i]^{-1}, \quad (1.147)$$

$$= E\{\underline{\delta}_{x^i}(\underline{\delta}_{x^i})^T\}. \quad (1.148)$$

- Central estimation

  Local estimates have to be transmitted together with their error covariances to the central processor at reference coordinate system using

$$\underline{\hat{x}}_R^i = \mathbf{J}_i^{-1}(\underline{\hat{x}}^i - \underline{T}_i), \quad (1.149)$$

  and

$$\sum_R^i = E\{(\underline{x} - \underline{\hat{x}}_R^i)(\underline{x} - \underline{\hat{x}}_R^i)^T\} \quad (1.150)$$

$$= \begin{array}{l} \mathbf{J}_{i0}^{-1}\sum^i(\mathbf{J}_{i0}^{-1})^T + \mathbf{J}_{i0}^{-1}(\mathbf{J}_{a1}\mathbf{Q}_{J_i}\mathbf{J}_{a1}^T + \mathbf{J}_{a2}\mathbf{Q}_{J_i}\mathbf{J}_{a2}^T)(\mathbf{J}_{i0}^{-1})^T \\ +2\mathbf{J}_{i0}^{-1}\mathbf{Q}_{T_i}(\mathbf{J}_{i0}^{-1})^T \end{array} \quad (1.151)$$

  The global estimate for the Fisher estimator is

$$\underline{\hat{x}} = [(\sum_R^1)^{-1} + \ldots + (\sum_R^N)]^{-1}[(\sum_R^1)^{-1}\underline{\hat{x}}_R^1 + \ldots + (\sum_R^N)^{-1}\underline{\hat{x}}_R^N], \quad (1.152)$$

which is an optimal combination of local estimates and is with a minimum error covariance for the global estimate

$$\sum = [(\sum\nolimits_R^1)^{-1} + \ldots + (\sum\nolimits_R^N)^{-1}]^{-1}. \tag{1.153}$$

A.5 Discussion

In summary:

- Centralized estimation involves a high dimension central processor. In distributed estimation, local processors can operate in parallel and the dimension of the central processor is low.
- Centralized estimation requires more complicated communication networks.
- When the networks have uncertainties, distributed estimation introduces additional errors.
- For some applications, different types of measurement quantities require distributed estimation.

The overall criterion for choosing centralized or distributed fusion algorithms is that the distributed fusion algorithms are preferred for the cases where the uncertainties of the communication networks are very large.

VITA

Lang Hong was born in Xiamen, Fujian, P.R. China on December 3, 1956. He graduated from the Fuzhou Second High School, Fujian, P.R. China in June 1974. He entered the Fuzhou University, P.R. China, in February, 1978 and received a Bachelor of Science degree in Electrical Engineering in January 1982, with the highest honors. From February 1982 to August 1984, he worked as a Teaching Assistant in the Fuzhou University.

In September 1984, Lang Hong was awarded a Graduate Assistantship and entered the Department of Electrical & Computer Engineering, the University of Tennessee, Knoxville. He received his Master of Science degree majored in Electrical & Computer Engineering in July 1984 and then started his Ph.D. program. From September 1986 to March 1988, he was employed as a Lecturer by the Department of Electrical & Computer Engineering. He also worked as a Research Assistant throughout his Ph.D. program. In December 1989, Lang Hong received his Doctor of Philosophy degree majored in Electrical and Computer Engineering.

Lang Hong is a member of the Honor Society of Phi Kappa Phi, and a member of the Institute of Electrical and Electronics Engineers.