

**Image Merging in a Dynamic Visual Communication  
System With Multiple Cameras**

*Ying Cui*

M.Sc., Beijing Institute of Technology, Beijing, China, 1989

B.Sc., Tsinghua University, Beijing, China, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

November, 1997

© Ying Cui, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical and Computing Engineer

The University of British Columbia  
Vancouver, Canada

Date Nov. 30, 97

## Abstract

In tele-operation, visual communication plays an important role as a source of information for control of a remote machine. The main objective of this thesis is to investigate the image merging in a dynamic visual communication system (DVCS) that can provide better visual presentation of the remote machine's working environment to the operator. The conventional VCS such as television cannot provide wide field of view (WFOV) and high resolution at the same time without significantly increasing the number of pixels and the bandwidth which is difficult and expensive. One of the proposed alternatives is to have a high resolution insert at the area of interest (AOI), determined by the observer's current eye orientation, projected into a cutout in the low resolution wide field of view (WFOV) background. This system is called a dynamic VCS (DVCS) in this thesis because of its active feedback control over the viewing scene. A DVCS requires a multi-channel imaging system, dual-resolution presentation, an eye tracker controlling the location of the AOI insert within pixel level accuracy, and an image merging system that can register and fuse AOI and WFOV images, all in real time.

This thesis discuss some of these issues, mainly focusing on the design and implementation of the image merging in such a system. Several possible approaches are analyzed with regard to the free parameters in the implementation, and experiments are carried out on seven sets of AOI and WFOV images. These images are taken by off-the-shelf cameras with different rotational angles, zooms (scale), and optical centres (translational change) (RST). The optical axis for AOI and WFOV imaging are kept parallel. Based on the analysis and experiments, a new multi-process approach was designed and implemented which can trade off performance characteristics for various imaging conditions. This approach requires only rough estimation of the RST values to start with and presents a registered and fused dual resolution image to the viewer. This processing is also calibration free and can relax the specification requirements of the position sensor and camera control devices. A new study of using corner attributes to recover RST values leads to a derivation of an analytical representation of the significance value for detecting scale-consistent corners. There are many other issues to be studied in the future for a better DVC system.

# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Figures</b>	<b>vi</b>
<b>Glossary</b>	<b>x</b>
<b>Acknowledgements</b>	<b>xiii</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	3
1.2 DVC System Structure . . . . .	3
1.3 Software Design . . . . .	5
1.4 Thesis Outline . . . . .	5
1.5 Thesis Contributions . . . . .	7
<b>Chapter 2 DVCS Specifications and Implementation Technology</b>	<b>8</b>
2.1 Human Visual Perception . . . . .	8
2.1.1 Visual Acuity . . . . .	9
2.1.2 Critical Fusion Frequency . . . . .	10
2.1.3 The Field Of View (FOV) . . . . .	11
2.1.4 Depth Perception and Stereopsis . . . . .	12
2.1.5 Eye Movements and Head Movements . . . . .	13
2.1.6 Expectations . . . . .	14
2.2 Implementation Technology and Design . . . . .	15
2.2.1 Current Technology Review . . . . .	15
1        Wide Field Approach . . . . .	15
2        Eye-tracking and Head-tracking . . . . .	16

2.2.2	CAE FOHMD . . . . .	16
2.2.3	CONDOR Advanced Visionics System . . . . .	20
2.2.4	Hardware System Design . . . . .	23
1	WFOV Imaging and Acquisition Using Intensity Array . . . . .	23
2	Image Fusion Using a Microlens . . . . .	23
3	Pixel Effect Removal For LCD Screen . . . . .	23
<b>Chapter 3 Processing System Design</b>		<b>25</b>
3.1	Imaging Condition and the Transform Matrix . . . . .	28
3.2	Image Registration . . . . .	29
3.3	Performance Evaluation . . . . .	29
3.4	Sampling and Interpolation . . . . .	30
3.5	Fusion . . . . .	31
<b>Chapter 4 Analysis of Correlation-based Image Registration Methods</b>		<b>32</b>
4.1	Estimation Accuracy . . . . .	32
4.2	Reliability of the Estimation Results . . . . .	36
4.3	Computational Cost . . . . .	37
4.4	Brute Force Search and Hierarchical Approach . . . . .	38
4.5	Other Options . . . . .	41
4.6	Experiments . . . . .	42
<b>Chapter 5 Using Corners as Control Points to Recover RST Changes</b>		<b>52</b>
5.1	Introduction . . . . .	52
5.2	Estimation Accuracy . . . . .	53
5.3	Reliability of the Estimation Results . . . . .	58
5.4	Computational Cost . . . . .	60
5.5	Other Factors . . . . .	60

5.6	Experiments . . . . .	61
<b>Chapter 6</b>	<b>RST Invariant Corner Detection</b>	<b>67</b>
6.1	Introduction . . . . .	67
6.2	Modeling a Corner . . . . .	69
6.3	Smoothing Effects and Definition of Significance Value . . . . .	70
6.3.1	Review of Multi-scale Corner Detection and Corner Behavior . . . . .	70
6.3.2	The Smoothing Effects on an Isolated Corners . . . . .	72
6.3.3	The Detectability and The Displacement of an Isolated Corner Over Smoothing Scale . . . . .	77
6.3.4	The Detectability and The Displacement of a DS-Isolated Corner Over Smoothing Scale . . . . .	78
6.3.5	Experiments . . . . .	80
6.4	Summary . . . . .	83
<b>Chapter 7</b>	<b>Image Fusion and Proposed Approach</b>	<b>85</b>
7.1	Radial Distortion Recovery . . . . .	85
7.2	Up-sampling & Down-sampling . . . . .	86
7.3	Normalization and Smoothing in Transition Band . . . . .	88
7.4	Proposed Approach . . . . .	90
<b>Chapter 8</b>	<b>Conclusions and Future Work</b>	<b>92</b>
8.1	Conclusions . . . . .	92
8.2	Future Work . . . . .	93
<b>Appendix A</b>	<b>Imaging System Using Intensity Array</b>	<b>94</b>
<b>Appendix B</b>	<b>Moving Insert Using a Microlens Array</b>	<b>95</b>
<b>Appendix C</b>	<b>Pixel Effect Removal for LCD Screen</b>	<b>99</b>
<b>Appendix D</b>	<b>Additional Experimental Results</b>	<b>101</b>
<b>Bibliography</b>		<b>120</b>

## List of Figures

Figure 1.1	Real Scene and DVCS Presented Scene . . . . .	2
Figure 1.2	The DVCS System Components . . . . .	3
Figure 1.3	The Ratio of the Pixel Numbers Required for Different Structure . . . . .	4
Figure 1.4	DVCS Software Diagram . . . . .	5
Figure 2.1	a. (left) Horizontal Section of Right Eyeball; b. (right) Schematic Diagram of Optic Media [Kelly, 56] . . . . .	8
Figure 2.2	a. (left) Visual Acuity vs Distance From Fovea [Mountcastle, 74] b. Simplified Regions . . . . .	10
Figure 2.3	a. (left) Effects of Retina Illumination on Critical Fusion [Mountcastle, 74]; b. (right) Simplified for AOI and WFOV Region . . . . .	11
Figure 2.4	a. (left) A Perimetry Chart Showing the Field of Vision From Left Eye [Kelly, 74]. b. (right) FOV in Practical System . . . . .	11
Figure 2.5	The Stereo Vision and Field of View Distribution [Mountcastle, 74] . . . . .	13
Figure 2.6	Typical Horizontal Eye Movements Recorded with a Photoelectric Monitor Showing (a) Saccadic Jumps, (b) Fixation Movement, (c) Smooth Pursuit [Young and Sheena] . . . . .	14
Figure 2.7	Comparison of Linear Projection and Fisheye View . . . . .	15
Figure 2.8	The CAE Fiber-Optic Helmet Mounted Display System Performance Specifications, [CAE product brochure] . . . . .	18
Figure 2.9	The Helmet Components, [CAE product brochure] . . . . .	19
Figure 2.10	AVS System Specifications Expected . . . . .	21
Figure 2.11	AVS Block Diagram . . . . .	21
Figure 2.12	CONDOR HMD Components, [Kanahele, 94] . . . . .	22
Figure 3.1	The Illustration of Image Fusion without RST Registration . . . . .	27
Figure 4.1	1D Illustration of Sample Points Distribution at Different Scales . . . . .	33
Figure 4.2	2D Illustration of the Range of Sample Point Difference (Outside Dotted Line: scale=3 and 4; Outside Dashed Line: scale=3 and 2.67) . . . . .	34

Figure 4.3	The $P_D$ vs Window Size & Scale Change with $S=4$ . . . . .	35
Figure 4.4	The $P_D$ vs Window Size & Scale Change with $S=3$ . . . . .	35
Figure 4.5	Computational Cost vs Factors ( $k=1000$ ) . . . . .	38
Figure 4.6	Building Multi-resolution Hierarchy . . . . .	39
Figure 4.7	Sample Points by Different RS Sampling . . . . .	39
Figure 4.8	Mapping Down-sampled AOI to WFOV . . . . .	40
Figure 4.9	The Flowchart of the Hierarchical Brute Force Search Approach . . . . .	41
Figure 4.10	The WFOV Images of the Seven Sets . . . . .	44
Figure 4.11	The AOI Images of the Seven Sets . . . . .	45
Figure 4.12	Correlation Surface at Coarsest Level with the RS Step Size Chosen for $P_s > 90\%$ . . . . .	46
Figure 4.13	Correlation Surface at Second Level with the RS Step Size Chosen for . . .	47
Figure 4.14	Comparison of the RS Estimation Error After Different Levels of Brute Force Search (BFS) . . . . .	48
Figure 4.15	Comparison of the X/Y Estimation Error After Different Levels of Brute Force Search (BFS) . . . . .	49
Figure 4.16	Comparison of Estimation Errors for CS-Lab (top) and Map (bottom) Pictures. . . . .	50
Figure 4.17	The Fused Images After Different Levels of BFS . . . . .	51
Figure 5.1	The Relation Between Performance Index and other Factors . . . . .	53
Figure 5.2	The Sensitivity to Scale vs the Corner Distance . . . . .	54
Figure 5.3	The Sensitivity to Rotation vs. Corner Distance . . . . .	55
Figure 5.4	The Location of Corners Bounded by the Windows . . . . .	56
Figure 5.5	Estimation of RST ( $T \rightarrow R \rightarrow S$ ) Using Two Control Points $A \rightarrow A' \rightarrow A'' \rightarrow (A)$ and $B \rightarrow B' \rightarrow B'' \rightarrow (B)$ . . . . .	56
Figure 5.6	The Search Range of Corner Matching vs RST Change . . . . .	59
Figure 5.7	Computational Cost vs Factors . . . . .	60
Figure 5.8	Comparison of the Frequency Domain Area Enhanced by Laplacian & Corner Map . . . . .	61

Figure 5.9	The Effects of Corner Numbers on the Estimation Error . . . . .	63
Figure 5.10	Comparison of the Estimation Error of Different Methods . . . . .	64
Figure 5.11	The Fused Images After Each Process . . . . .	66
Figure 6.1	Comparison of The Corner Sets for Different Resolution Images . . . . .	68
Figure 6.2	Corner Model . . . . .	69
Figure 6.3	Gaussian and Cylindrical tube as Smoothing Function . . . . .	73
Figure 6.4	The Approximation of The Gaussian Smoothing Function . . . . .	74
Figure 6.5	The Matlab Simulation Results for Smoothing Effects (a) $\sigma=4$ ; (b) $\sigma=2$ . . . . .	75
Figure 6.6	The Top View of the Smoothing Effects on the Corner . . . . .	75
Figure 6.7	The Step Edge Before (Solid Line) /After (Dotted Line) Smoothing . . . . .	76
Figure 6.8	The Gradient of Step Edge Before (solid line) /After (dotted line) Smoothing . . . . .	77
Figure 6.9	Test Image . . . . .	80
Figure 6.10	Solid: partial smoothing+selection; Dashdot: full smoothing; Dashed: Selection; Dot: no selection. . . . .	83
Figure 7.1	Comparison of Up-sampling and Down-sampling Results . . . . .	87
Figure 7.2	Comparison Between Rectangular transition & Circular Transition . . . . .	88
Figure 7.3	Comparison of Different Smoothing Methods . . . . .	89
Figure 7.4	Flowchart of Multi-process Approach for RST Estimation . . . . .	91
Figure A.1.1	The Intensity Element and the Relations Between its Parameters . . . . .	94
Figure A.2	Simple Designs of the Intensity Array (a) left: plate form; (b) right: top view of a semisphere . . . . .	94
Figure B.1	The Setup of the Optical System with Moving Insert . . . . .	95
Figure B.2	The Microlens Projection . . . . .	97
Figure B.3	Relation of The System Parameters . . . . .	98
Figure C.1	The Neighboring Pixel Intensity Before/After Vibration . . . . .	100
Figure D.1	RS Estimation Error Comparison Among Methods . . . . .	103
Figure D.2	X/Y Estimation Error Comparison Among Methods . . . . .	104
Figure D.3	Comparison of the Estimation Error of the Steps in Proposed Approach . . . . .	105

Figure D.4	Estimation Error of Methods on the Lab Picture Pairs . . . . .	106
Figure D.5	Comparison of Estimation Error of Methods on the Toys Picture Pairs . . .	106
Figure D.6	Comparison of Estimation Error of Methods on the Map Picture Pairs . . .	106
Figure D.7	Comparison of Estimation Error of Methods on the Tool Picture Pairs . . .	107
Figure D.8	Comparison of Estimation Error of Methods on the Plant Picture Pairs . . .	107
Figure D.9	Comparison of Estimation Error of Methods on My Picture Pairs . . . . .	108
Figure D.10	Comparison of Estimation Error of Methods on the Box Picture Pairs . . . .	108
Figure D.11	Comparison of Estimation Error Between 1-BFS & 2-BFS on All Pictures . . . . .	108
Figure D.12	Comparison of 1-BFS & 1-BFS+Corner Match on All Pictures . . . . .	109
Figure D.13	Comparison of Estimation Error Between 2-BFS with nonmaximum & 2-BFS with QLF on All Pictures . . . . .	109
Figure D.14	Comparison Between 4-BFS & 4-BFS with Corner Map on All Pictures . . .	110
Figure D.15	Comparison Between 4-BFS & 4-BFS+Optimization on All Pictures . . . .	110

## Glossary

As defined in [Burt, 96], or used in this thesis (\*):

**Blur** — a loss of image sharpness, introduced by defocus, lowpass filtering, camera motion, etc.

**Border** — the first and last row and column of a digital image.

**Corner** — a two dimensional discontinuity point. \*

**Digital Image** — (1) an array of integers representing an image of a scene; (2) a sampled and quantized function of two or more dimensions, generated from and representing a continuous function of the same dimensionality; (3) an array generated by sampling a continuous function on a rectangular (or other) grid and quantizing its value at the sample points.

**Digital Image Processing** — digital processing of images, the manipulation of pictorial information by computer.

**Edge** — (1) a region of an image in which the gray level changes significantly over a short distance; (2) a set of pixels belonging to an arc and having the property that pixels on opposite sides of the arc have significantly different gray levels.

**Feature** — a characteristic of an object, something that can be measured and that assists in classification of the object (e.g., size, texture, shape).

**Feature Extraction** — a step in the image processing in which measurements of the features are computed.

**Foveal Vision:** the viewing angle that has the highest resolution. \*

**Hierarchical Approach:** Multi-resolution approach using the low resolution result to limit the search space of the higher resolution search.

**Gray Level** — (1) the value, associated with a pixel in a digital image, representing the brightness of the original scene at the point represented by that pixel; (2) a quantized measurement of the local property of the image at a pixel location.

**Low-pass Filtering** — an image smoothing (usually convolution) operation in which the low frequency components are emphasized relative to the high frequency components.

**Image** — any representation of a physical scene or of another image.

**Image Matching** — any process involving quantitative comparison of two images in order to determine their degree of similarity.

**Image Processing Operation** — a series of steps that transforms an input image into an output image.

**Image Registration** — a geometric operation intended to position one image of a scene with respect to another image of the same scene so that the objects in the two images coincide.

**Interpolation** — the process of determining the value of a sampled function between its sample points.

**Neighborhood** — a set of pixels located near a given pixel.

**Non-maximum Suppression** — a way to obtain the maxima by selecting the pixels that have higher value than all the eight immediate neighborhood pixels. \*

**Peripheral Vision** — the viewing angle with low resolution and wide field of view. \*

**Pixel** — the smallest element of a digital image. The basic unit of which a digital image is composed.

**Region** — a connected subset of an image.

**Registered Images** — two or more images of the same scene that have been positioned with respect to one another so that the objects in the same scene occupy the same positions.

**Resolution** — (1) in optics, the minimum separation distance between distinguishable point objects; (2) in image processing, the degree to which closely spaced point objects in an image can be distinguished from one another.

**Sampling** — the process of dividing an image into pixels (according to a sampling grid) and measuring the local property (e.g., brightness or color) at each pixel.

**Scale** — the ratio of the zooms in camera used to take WFOV and AOI.

**Scene** — a particular arrangement of physical objects.

**Sharp** — pertaining to the detail in an image, well defined and readily desirable.

**Smoothing** — any image-processing technique intended to reduce the amplitude of small details in an image. Smoothing is often used for noise reduction.

**System** — anything that accepts an input and produces an output in response.

**Texture** — in image processing, an attribute representing the amplitude and spatial arrangement of the local variation of gray level in an image.

Asynchronous used in this thesis:

**AOI:** Area Of Interest

**AVS:** Advanced Visionics System

**BFS:** Brute Force Search

**CAE:** CAE Electronics LTD

**CFF:** Critical Fusion Frequency

**CONDOR:** The Covert Night/Day Operations for Rotorcraft Program

**DVC:** Dynamic Visual Communication

**DVCS:** DVC System

**FOHMD:** Fiber Optics Helmet Mounted Display

**FOV:** Field Of View

**HDD:** Head Down Display

**HMD:** Helmet Mounted Display

**IFOV:** Instantaneous Field Of View

**VC:** Visual Communication

**VCS:** VC System

**WFOV:** Wide Field Of View

## Acknowledgements

I would like to express my gratitude to my advisor, Dr. Peter D. Lawrence, for his guidance, support and advice throughout the course of this thesis. His advice has always been appropriate and convincing, and inspired me to pursue my profession. I also had the good fortune to have the advice of Dr. Lynn Kirlin, professor in University of Victoria, who has been encouraging and understanding. Mr. Haig Farris, Faculty of Commerce and venture capitalist, has broadened my view and knowledge through his course and his involvement in Green College, and has influenced my career orientation.

Thanks to my committee members, Dr. Jim Little, Rabab Ward, Matt Palmer, who supported my thesis work by offering their knowledge and opinions through many stages. Dr. David Lowe offered his expertise and time through his course and discussion on this thesis work.

My friends and colleagues have helped me through by sharing their ideas and caring for my progress. A research engineer in the Robotics Lab, Niall Parker, the first person I met in my office, introduced me to the new environment; Dan McReynolds, a Ph.D candidate in LCI lab of the Computer Science department, shared his ideas and experiments, especially provided his valuable opinions on my work. My good friends, Ying Fang, Graham Bell, Shyan Ku, Shahram Tafazoli, Terence Gilhuly, Hossein Saboksayr, Ray Burge, Atousa Soroushi, Taming Yang, Yi Guo, Wayne Lee, Dave Michelson, Diane Sun, fellow students in Green college, all shared many charitable times with me. Special thanks to my volleyball team with whom I won my biggest title in school, the Intramural Championship.

Thanks to my parents, who dedicated their career to education, made me enjoy campus environment and choose to pursue schools as long as it takes. My father, who passed away years ago, had the vision that I can be an engineer, and I am glad that he made that the decision.

My pride is my son, Daniel Hsiao, who has been busy growing up, motivated me to make my own progress.

## Chapter 1 Introduction

The real-time communication of visual information, known as television, has spread from the entertainment industry into many commercial and industrial applications, see Fig. 1.1(a) and (b). Improvements have been made to incorporate wider view and higher resolution into cameras, transmission, and display systems. There are also many patented systems for binocular stereo, although none are widely used.

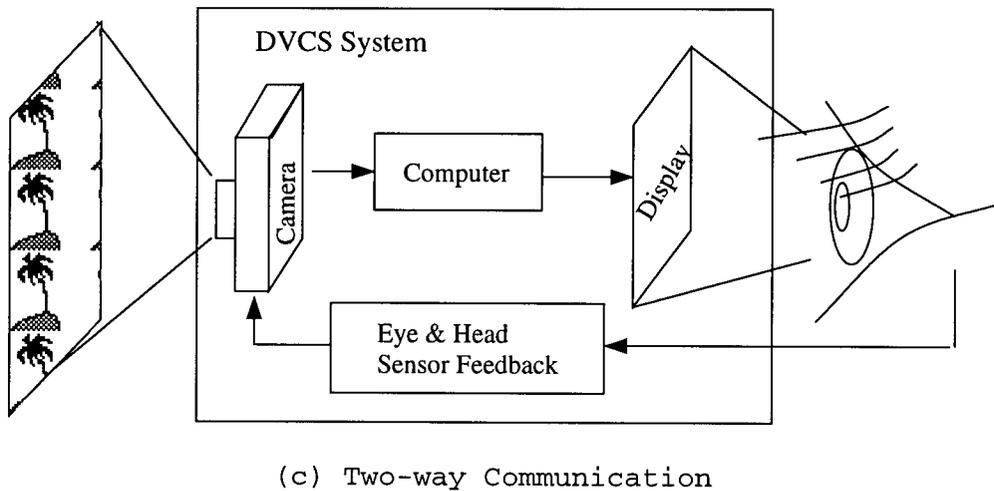
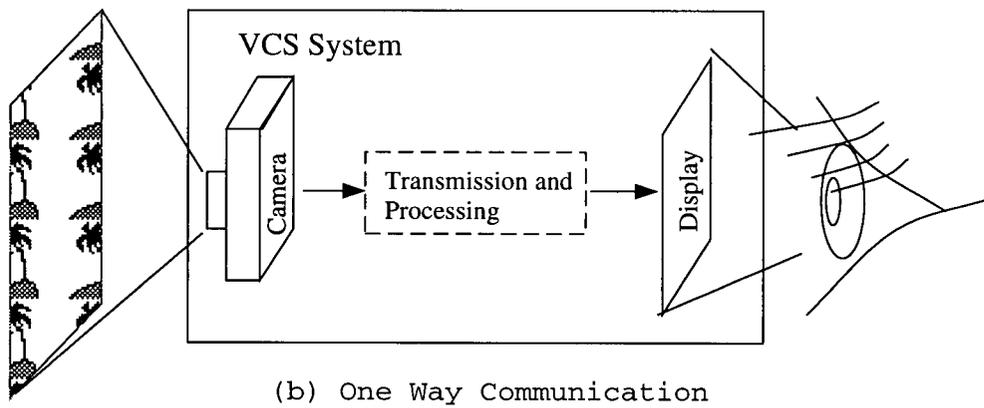
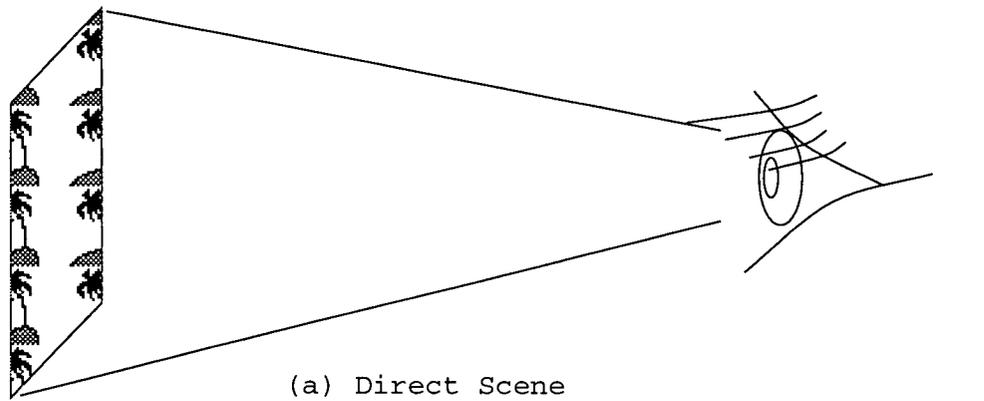
One application of television is as a part of a system for manipulation of a remote environment. Currently such "teleoperation" systems are in use in remote handling of nuclear and other hazardous materials, as well as other applications. In some proposed applications, visual acuity is especially important such as for remote display. Since image sensors and displays have fixed rasters (e.g., 640x480, 800x600, etc.), improving the resolution is achieved only by reducing the angle of view of the whole scene.

The human visual system has a structure that allows it to acquire wide angle low resolution information in the natural periphery and narrow angle high resolution information at the fovea in a "seamless" overall view. Details of the scene are acquired by moving the fovea (by eye rotation) to an area of interest.

A high resolution television system must either:

1. Create a wide angle display of high resolution using multiple screens or projectors (the approach taken in the "cave" simulation at the University of Illinois, [www.beckman.uiuc.edu/themes](http://www.beckman.uiuc.edu/themes)) (see Fig. 1.1.(b)).

2. Using a head-mounted display to create a high (a foveal) resolution view of the world only where the viewer is currently looking and elsewhere create a low resolution view. This approach attempts to match the human visual system capabilities. To accomplish this, one requires a high resolution camera to follow the viewer's current line-of-sight (Fig. 1.1(c)) and a means of merging the high resolution and low resolution views in the display.



**Figure 1.1** Real Scene and DVCS Presented Scene

The objective of this thesis is to investigate the issues involved in fusing a high resolution insert into a low resolution background, to obtain high resolution in the area of interest (AOI), while keeping the wide field of view (WFOV) background. The AOI is designed to follow the

eye-movement (in a Helmet Mounted Display, HMD) or head-movement (in a Head Down Display, HDD). This is the model to be investigated in this thesis and is defined as a dynamic VCS (DVCS) because of the active feedback control over the scene.

## 1.1 Background

This DVCS model originated from flight simulator display systems and has been adapted for tele-operation application. A wide field of view is necessary for the awareness of the changes over the background while high resolution is required in the area of operation. One major structural difference for tele-operation VCS is that the AOI and WFOV images are taken from remote cameras instead of computer generated as in a flight simulator. There are many detailed issues to be investigated in this adaption such as wide field of view imaging, software systems for real-time, calibration-free registration and fusion of the AOI and WFOV images.

## 1.2 DVC System Structure

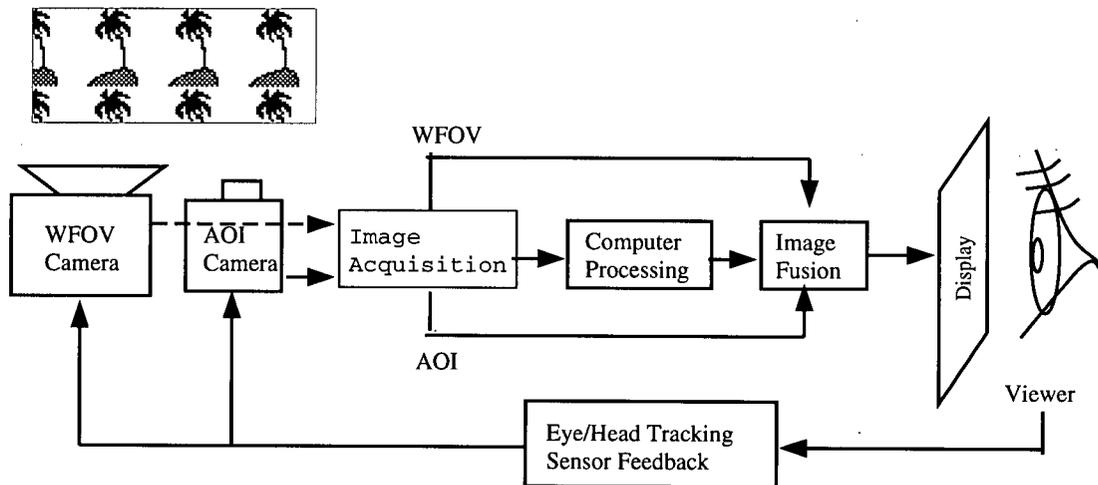
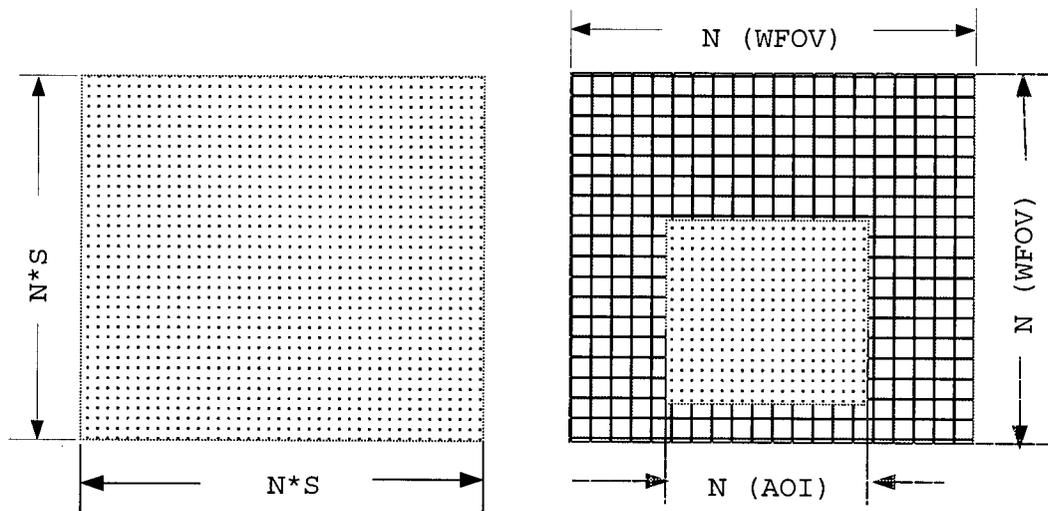


Figure 1.2 The DVCS System Components

A DVC system consists of four major components: imaging, transmission and processing, display and feedback control. The imaging system is normally one or multiple cameras, with optical lenses and 2-dimensional optical sensors (e.g., CCD array). The transmission system can be a cable or an RF link and the processing system is usually a computing system (or processor). The display system

can be a CRT or LCD screen, fibre optic bundle, or projector. In our study, we will concentrate on a DVCS model with two video cameras as the imaging system, one with wide field of view (WFOV) but low resolution, another with a narrow field of view but high resolution tracking the area of interest (AOI), a cable as the transmission line, a computer as the processing system, and a helmet-mounted-display as the display system. The feedback loop obtains position information from position sensors tracking the movement of the eye and head and goes to a camera control device. This is shown in Fig. 1.2.

The slaved AOI insert is able to shift from one location to another. This is necessary in a practical VCS system since the resources (including sensors, actuators, and processing) are usually limited, and must be used selectively in order to enhance the focused information without losing the background. Since a display device like a LCD screen has only one fixed resolution, the images from separate LCD screens must be combined optically.



**Figure 1.3** The Ratio of the Pixel Numbers Required for Different Structure

To illustrate in detail, this dual resolution presentation has the advantage of being able to present high resolution image with a wide field of view without having high resolution over the peripheral vision, which reduces the total number of pixels required for the same result by many times. This is shown in Fig. 1.3. If one wants to increase the low resolution scene by a resolution factor of  $S$ , then the whole scene has  $NS \times NS$  pixels in the first approach as shown on the left,

and  $N^2 + N^2$  pixels (assuming the low resolution pixels under the insert are generated and then discarded), as shown on the right. The ratio is then  $(NS)^2/(2N^2) = S^2/2$ .

### 1.3 Software Design

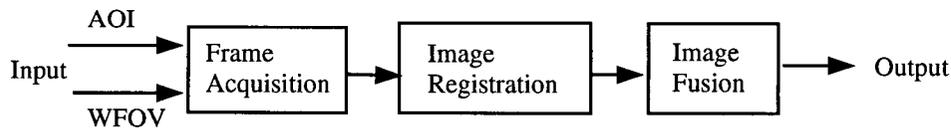


Figure 1.4 DVCS Software Diagram

In the DVC system defined earlier, the AOI and WFOV are taken with different zoom, angle and optical centres, and need to be registered and fused before being displayed. That is, the software system needs to acquire the AOI and WFOV images from the cameras and recover the rotation/scale/translation (RST) changes between the two images, followed by a fusion process. This is shown in Fig. 1.4.

The goal is to perform calibration-free registration and fusion of images with RST changes in real-time. The feasibility and performance of such a software system are studied in detail in this thesis. Experiments are carried out on seven sets of images with various content. As a result, an optimized approach is proposed and implemented. This multi-process approach is able to adapt to the different imaging conditions and minimize the computational cost accordingly.

### 1.4 Thesis Outline

The thesis work includes the investigation of the DVCS system specifications from the end user's point of view, examination of the technology involved in implementation, and a feasibility study of the software part of the DVCS model, so that the requirements for the hardware can be understood better.

Chapter 2 will briefly review human visual perception, with an emphasis on human visual capacities in the areas that are related to the VCS system specifications. This information will help to define the DVCS system specifications. This chapter will also investigate the current technology under development or developed to achieve these goals and the limits of the current technology. The

two most advanced systems, CAE FOHMD (commercial) and CONDOR (military), are introduced as examples of the current stage of development, followed by some design ideas.

Chapter 3 will discuss the software design for this system. This includes the requirements, review of the related literature and the basic steps and functions involved. Among the steps in the software implementation, the RST estimation is the most complex step and is discussed in more detail. A review of the RST estimation algorithms leads to three possible approaches which will be examined in the next three chapters. The evaluation methods are introduced.

Chapter 4 will start to investigate the details of a correlation-based approach. The performance is analyzed with respect to the free parameters, and a hierarchical brute force search is proposed and tested. Experimental results are given as well as the computational cost.

Chapter 5 will study a second approach which is feature-based. In this approach, corner points are used as control points for the image registration. Among all the features, corners are selected because they are 2D features and are invariant to RST changes, which is especially relevant in our case. Corners are used as control points and correlation is used in corner point matching in our analysis. The reliability, accuracy, and computational cost are analyzed and experiments are carried out. The results suggest that only small RST changes can be recovered reliably and efficiently with this approach.

Chapter 6 will study scale-consistent corner detection, as the first step in using corner attributes for image registration of large RST changes. As analyzed in Chapter 5, corner matching using correlation without any knowledge of RST values requires a large range of search space which is not computationally efficient nor reliable. So, alternative ways of corner matching, such as geometric, rigidity, disparity, attributes and invariants have been studied over years. This chapter will study the corner behavior over scale so that scale consistent corners can be identified and used as control points. A significance value associated with scale-consistency is defined based on the study of the smoothing effects on the corners.

Chapter 7 will study the fusion of the registered images. In hardware, they can be projected onto one screen. In this chapter, software fusion is discussed and implemented. This involves recovering possible relative distortion, resampling, normalization and transition band smoothing.

Also, a complete approach of software implementation is summarized in Chapter 7, based on the analysis and experiments for the previous chapters. This approach combines different process to obtain best results for various imaging conditions.

## 1.5 Thesis Contributions

This thesis has made contributions in the following subjects:

- \* Study of a feasible DVCS model, the system specifications, implementation technology, and processing system;
- \* Analysis of the performance of the image registration algorithms in recovering rotation/scale/translational (RST) changes between insert AOI and background WFOV images;
- \* Experiments on seven sets of images taken with various image contents and RST changes to study the effects of the free parameters in implementing the registration algorithms;
- \* Implementation of a processing system that can trade off performance characteristics based on the imaging condition using an adaptive approach;
- \* Study of the scale-space behavior of corners and derivation of the analytical presentation of the significance value for detecting the scale-consistent corners;
- \* Three novel ideas for improving the DVCS implementation are proposed: a) an imaging system using an intensity array; b) an eye-tracking insert using a microlens; c) pixel effect removal for LCD screen display.

## Chapter 2 DVCS Specifications and Implementation Technology

The ultimate goal of a DVC system is to present as realistic a scene as is necessary to carry out a task. There are always some trade-offs in implementation, based on applications and affordability. To make the best trade-offs, there is a need to understand how the human eye perceives the scene, as well as how well current technology performs. This chapter will discuss the facts from both perspectives.

First, a brief review of human visual perception is given, with an emphasis on human visual capacities in the areas that are related to the VCS system specifications. As these set the expectations of the system specifications, the following section will investigate the feasibility of and technology for implementation. Technical difficulties and limits are discussed, followed by some design ideas. The two most advanced systems, CAE FOHMD (commercial) and CONDOR (military), are introduced as examples of the current stage of development.

### 2.1 Human Visual Perception

The optical structure of the eye is shown in Fig. 2.2(a). The simplified “reduced eye” is shown in Fig. 2.2(b), from [Mountcastle, 74].

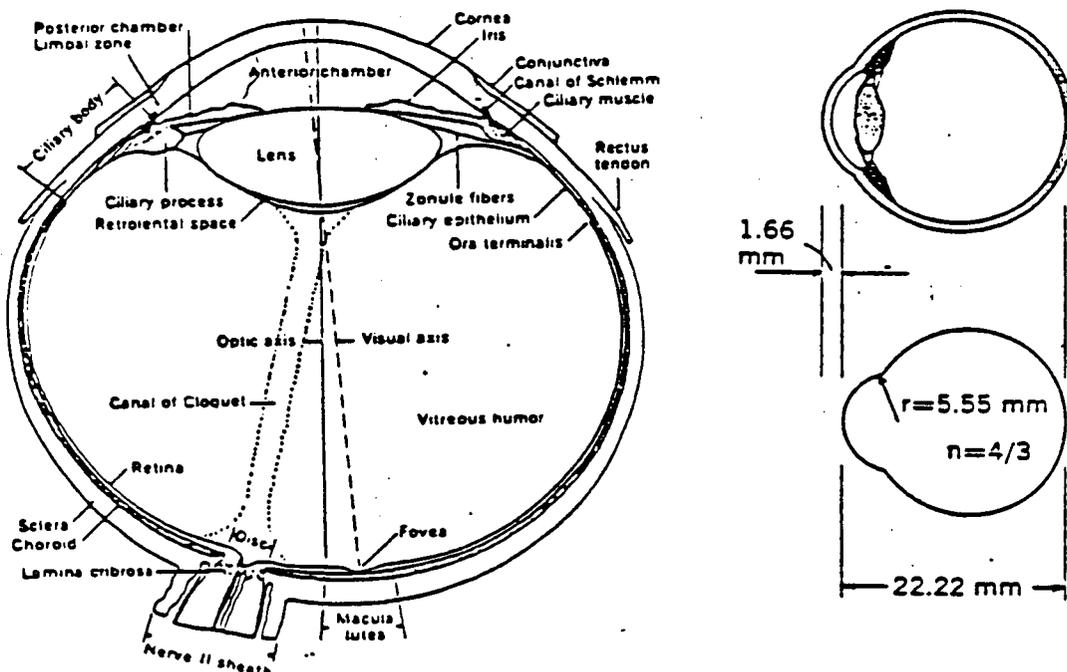


Figure 2.1 a. (left) Horizontal Section of Right Eyeball; b. (right) Schematic Diagram of Optic Media [Kelly, 56]

As shown in Fig. 2.2, the optics of the eye (in good focus) is a space-invariant linear optical system which can be represented by a point-spread function. Unfortunately, the other parts like receptors (sensor) and visual pathways (processor) are neither space-invariant nor linear. So, it takes more than just a point-spread function to model the whole system, and a complete model is not possible. What is needed is to identify a set of significant visual phenomena that can be duplicated to some extent with current technology. State of the art DVCS systems can accommodate field of view (aperture), visual acuity (spatial resolution), critical fusion frequency (frame rate), color, stereo parallax (disparity), eye-movement and head-movement to some extent. So, in the following sections, we will discuss the visual capabilities in these functionality.

**Visual Acuity** The ability to detect a border, or an abrupt change in the visual field is called the resolution of a stimulus in space. This is illustrated in Fig. 2.3(a). In practice, they are simplified into three regions as shown in Fig. 2.3(b)

- The foveal region which is the hyperacuity area of 26 arc seconds or one arc minute approximately;
- The transition band where the resolution drops rapidly from 1 arc minute to about 10 to 30 arc minutes;
- The peripheral region where the resolution stays reasonably stable at about 30 arc minutes.

As discussed in [Mountcastle, 74] and [Kelly, 56], the central retina, particularly the fovea, contains a much greater information density and provides most fine details of the scene, while the peripheral region is mainly responsible for target detection, motion detection, pattern recognition in wider range, and so on. So it's equally important to present both the high resolution foveal vision and wide FOV peripheral vision. This is the major reason for adopting a two level resolution DVCS, with an eye-tracking high resolution insert covering the foveal region called the area of interest (AOI), while keeping the wide field of view (WFOV) background covering the peripheral region.

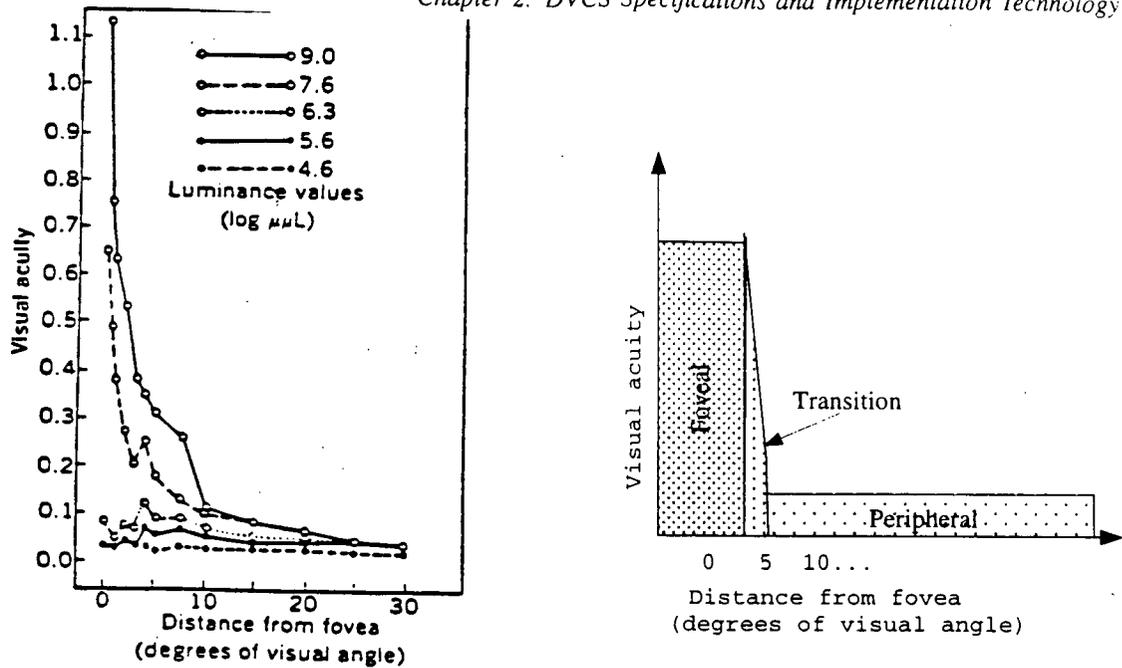


Figure 2.2 a. (left) Visual Acuity vs Distance From Fovea [Mountcastle, 74] b. Simplified Regions

**Critical Fusion Frequency** The ability to discriminate the stimuli that are separated in time is called the temporal resolution of stimuli. The critical fusion frequency (CFF) is the minimum frequency at which repeated stimuli appear to fuse together into a continuous stimulus. The resolution of stimuli in time is limited because the response to a given stimulus does not cease exactly when the stimulus ceases but persists for a time thereafter. Fig.2.4(a), as in [Mountcastle, 74], shows CFF as a function of both retinal illumination (intensity of the flickering stimulus) and retinal location.

From Fig.2.4(a), we can see the critical fusion rate distribution function can be simplified into a piecewise function, 30Hz for the peripheral region and 50 Hz for the foveal region. This is the reason behind the 30 frame/second TV screen and 50~60 frame/second refresh rate for high performance monitors. This also suggest that AOI frame rate should be 50~60, while WFOV can just have 25~30 frame/second in DVCS system as shown in Fig.2.4(b).

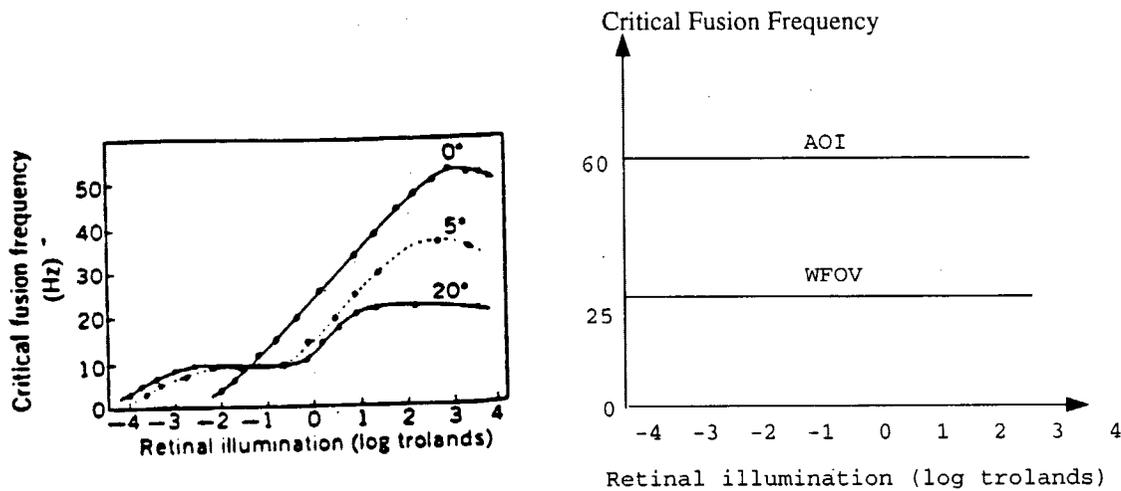


Figure 2.3 a. (left) Effects of Retina Illumination on Critical Fusion [Mountcastle, 74]; b. (right) Simplified for AOI and WFOV Region

**The Field Of View (FOV)** The field of view (FOV) is the area seen by an eye at a given instant, also called the instantaneous FOV (IFOV). The perimetry chart is a chart that shows the FOV, the area not covered in the dark in Fig. 2.5(a). In all perimetry charts, a blind spot caused by lack of rods and cones in the retina over the optic disc is found approximately 15 degrees lateral to the central point of vision, as illustrated in Fig.2.5(a), [Kelly, 56].

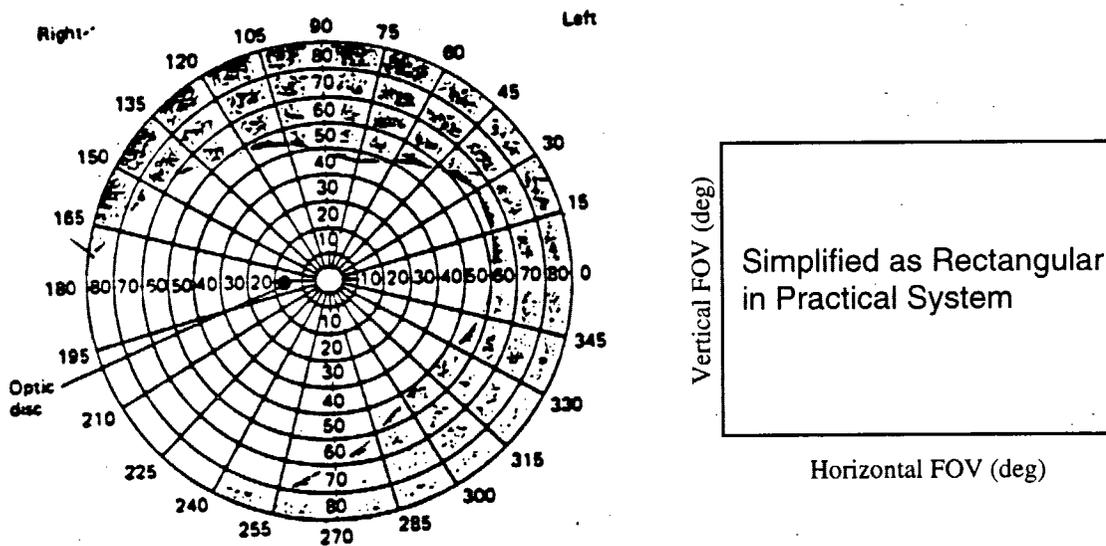


Figure 2.4 a. (left) A Perimetry Chart Showing the Field of Vision From Left Eye [Kelly, 56]. b. (right) FOV in Practical System

Visual phenomena are closely related to the FOV, in the sense they are functions of FOV. As discussed earlier, CFF and visual acuity are subjective to the region of the field, for example.

Research by [Mitsuo Ikeda and Tetsuji Takeichi, 75] show that the functional visual field size shrinks with foveal loads of greater recognition difficulty. They also noticed that this effect is less obvious with trained subjects, indicating that they learned to shift their attention towards the periphery without sacrificing detectability at the fovea. Since this is more subjective, we wouldn't consider this effect in our module.

To be able to provide a display that covers the IFOV, or the greater part of it, is very essential in creating a tele-presence. But most systems can only provide a WFOV (defined as minimum 90 degree), and a larger FOV presentation is still under investigation.

**Depth Perception and Stereopsis** Stereoscopic acuity, is a measure of the ability to discriminate differences in the distance of stimuli from the eyes. The visual apparatus normally perceives distance by three major means [Mountcastle, 74]. These means are (1) determination of distance by sizes of retinal images of known objects (2) determination of distance by moving parallax, and (3) determination of distance by stereopsis—binocular vision. Fig.2.6, as in [Mountcastle, 74], shows the stereo vision FOV. Stereo is an important aspect of vision, and usually implemented in a HMD by presenting two eyes with a pair of images taken with disparity. Details are not discussed in this thesis.

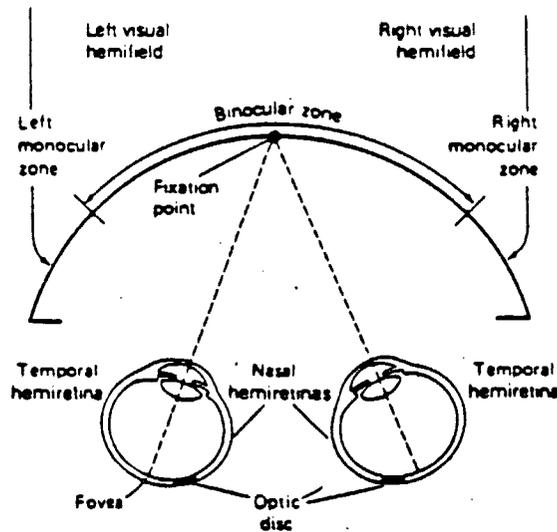


Figure 2.5 The Stereo Vision and Field of View Distribution [Mountcastle, 74]

**Eye Movements and Head Movements** Eye movements help the visual system by orienting the fovea to the center of the visual task [Young&Sheena75]. Besides that, eye-movement and head movement also enhance the dynamics of the visual field, accommodation, and hyperacuity recognition tasks. Fig. 2.7, from [Young&Sheena75], shows typical single-eyed horizontal eye movements recorded by the photoelectric technique. The figure gives all the basic types of eye-movements and their speed and range.

Eye movements that are in the range of 0.5 to 1 degree like fixation movements are not tracked in a VCS system. The reason is the AOI display is designed to be slightly larger than the foveal region, so even with the presence of the fixation movement or other small range eye-movements, the foveal vision will still be in the same AOI region. This makes the eye-tracking task easier to implement. Time delay becomes a significant problem in eye-tracking systems since the overall system time lag should be less than one frame time, and the sampling rate needs to be at least higher than the fusion rate in the foveal region which is 50Hz.

The range of eye movements are usually within  $\pm 30$  deg., though different types of eye movements have different ranges of movement. Head motion is often involved when the target

motion exceeds 30 degrees.

In a helmet-mounted DVC system, the AOI follows eye-tracking while the WFOV follows head-tracking. The hard edges (abrupt resolution changes ) between the AOI and the WFOV should be smoothed with a transition region. An AOI width of 25 degrees within which is a 5 degree wide smoothly varying transition region and IFOV of 130 degree are recommended by Eric. M. Howlett, see [Howlett, 92] & [ Kelly, 92].

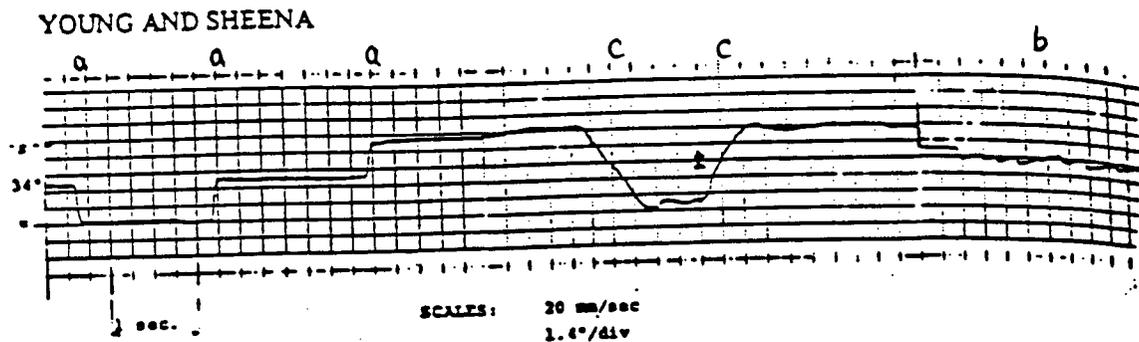


Figure 2.6 Typical Horizontal Eye Movements Recorded with a Photoelectric Monitor Showing (a) Saccadic Jumps, (b) Fixation Movement, (c) Smooth Pursuit [Young and Sheena]

**Expectations** The above review on human vision capabilities provides information on the expectations and evaluation measures of a DVCS system. That is, the system is perfect when the performance matches the direct human visual capabilities. Since a practical system can't afford to meet all these expectations, some kind of trade-offs have to be made. In our study, we adopt the same system structure used for flight simulator display, except the source of images are from remote cameras instead of being computer generated. Some adjustments are made to reduce the cost of the system for feasible industrial applications. These can be justified by lowering some of the performance criteria that are less critical in industrial applications, such as system time-delay, resolution and field of view.

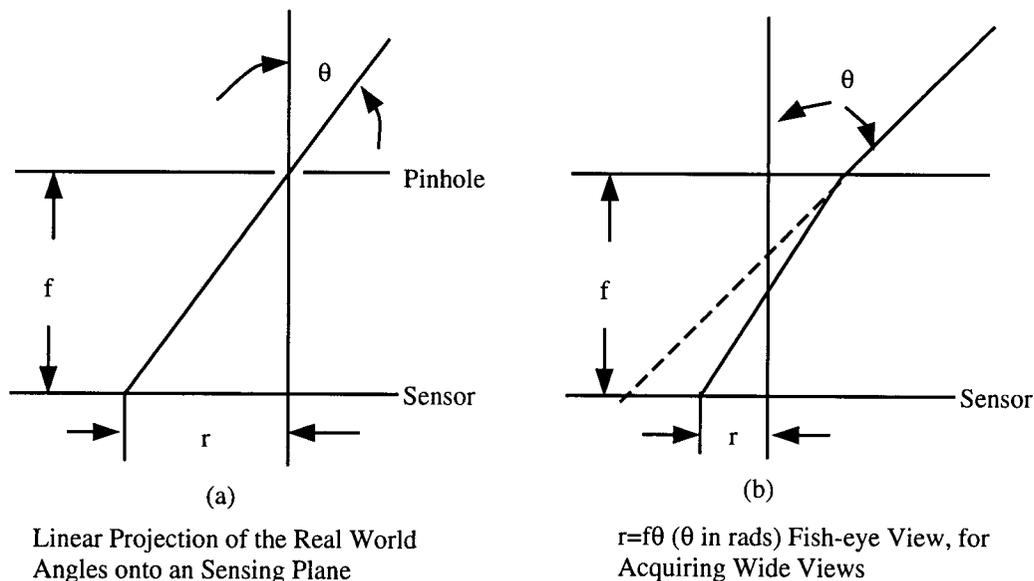
## 2.2 Implementation Technology and Design

The limits of some technical difficulties and affordability concerns determine that some kind of trade-offs have to be made for a practical system.

An extensive review of the literature in this area was carried out. No papers are found describing the same research in this area. However, there are several commercial products that use a AOI over a WFOV for only flight simulators and cockpit target display. This section will discuss the related implementation issues, starting with reviews of the current technology and state of the art designs, followed by some improvement ideas.

### Current Technology Review

**1 Wide Field Approach** One of the technical difficulties is in imaging/displaying a WFOV image. The traditional method is a mosaic of several LCD screens. Given a large enough number of LCD or CRT screens, a very wide field of view can be obtained at high cost.



**Figure 2.7** Comparison of Linear Projection and Fish-eye View

A fisheye lens can also provide WFOV. This is achieved by compressing the wide angle image in a way that looks like a fisheye. When the 3-D real world is projected to a 2-D plane sensor,

usually they are linear projections as shown in Fig. 2.8(a), as in a pinhole camera, or an ordinary camera with a distortion-free lens. The problem is a wide bandwidth is required for transmission and storage, and a high resolution and wide angle lens is hard to make.

Fig. 2.7(b) shows the fisheye view projection, which is used for acquiring very wide FOV. Recently, this kind of fisheye system has become more popular because its resolution pattern can be made to match the human visual system, i.e., higher priority is given to the foveal field to make the best possible use of the total pixels, and the peripheral field is compressed into the edges of the image. The image is distorted radically and has to be restored before being presented, as discussed in [Eric M. Howlett, 89]. The problem with this system is that the focus is always centered and fixed.

**2 Eye-tracking and Head-tracking** Eye-tracking is used to record the eye movement and control the positioning of the high resolution insert so that the insert always follows the foveal vision of the eye, or AOI. Many techniques for monitoring eye movements have been developed, and a good review can be found in [Geoffrey R. Loftus, 74].

Head-tracking is used to record the head movement so that the background or WFOV always follows the head movement. See [Howlett, 89] for details.

## CAE FOHMD

There are currently two types of dual-resolution image presentations. One uses an eye-tracked insert fused with a head-tracked background all in one helmet-mounted-display (HMD). A recent CAE product called a fibre optic helmet mounted display (FOHMD) is an example. Another way is to have a head-down WFOV (background is displayed on the surrounding screen) and head-mounted AOI, as in the CONDOR system, which will be introduced in next section.

The CAE FOHMD is the most advanced commercial product designed for a flight simulator display system. The performance specifications for the FOHMD are shown in Fig. 2.8, from a CAE product flyer. The detailed implementation is not known to us. This system is introduced briefly as a reference of HMD model and the performance achieved. The system helmet components is shown in Fig. 2.9, [CAE product brochure]. In this picture, the pilot's head motion are sensed

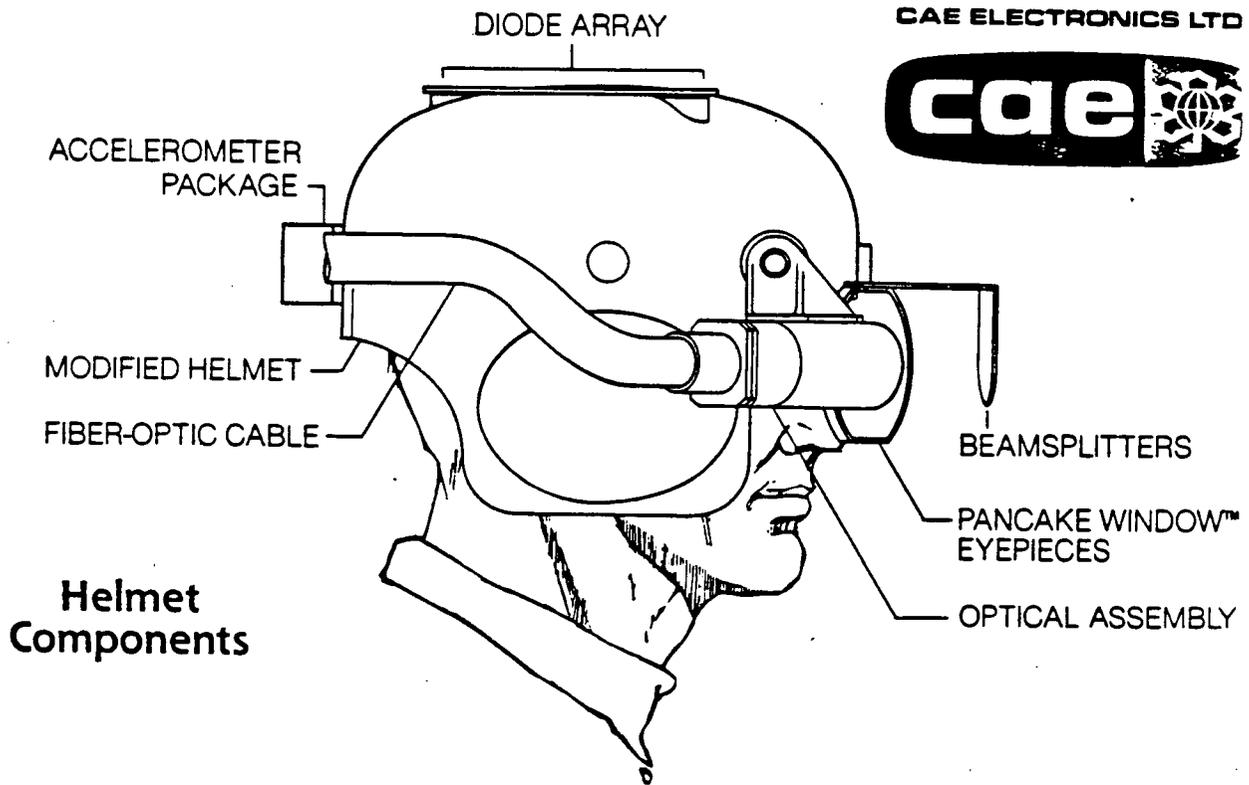
with a diode array and accelerometer. Fiber-optic cables carry the image which are displayed via an optical assembly to the eyes.

The performance specifications for the Fiber-Optic Helmet Mounted Display are as follows:

(Subject to change without notice)

- (a) **Instantaneous FOV:** 135° horizontally by 64° vertically, consisting of: a 135° by 64° low-resolution field, and 55° by 30° or 25° by 19° high-resolution field.
- (b) **Total FOV:** unlimited
- (c) **Brightness:** minimum 30 footlamberts
- (d) **Contrast:** greater than 30:1
- (e) **System Resolution:** Background: 135° x 64° – 4.6 arcmin/TV line (horizontal)  
Inset: 25° x 19° – 1.4 arcmin/TV line (horizontal)
- (f) **Geometric Distortion:** less than 1.5%
- (g) **Registration:** inherently registered at all times
- (h) **Color:** full color system
- (i) **Optical Helmet Tracker:**
  - (1) **Range:** Pitch  $\pm 90^\circ$   
Yaw  $\pm 180^\circ$   
Roll  $\pm 90^\circ$   
X 24 inches  
Y 24 inches  
Z 12 inches
  - (2) **Resolution:** Pitch 0.05°  
Yaw 0.05°  
Roll 0.05°  
X 0.01 inches  
Y 0.01 inches  
Z 0.01 inches

**Figure 2.8** The CAE Fiber-Optic Helmet Mounted Display System Performance Specifications, [CAE product brochure]



**Helmet  
Components**

Figure 2.9 The Helmet Components, [CAE product brochure]

## CONDOR Advanced Visionics System

Another way of presenting the insert and background is to have a head-down display (HDD) as background and a helmet-mounted insert projection slaved to the head motion. This is proposed for the most advanced military system in CONDOR program. The project is discussed in [Kanahele, 96].

As explained in [Kanahele, 96], "The Covert Night/Day Operations for Rotorcraft (CONDOR) program is to develop and demonstrate an advanced visionics concept coupled with an advanced flight control system to improve rotorcraft mission effectiveness during day, night, and adverse weather conditions in a Nap-of-the Earth (NOE) environment. The Advanced Visionics System (AVS) for CONDOR is the flight-ruggedized head mounted display and computer graphics generator with the intended use of exploring, developing, and evaluating proposed visionic concepts for rotorcraft including the application of color displays, wide field-of-view, enhanced imagery, virtual displays, mission symbology, stereo imagery, and other graphical interfaces."

"AVS was designed for use as a research tool to be integrated into both in-flight helicopter and ground-based simulator environments. The AVS HMD will enable researchers to test the impact and interaction of various display parameters such as color, FOV and overlap and as such, requires enhanced performance capabilities for image resolution, color and accuracy. To accomplish this, state of art technology has been incorporated into AVS design."

The system specifications are given in Fig. 2.12. Again, no details are known to us, it is introduced as a reference for an HMD model.

"The AVS design includes a Programmable Display Generator (PDG), Head Down Display (HDD), and Helmet Mounted Display (HMD) system. Both the PDG and the HDDs are commercial off-the-shelf systems modified to meet AVS requirements. The HMD system, however, uses new technology and subsystems developed specifically for the AVS program." The system block diagram is shown in Fig. 2.11, and the HMD components is shown in Fig. 2.12.

Luminance:	120 fL
Contrast:	12:1
Field-of-View	50 deg. V x 60 deg. H per eye
Overlap	20 deg., 30 deg., 40 deg. selectable
Resolution	1280 H x 1024 V pixels (0.5 lp/mrad)

Figure 2.10 AVS System Specifications Expected

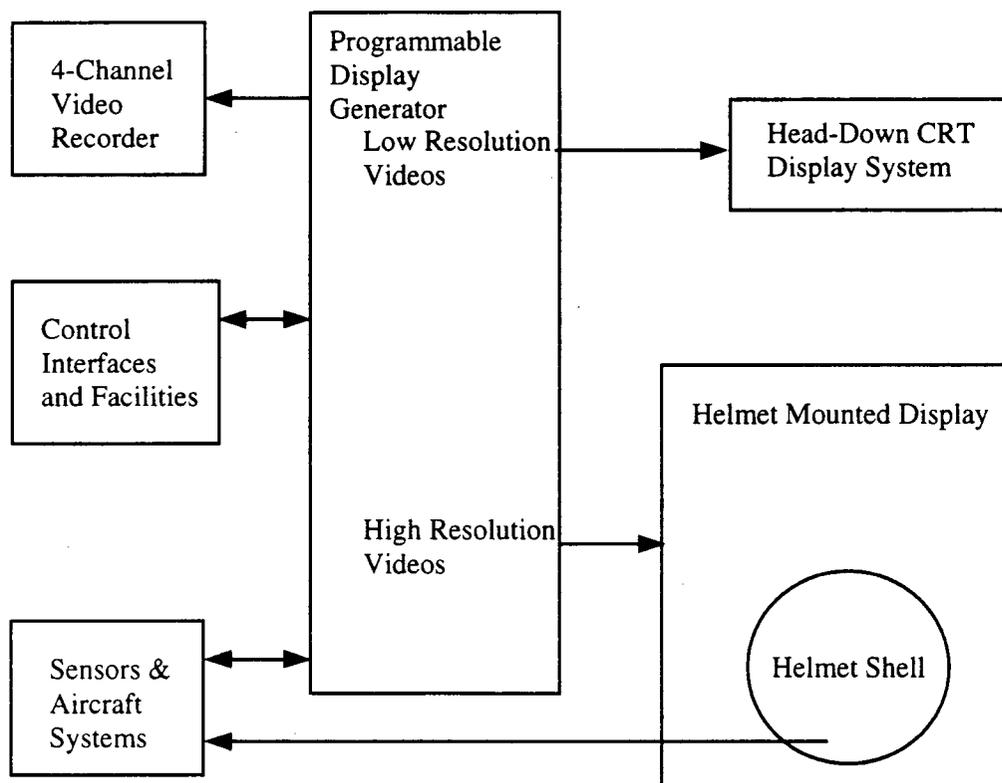


Figure 2.11 AVS Block Diagram

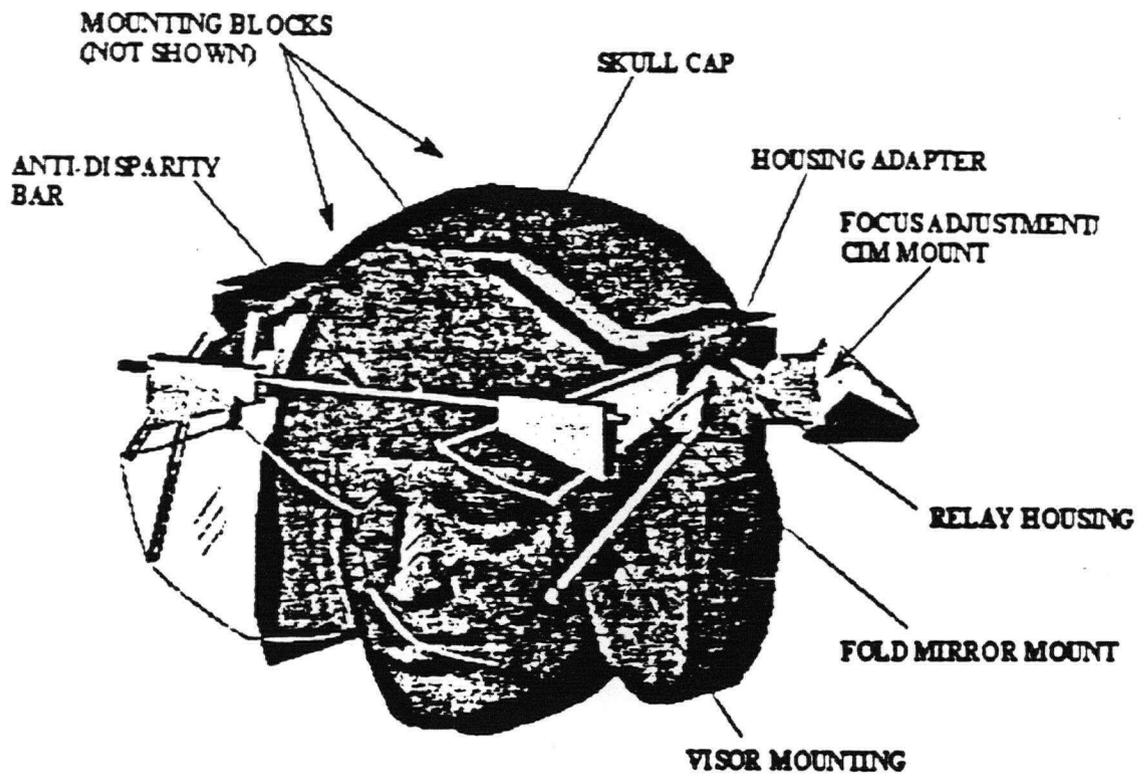


Figure 2.12 CONDOR HMD Components, [Kanahele, 96]

## Hardware System Design

As discussed earlier, even the most advanced DVC system can't meet all the expectations of a perfect presentation. On the one hand, this means trade-offs, on the other hand, this has also been a driving force for new technology and design ideas. Also, note that the display systems discussed above are for ground flight simulator, the images are computer generated and do not require the merging of AOI and WFOV video images.

But for tele-operation, the images are taken from remote scene in real time, and accordingly, the hardware for imaging WFOV and AOI as well as a processing system that merge the two images are required. There is today no system that provides these functions. Although not directly the topic of this thesis, some of the issues involved in the hardware implementation of WFOV and AOI imaging, merging and display have been studied, and a few design ideas are proposed. They are discussed in the appendices, and here is a brief introduction of the topics addressed.

**1 WFOV Imaging and Acquisition Using Intensity Array** As discussed earlier, one implementation difficulty is to provide wide field of view imaging. A pinhole camera is able to provide a wide field of view, but is limited by its aperture. One solution is to have a pinhole array and recover the image by deconvolution. If the array is properly arranged, there exists a deconvolution function to allow the recovery of the original scene from the overlapping multiple images obtained. In general, the array provides a  $\sqrt{N}$  times effective aperture compared to one pinhole where  $N$  is the number of pinholes. An imaging system based on a similar design principle is proposed, see Appendix A for details.

**2 Image Fusion Using a Microlens** This system is designed to project the insert image to the registered position over the background with optical and electronics devices instead of mechanical devices. The advantage is that it is less expensive and easier for real-time implementation. The idea is discussed in Appendix B.

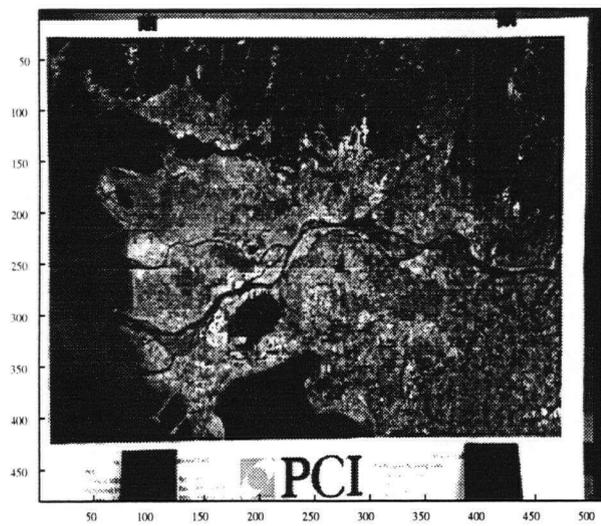
**3 Pixel Effect Removal For LCD Screen** In an HMD system, the viewing distance is close enough that the pixel details on an LCD screen become obvious and can affect the perception of the image features. That is, the gaps between the pixels can be seen and cause discontinuously in the image.

To solve this, a vibration based smoothing mechanism is proposed. This is based on the knowledge that if the vibration rate is higher than the critical fusion rate, the viewer will not be able to notice the motion, instead, the perceived image is the fusion of the consecutive frames. This will give the same visual effect as smoothing along the gap if the motion is crossing the pixel gaps. Further discussion of this is given in Appendix C.

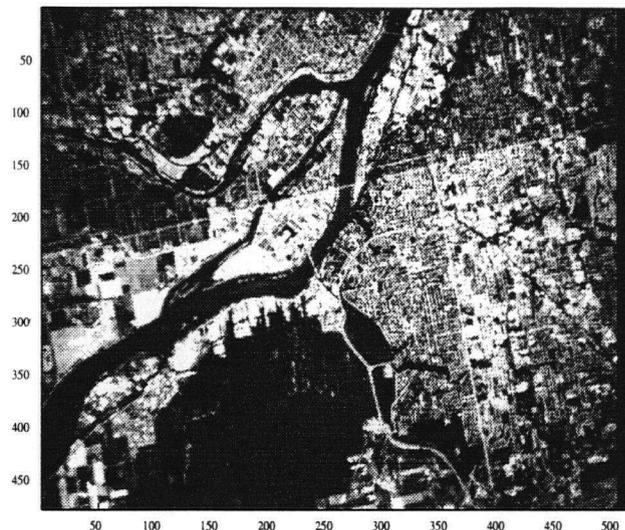
## Chapter 3 Processing System Design

We have discussed the DVC system specifications and implementation technology, next we will study the implementation of the processing system. As shown in Fig. 1.5, the computer acquires two video channel inputs AOI and WFOV, which have real-time rotation/scale/translational (RST) changes. Registration and fusion of AOI and WFOV are required before outputting to a display.

The requirements of the processing system are: given an initial RST estimation and their change range, a registered and fused dual resolution image should be generated in real-time. During this process, the RST estimation is the important step and also the most complex. Details are examined in this Chapter. This chapter will review the literature on this subject, the functions involved, and the problems to be solved. A brief introduction of fusion methods is also given. Fig. 3.1 illustrates why the processing system is necessary, since without RST estimation, the superimposed image is misaligned and not readable, as shown.



Background



Insert

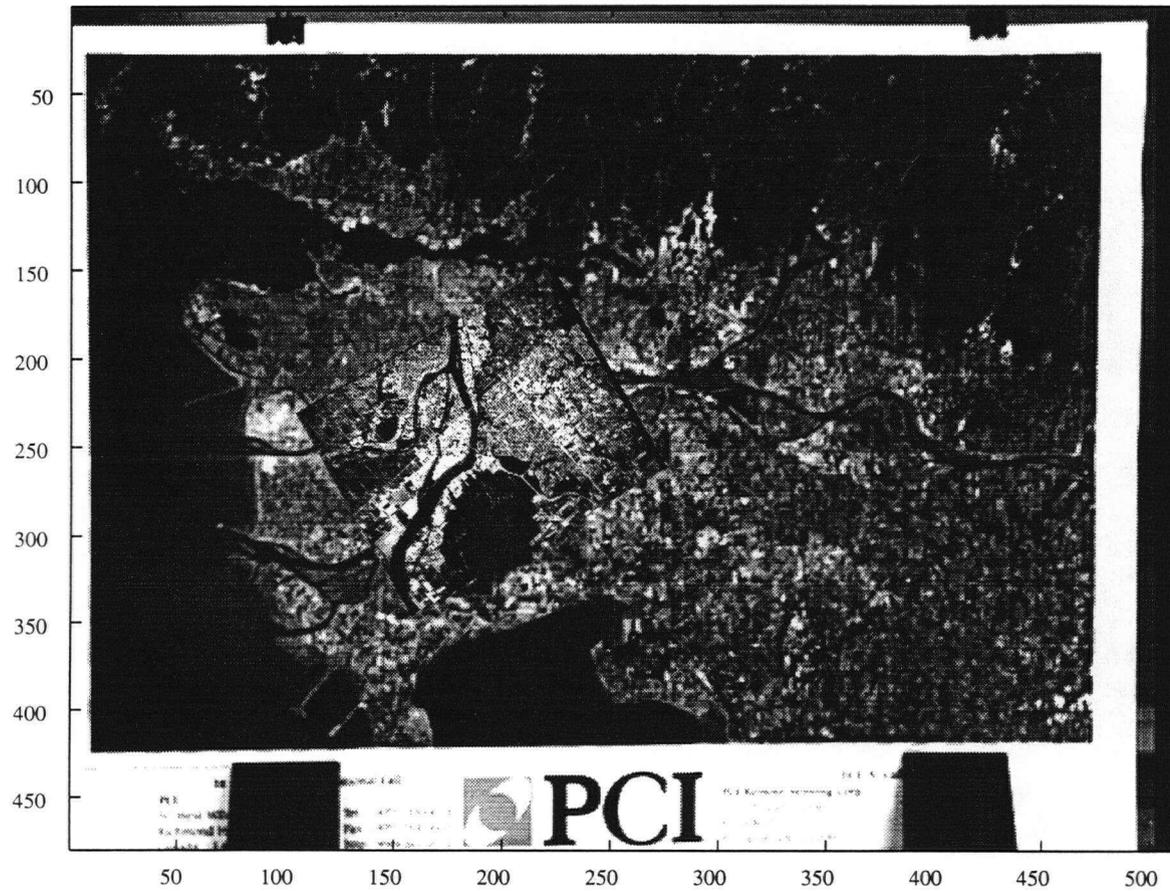


Figure 3.1 The Illustration of Image Fusion without RST Registration



### 3.1 Imaging Condition and the Transform Matrix

As discussed earlier, the AOI and WFOV images are obtained with relative RST changes. Their relation can be expressed with a transform matrix in Cartesian coordinates. That is,  $X' = T * X$ , where

$$X = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3.1)$$

and

$$X' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (3.2)$$

are the corresponding points in original image and transformed image, T is the transform matrix.

**Translation:** the matrix for translational change is:

$$T_T = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.3)$$

where  $x_0$  and  $y_0$  are the amount of shifts in X and Y directions.

**Scale:** scaling by factor of  $S_x$  and  $S_y$  along X and Y is given by the transform matrix

$$T_S = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.4)$$

**Rotation:** rotation transformation around the origin of the coordinates is given by:

$$T_R = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.5)$$

Finally, for RST, the transformation matrix, it is simply the multiplication of the above three matrixes

$$T_{RST} = T_R * T_S * T_T \quad (3.6)$$

### 3.2 Image Registration

Image registration is the process of recovering the unknown parameters in the transform matrix, i.e., the RST values. The purpose is to present two images at the same RST scale so that further similarity measure and the fusion process can be performed. A good review of image registration methods can be found in [Fua, 93]. There are two main categories of registration methods, one is correlation-based, the other feature-based. Correlation is generally regarded as a reliable and precise method in recovering translational (T) change [Guelch88], but the computational cost can be high depending on the correlation window size and search region. It is not very well known as to the performance of this method in recovering the RS values. In [Schmid & Mohr, 95], it is pointed out that correlation is invariant to small RS change, for up to 20% scale change, and a 15 degree rotational change. But, no details and rationale were given. This is enough if only translational change needs to be recovered. But when RS changes need to be recovered, it become necessary to understand in detail how well the algorithms will perform under these circumstances. This will be the topic of the next chapter.

Another category of image registration algorithms employs features. These algorithms extract features of interest from images, such as edge segments, closed-boundaries, contours or corner points and match them with some similarity measurements. The reliability and precision of these approaches depend on the feature contents of the two images to be matched, and therefore lack consistency over different pictures. Nonetheless, its robustness to changes, such as RST, motion and view changes, make it a popular tool in areas where the correlation fails, or the computational cost is too high.

Among the features, corner points have been widely used as control points in practice, because they are 2D features and are invariant to RST changes, which is especially relevant in our case. This is why corners are chosen in our study of the feature-based approach in recovering RST values. This is studied in Chapter 5, using correlation window in matching, and in chapter6, using attributes in matching.

### 3.3 Performance Evaluation

There are many factors that affect the performance, but free parameters are what we can adjust

to obtain the best results. So, understanding how to select these parameters become very important part of implementation.

Theoretically, free parameters should be selected based on their contributions to the performance index such as estimation accuracy, the reliability of the estimation results, and the computational cost. Chapters 4 to 6 study the relation between the performance index and the free parameters. In some cases, analytical equations are derived. These will provide the guidelines in selecting the free parameters in applications involving RST changes.

### 3.4 Sampling and Interpolation

Given two images and the transform matrix, one of the images needs to be transformed to match another one. This process involves sampling and interpolation which is introduced in this section.

The re-sampling with RST changes are as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = S \begin{bmatrix} \cos R & -\sin R \\ \sin R & \cos R \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (3.7)$$

or

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} S * \cos R * x - S * \sin R * y + x_0 \\ S * \sin R * x + S * \cos R * y + y_0 \end{bmatrix} \quad (3.8)$$

This equation shows that the new pixel values are not necessarily integer numbers, which means some kind of interpolation is required to obtain the new pixel values. The transformed pixels generally fall into the space between four neighboring pixels. The interpolation process is to determine the gray level value using these four pixels. Many possible interpolation schemes can be used.

**Nearest Neighbor Interpolation:** The simplest interpolation scheme is the so-called zero-order, or nearest neighbor, interpolation. In this case, the gray level of the output pixel is taken to be the value of the input nearest pixel. This is computationally simple and produces acceptable results for many applications.

**Bilinear Interpolation:** First order, or bilinear, interpolation produces more desirable results than zero-order interpolation, with only a slight increase in programming complexity and computation time. The bilinear function is given as:

$$f(x, y) = [f(1, 0) - f(0, 0)]x + [f(0, 1) - f(0, 0)]y + [f(1, 1) + f(0, 0) - f(0, 1) - f(1, 0)]xy + f(0, 0) \quad (3.9)$$

**Higher Order Interpolation:** A function with more than four coefficients is made to fit through a neighborhood of more than four points. The obtained subpixel value usually gives higher accuracy result at the cost of extra computation.

### 3.5 Fusion

Smoothing is needed in the transition band to bridge the gap between high resolution and low resolution images. There are several ways of doing the smoothing transition:

- (i) An average of the intensity values of the overlapped pixels.
- (ii) A median filtering of the intensity values of the overlapped pixels.
- (iii) A weighted median or a weighted average where the weights can correspond to one of the several choices, yielding very different types of fusion. For example, the weights can be chosen to decrease with the distance of a pixel from its frame centre.

## Chapter 4 Analysis of Correlation-based Image Registration Methods

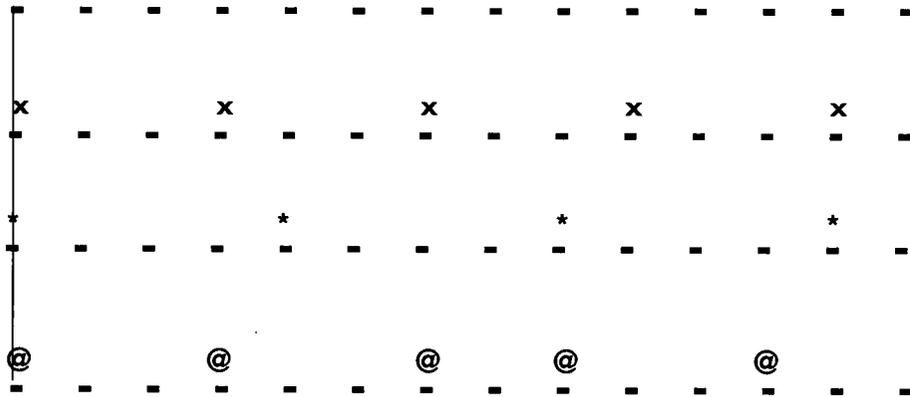
As discussed earlier, the first category of registration algorithms is correlation-based, and the free parameters consist of correlation window size, RST step sizes, the RST change ranges, and filter size. In the following sections, we will analysis how these free parameters affect the performance and experiments are carried out accordingly.

### 4.1 Estimation Accuracy

The first performance index is the estimation accuracy. It depends on how sensitive the correlation function is to the changes of variables to be estimated, since the correlation value is used to recover these variables. The higher the sensitivity, the higher the estimation accuracy. In applications where images have relative RST changes (i.e., RST values are the variables of correlation function), we look at the sensitivity of the correlation value to RST changes and establish the relation between the sensitivity and the free parameters.

**Sensitivity to S (scale) change:** Given two images I and I' with a scale difference of S, I' needs to be down-sampled by S to be correlated with I. If the scale difference between I and I' is  $S + \Delta S$  instead of S, then I' needs to be down-sampling by  $S + \Delta S$  to be correlated to I. If  $\Delta S$  is small enough, then down-sampling by S or by  $S + \Delta S$  should give the same sampling point locations. To illustrate how the sampling point locations are distributed as scale changes, a map is drawn as shown in Fig. 4.1. The figure illustrates the concept that down-sampling the same image by a different scale usually generates different sample points, if the scale difference is large enough. For example, as shown in Fig. 4.1, for scale=3 and scale=4, most of the sample locations are different. On the other hand, if the scale difference is small enough, all the sample points will be at the same location. In between, there will be partial overlap of the sample points between two down-sampling scales, which is the case for scale=3 and scale=2.67. Scale=2.67 has the same sample points as scale=3 up to the fourth sample point. So, if a correlation window of less than 8x8 is used, the correlation by these two down-sampling scales will give the same value, that is why correlation is invariant to small scale change, especially with a small correlation window. As  $\Delta S$  increases, the correlation value

difference between the two down-sampled images will increase accordingly, due to an increasing number of different sample points.

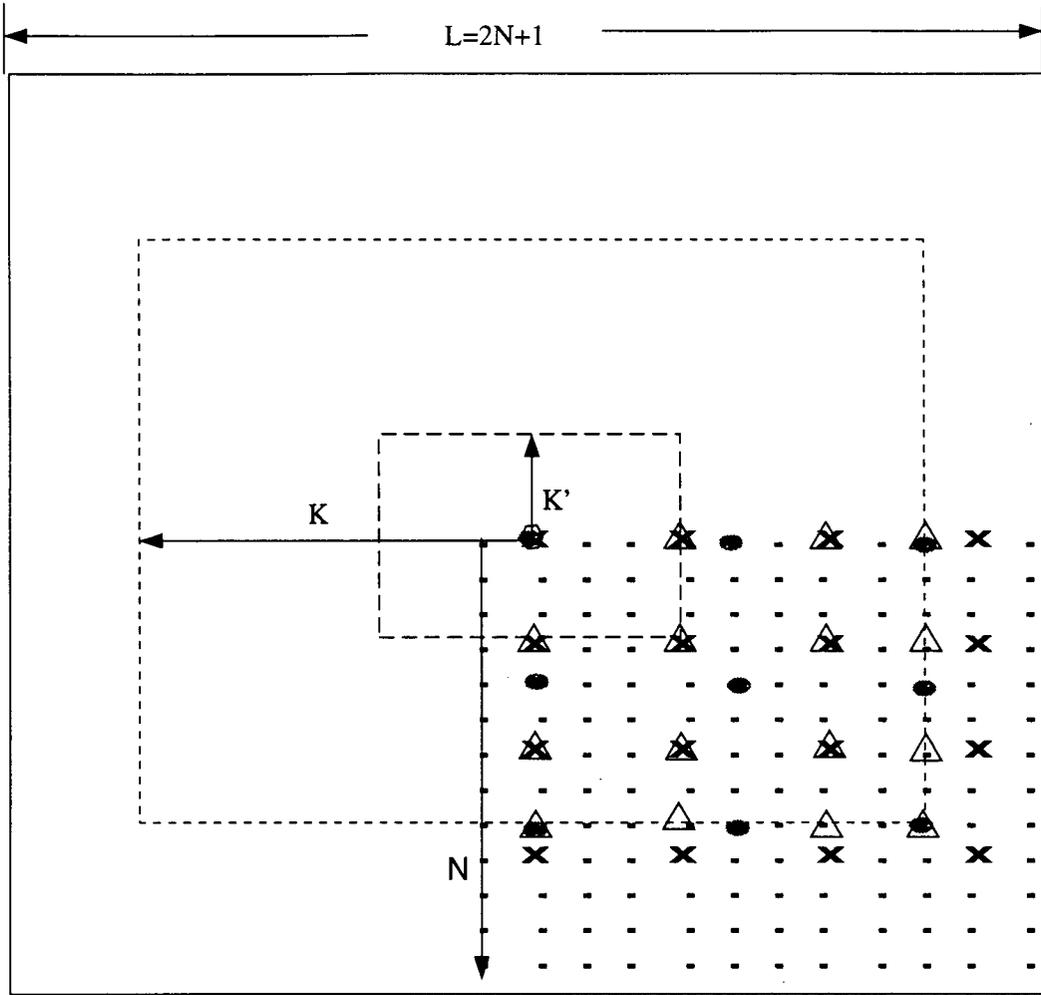


Vertical line represents the centre of the image  
 . are the original image points  
 x are the sampling points for scale=3  
 \* are the sampling points for scale=4  
 @ are the sampling points for scale=3\*8/9~2.67

**Figure 4.1** 1D Illustration of Sample Points Distribution at Different Scales

In general, the correlation value change is proportional to the percentage of sample point changes, given a uniformly random distribution of the pixel intensity level. This is illustrated as follows:

Given images  $I$  and  $I'$  to be correlated, the correlation value  $f_{corr} \propto \sum_i \sum_j (I_{ij} * I'_{ij})$ ; given images  $I$  and  $I''$  to be correlated,  $f'_{corr} \propto \sum_i \sum_j (I_{ij} * I''_{ij})$ , the difference of  $f_{corr}$  and  $f'_{corr}$  will be  $\Delta f_{corr} \propto \sum_i \sum_j (I_{ij} * (I'_{ij} - I''_{ij}))$ . Note,  $\Delta f_{corr} = \max$  if  $I'$  and  $I''$  have all different sample points, and  $\Delta f_{corr} = \min$  or 0 if  $I'$  and  $I''$  have all the same sample points. This explains the assumption we made earlier that the correlation value change is, in general, proportional to the percentage of sample changes.



**Figure 4.2** 2D Illustration of the Range of Sample Point Difference (Outside Dotted Line: scale=3 and 4; Outside Dashed Line: scale=3 and 2.67)

From Fig. 4.2, we can derive the mathematical relation of the changes. Given an original image with pixel numbers of  $L \times L$  or  $(2N+1) \times (2N+1)$ , and the sampling scale of  $S$ , total sample points of the resampled image is:  $((2N/S) + 1)^2$ .

Given another sampling scale at  $S + \Delta S$ , the first different sample point will start at  $K = S/\Delta S$  (note the unit of  $K$  is in pixels as in the original image  $L$ ). So, the total number of the same sample points are:

$$S_S = ((2K/S) + 1)^2 - 4((2K/S) + 1) \quad (4.1)$$

The total number of different sample points is:

$$S_D = S_T - S_S = ((2N/S) + 1)^2 - ((2K/S) + 1)^2 + 4((2K/S) + 1) \quad (4.2)$$

The percentage of the different sample points  $P_D$  is equal to  $S_D$  divided by  $S_T$ , i.e.,

$$P_D = 1 - ((2K/S) + 1)^2 / ((2N/S) + 1)^2 + 4((2K/S) + 1) / ((2N/S) + 1)^2 \quad (4.3)$$

or

$$P_D = 1 - ((2/\Delta S) + 1)^2 / ((2N/S) + 1)^2 + 4((2/\Delta S) + 1) / ((2N/S) + 1)^2 \quad (4.4)$$

From the equations, we can see given  $\Delta S = 1/N$ ,  $P_D \approx 0$ , and given  $\Delta S = 1$ , then  $P_D = 1$  or 100%. To understand these relations better, some real values are given in Fig. 4.3 & 4.4.

Window Size\ $P_P$	$\Delta S=0.2$	$\Delta S=0.1$	$\Delta S=0.05$
L=32, N=16	$\approx 1$	0.85	0.05
L=64, N=32	$\approx 1$	0.96	0.73
L=128, N=64	$\approx 1$	0.99	0.93

Figure 4.3 The  $P_D$  vs Window Size & Scale Change with  $S=4$

Window Size\ $P_s$	$\Delta S=0.2$	$\Delta S=0.1$	$\Delta S=0.05$
L=32, N=16	0.99	0.79	$\approx 0$
L=64, N=32	$\approx 1$	0.94	0.70
L=128, N=64	$\approx 1$	0.99	0.92

Figure 4.4 The  $P_D$  vs Window Size & Scale Change with  $S=3$

Note,  $P_D \approx 1$  or 100% is approximately equivalent to X/Y shift of one pixel in contribution to correlation change, so  $P_D < 0.8$  is regarded as not acceptable, unless with very high feature contents within the window. Note as  $S$  changes, not all the mislocated sample points are mislocated at the same distance. The larger the displacement, the larger the contribution to the correlation value change, but this is not counted here due to the calculation complexity. In general, we can assume that  $P_D=90\sim 100\%$  is approximately equivalent to one pixel shift in space in the amount of correlation change introduced.

In this and following analysis, the sensitivity values are derived with the nearest neighbor as the interpolation scheme, if linear or higher order interpolation is used, the sensitivity will increase accordingly.

**Sensitivity to R (rotational) change:** Similarly, the relation between the rotation change and the  $P_D$  values can be derived as: given a rotational change of  $\Delta R$ , the nearest distance of the different sample location will be

$$K = 1 / \tan(\Delta R) \quad (4.5)$$

$$N_D = ((2N/S) + 1)^2 - \pi(K/S)^2 + 2\pi(K/S) \quad (4.6)$$

where S is the base scale change and  $L \times L = 2N \times 2N$  is the original image size. The percentage of the sample point location change  $P_D$  is equal to  $N_s$  divided by the total pixel number in the correlation window, i.e.,

$$P_D = 1 - \pi(K/S)^2 / ((2N/S) + 1)^2 + 2\pi(K/S) / ((2N/S) + 1)^2 \quad (4.7)$$

or

$$P_D = 1 - \pi(1 / \tan(\Delta R) / S)^2 / ((2N/S) + 1)^2 + 2\pi(1 / \tan(\Delta R) / S) / ((2N/S) + 1)^2 \quad (4.8)$$

**Sensitivity to T (translational) change:** Correlation can recover space shift T up to one pixel precision, with  $P_D=1$  or 100%, or sub-pixel precision can be obtained if a certain interpolation scheme is applied.

As analyzed above, we established the relation among the window sizes, the RS step size and the correlation change which provides the criteria in selecting the proper window size, step size for the precision required, and avoid unnecessary detailed calculation.

## 4.2 Reliability of the Estimation Results

The reliability of the correlation-based methods in relation to the window size is well studied, as in [Fua, 93]. The larger the window size, the higher the reliability, since larger window size covers more feature contents in general.

The reliability of the correlation results is also related to feature content, the higher the feature bandwidth, especially the higher feature components, the higher the reliability. So, some kind of high pass filtering, like Laplacian, can be introduced to enhance the high frequency feature contents. This process can be regarded as feature extraction. Among the features, the corner map is another choice because corners are 2D features and sensitive to changes in all directions (in RS space). Our experiments shows that if correlation is done on the corner map, the reliability will increase accordingly. The disadvantage of this process is that a high pass filter tends to increase the noise level, so some kind of smoothing should be applied before feature extraction, and all these make it computationally expensive.

Another way to increase the reliability of the correlation is to have the feature density measurement beforehand. For example, variance can be calculated first to avoid unnecessary correlation in the featureless windows.

### 4.3 Computational Cost

The computational cost depends on correlation window size, search range, and what kind of preprocess is applied. The computation cost is proportional to the correlation window size, filter size and search range, and inverse proportional to the RST step sizes. These are summarized in Fig. 4.5. The numbers given are based on the space domain calculation, without considering possible computing skills that can reduce the computation cost. Our main purpose here is to present the general relation between the free parameters and the computational cost.

Methods (Image LxL, Window WxW (=K/S)X(K/S) in Fig. 4.2), No. of Scale Steps $S_N$ , No. of Rotation Steps $R_N$ , Translational Change $T_x$ & $T_y$ , Filter Size FxF)	Multiplications Required	Typical Number with W=8; F=3; L=480; $S_N = 10$ ; $R_N = 10$ ; $T_x = 10$ ; $T_y = 10$
Direct Correlation	$(2W^2 + 3) * T_x * T_y * S_N * R_N$	1300k
Smoothing	$(2F^2 + 3) * L^2$	4838k
Variance Normalization	$(3W^2 + 3) * T_x * T_y * S_N * R_N$	1950k
Laplacian Filtering	$(2 * F^2 + 3) * (W^2 * S_N * R_N + L^2)$	69350k
Plessey Corner Map	$5L^2 + 3 + (5W^2 + 3) * S_N * R_N$	5792k

Figure 4.5 Computational Cost vs Factors (k=1000)

In Fig. 4.5,  $S_N$  stands for the number of scale steps,  $R_N$  stands for the number of rotation steps, and  $T_x$  &  $T_y$  stand for the translational changes in X and Y dimensions.

#### 4.4 Brute Force Search and Hierarchical Approach

Given an RST change, the natural approach to estimation is by searching through RST space, which is a brute force search. In a brute force search, for each set of RST values, warping/resampling is needed for correlation based matching to give meaningful results. That is, a counter rotation, counter scale is done to recover the similarity of the gray level images. The brute force search can be computationally expensive, requiring warping, resampling and correlation for each set of RST values. A hierarchical approach can be introduced to address these concerns and to give significant functional and computational advantages. A coarse-to-fine approach reduces the search space at finer levels by using the results of matching at coarser resolutions. This approach has been used in many applications involving searching over large space to reduce the computational cost, [Burt, 93].

Fig. 4.6 show the buildup of the multi-resolution hierarchy. The lower resolution is simply obtained by replacing four neighboring pixels by one averaged pixel. The picture size is truncated to square and power of two before building the hierarchy.

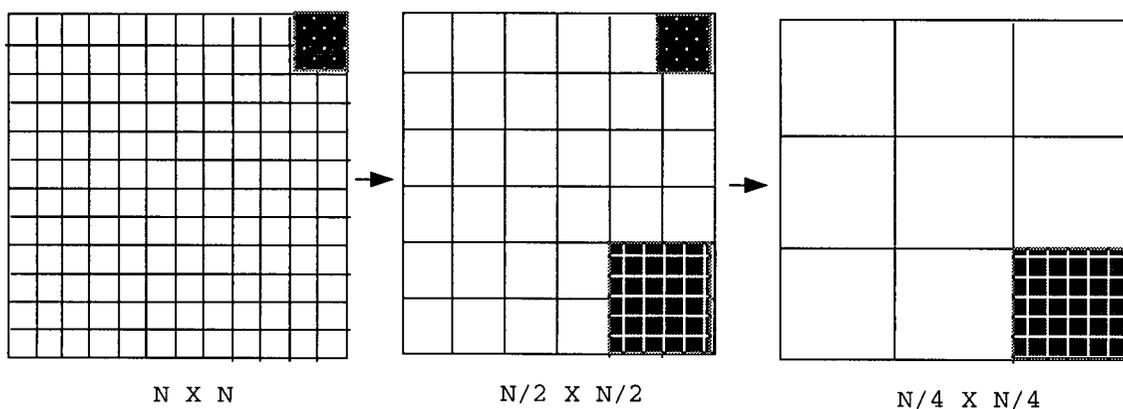


Figure 4.6 Building Multi-resolution Hierarchy

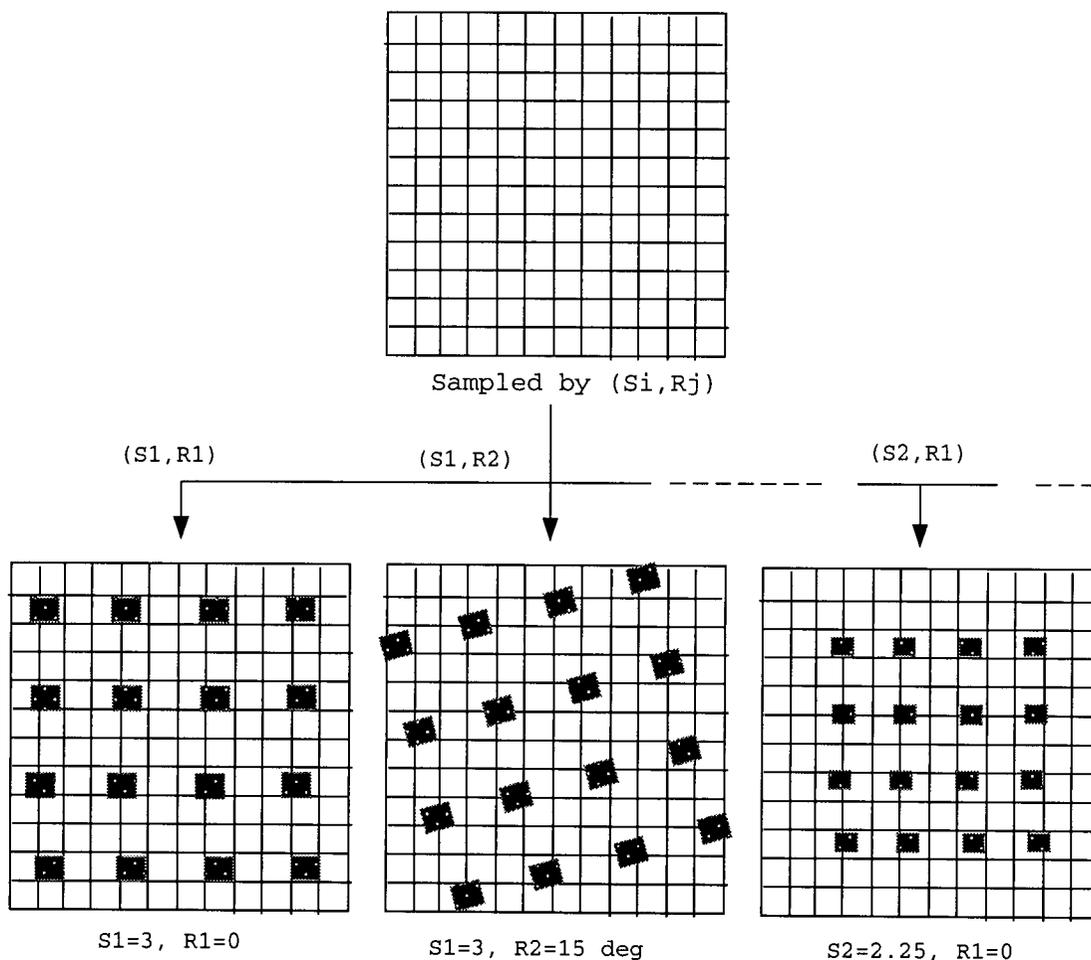
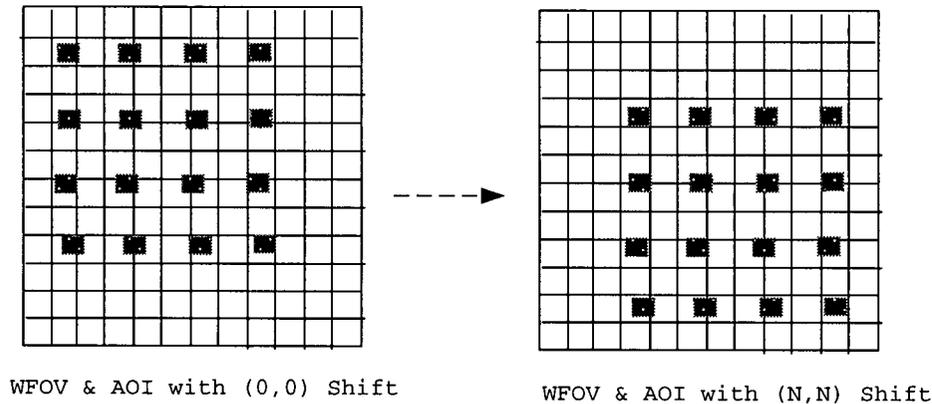


Figure 4.7 Sample Points by Different RS Sampling



**Figure 4.8** Mapping Down-sampled AOI to WFOV

As shown in Fig. 4.7, the AOI is down-sampled by different RS values, these are selected based on the initial estimation of RS value, and the change range of RS. The step size for R at lowest resolution level is chosen to be  $2/N$  (in radians) where N is the correlation window size. Step size of S is chosen to be  $2/N$ . Step size for T is chosen to be equivalent to the number of the hierarchy at the bottom level search and decreased to one at top resolution level search. The RS step size at higher levels will always be half the step size of the previous level. This is shown in Fig. 4.8. Note, T is chosen to be as the maximum point in T search space, i.e., the amount of shift that gives highest correlation value will be chosen as the T value. The search over RS will generate an RS surface map, which will be shown in experiments, and the maxima over the RS surface will give the new initial RS values for the next level search.

This approach also make it possible to use a different step size and window size for the different resolution levels, using the analytical results derived in the previous sections. Proper window size, step size, and resolution is the key to the success of this process. As analyzed earlier, a window that is too small is not sensitive to RS change, and has low reliability. A window size that is too large can be computationally expensive and unnecessary, as analyzed previously. Similarly, too small a step size for RST change can be redundant since the change is too small to make a difference in correlation, and too large of a change of RST can result in missing the global maximum and end up with a local maximum. The flowchart of this process is shown in Fig. 4.9.

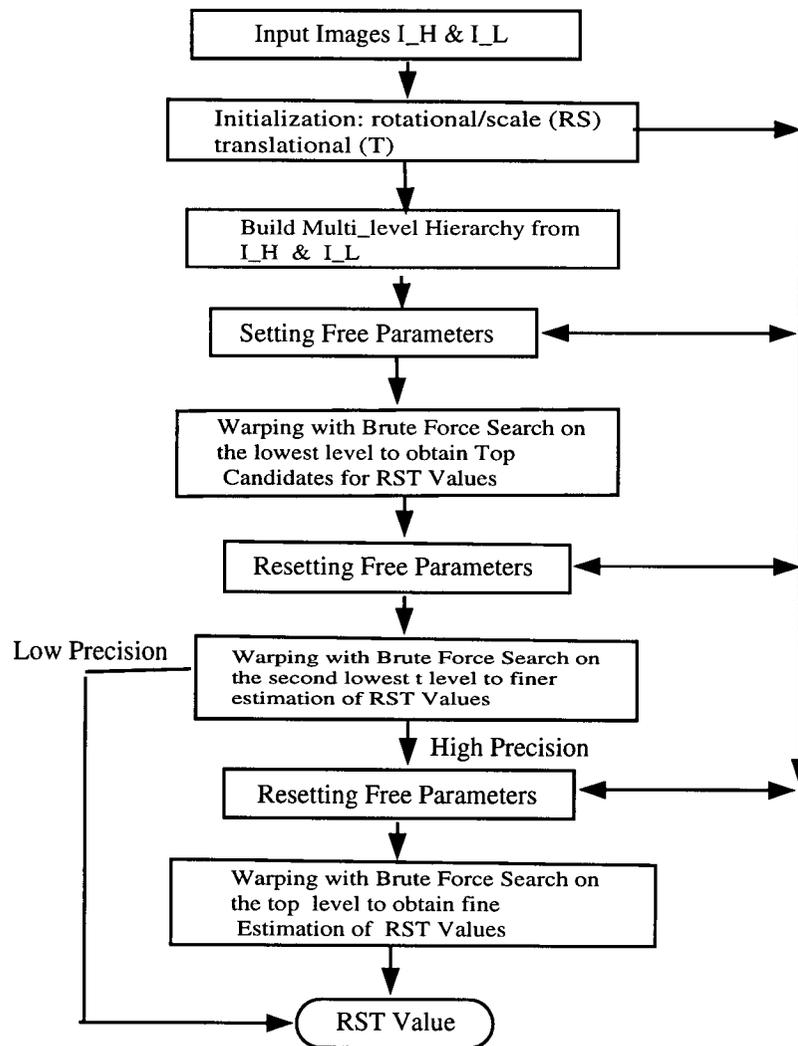


Figure 4.9 The Flowchart of the Hierarchical Brute Force Search Approach

#### 4.5 Other Options

An alternative to the above approach is an optimization approach. The basic idea behind the optimization approach is to fit the known points to a certain line/surface, and locate the next point according to this approximated line/surface. A global optimization approach can be used to locate a maximum point without any knowledge of the location. Unfortunately, all the global optimization approaches need intensive computations, since usually multi-starting points are used to locate global maxima.

If global optimization approach is not a good alternative to a brute force search at the coarsest level, there is still a possibility to use optimization after a first level brute force search when a global maximum point region is narrowed down. There are many local optimization approaches. We tried F\_R approach in our experiment. This approach is developed directly from the conjugate gradient method for solving linear systems, and it is regarded as a good approach for a non-quadratic function.

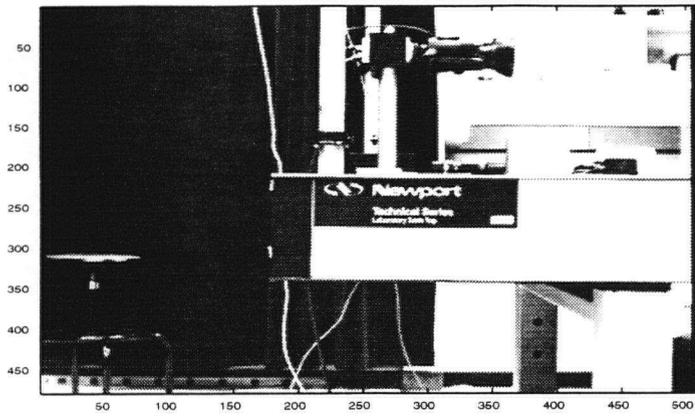
The minimum steps of an F\_R search is four RS space points to provide initial gradient values. And it takes more to converge to the true values. The converging speed depends on the image features and the step size of the line search. To adjust for all the images, a relatively smaller step size has to be chosen in a line search to guarantee the convergence. This means slower convergence speed in general, and no significant benefit in using this approach. In the case of restricted types of images, proper step size can be chosen to speed up the convergence.

An alternative is to use quadratic surface fit (QSF) or quadratic line fit (QLF) instead of going through the optimization loop. As we mentioned earlier, most optimization approaches use some kind of surface/line fit to determine the direction of the next step. Instead of using this surface/line fit to choose next direction, we use it to obtain finer RST estimation, or one step optimization. QLF is proved to be more efficient than QSF in our experiments. The steps applied to obtain finer maximum points are as follows: calculate the correlation surface through BFS, locate a maximum point over the surface using non-maximum suppression as discussed earlier, calculate 1D quadratic line along the maximum R and S point, locate a maximum point along the line, and the substep estimation of RS values are obtained.

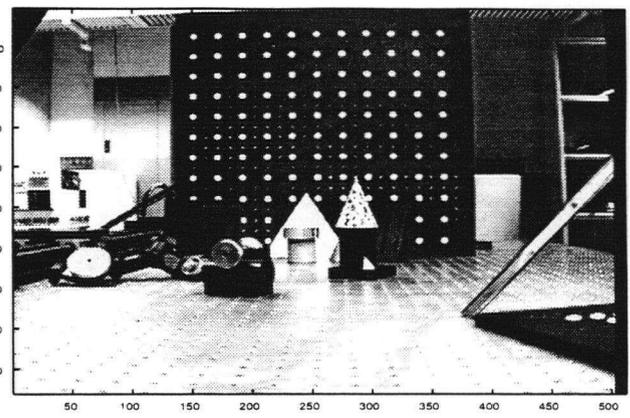
## 4.6 Experiments

In the experiments, 7 sets of images are taken with different zoom, different rotation, and different space shift, in the LCI lab of computer science department, in University of British Columbia. The optical axes of the two pictures are parallel, and the steps shown in Fig. 4.9 are applied to these images. The seven WFOV images are shown in Fig. 4.10, and the seven AOI images are shown in Fig. 4.11. & 4.12 shows the correlation surface over the RS steps chosen, and it shows that the step size can't be reduced further without missing the global maximum; at the same time, they are also sufficient to provide enough details for reliable estimation without introducing significant noise, i.e.,

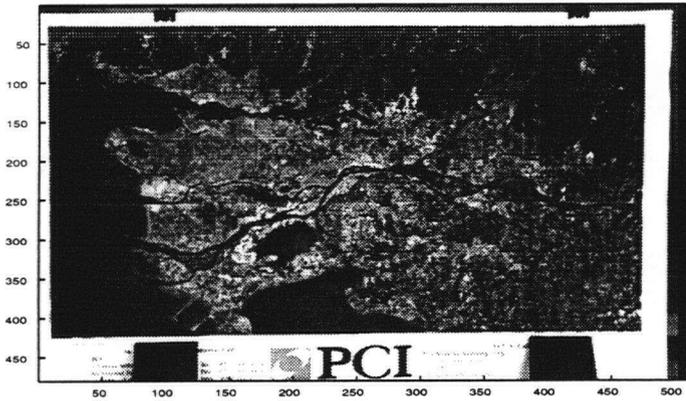
sufficient to provide enough details for reliable estimation without introducing significant noise, i.e., the surface is smooth enough. Note, this selection must suit all the picture, without any knowledge of the feature contents, which is the case for general tele-operation systems. So the RS step size might not be optimal for each individual image, but it is optimal for general images. A non-maximum suppression process is used to obtain the global maxima. First, over the correlation surface obtained after a brute force search over RS space, points that have values higher than all their eight immediate neighbors are assigned as local maxima, others are set to zeros. Then, the maximum among the local maxima is chosen as the global maximum by simply sorting the values. This global maximum point is used as a new set of initial RST values to set the next level RST search. Substep accuracy can be obtained by some kind of interpolation, here a quadratic line fit is used as discussed earlier.



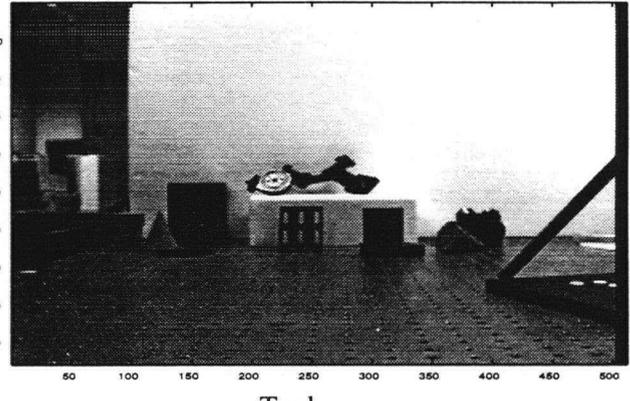
CS\_LAB



Toys



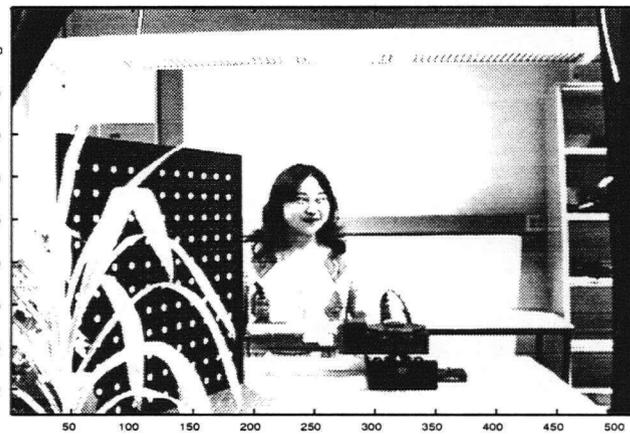
Map



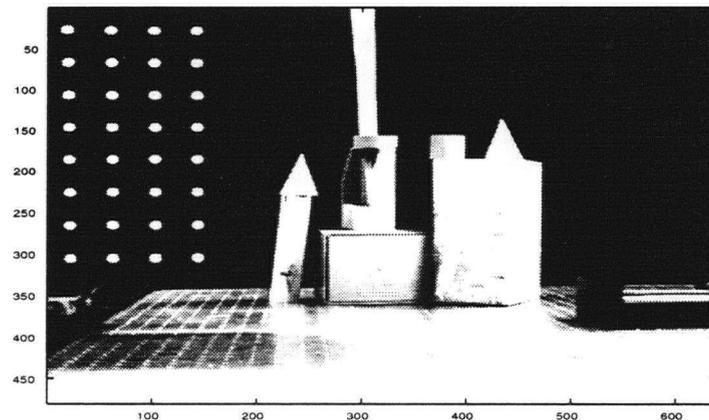
Tool



Plants



Me



Box

Figure 4.10 The WFOV Images of the Seven Sets

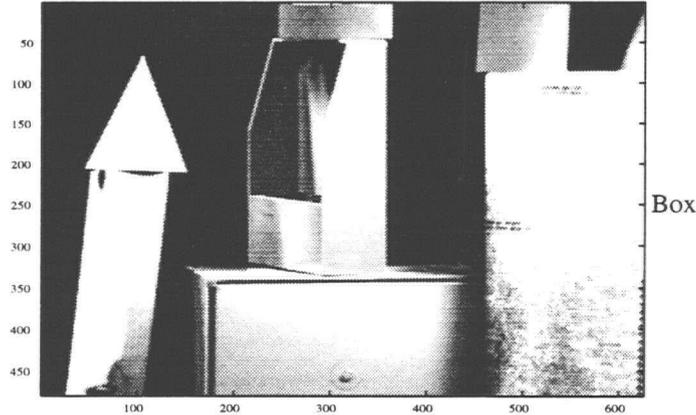
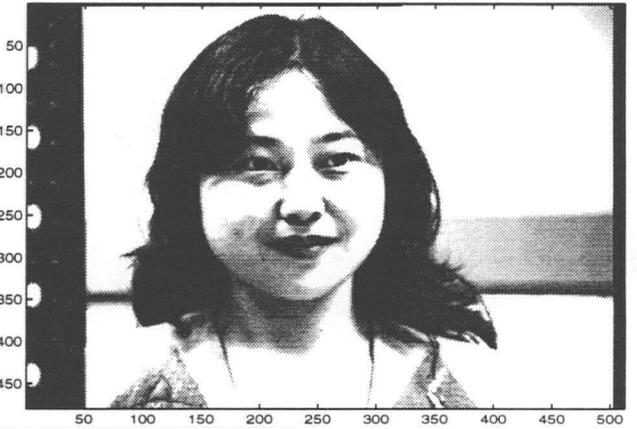
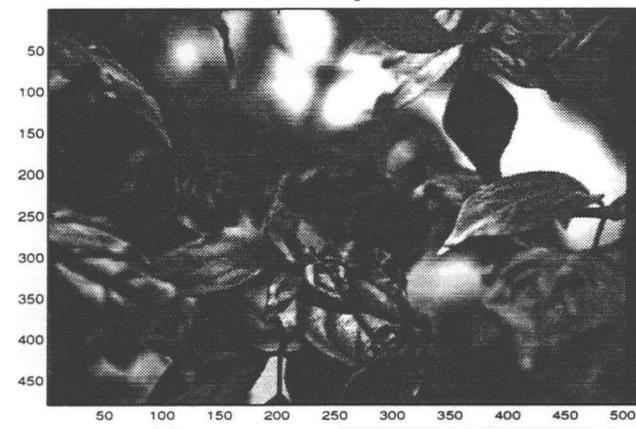
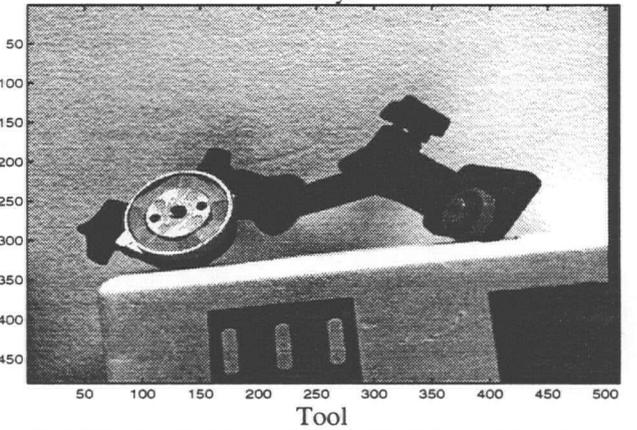
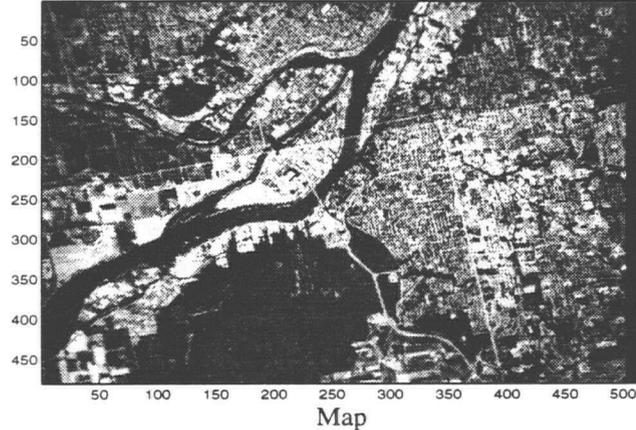
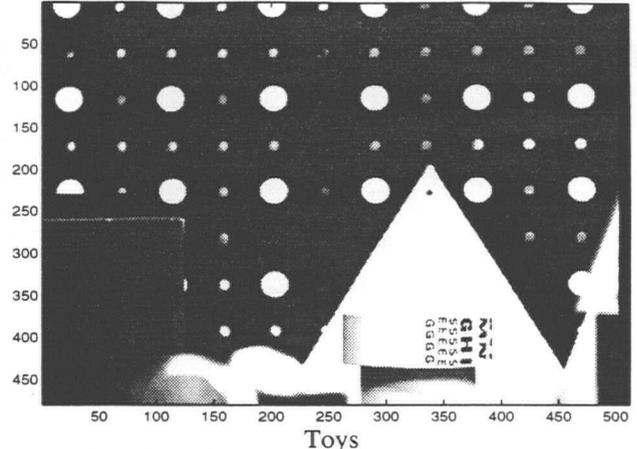
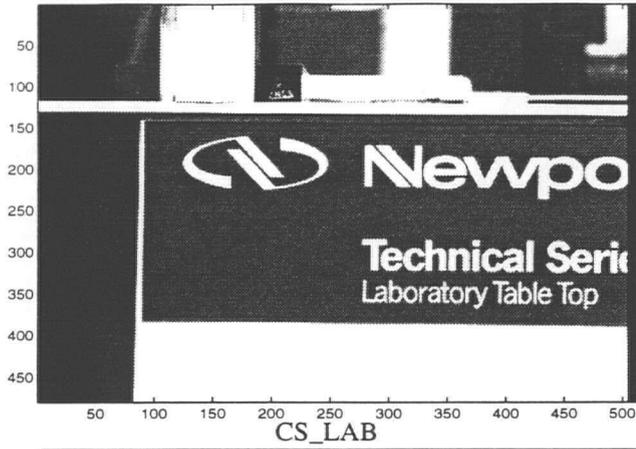


Figure 4.11 The AOI Images of the Seven Sets

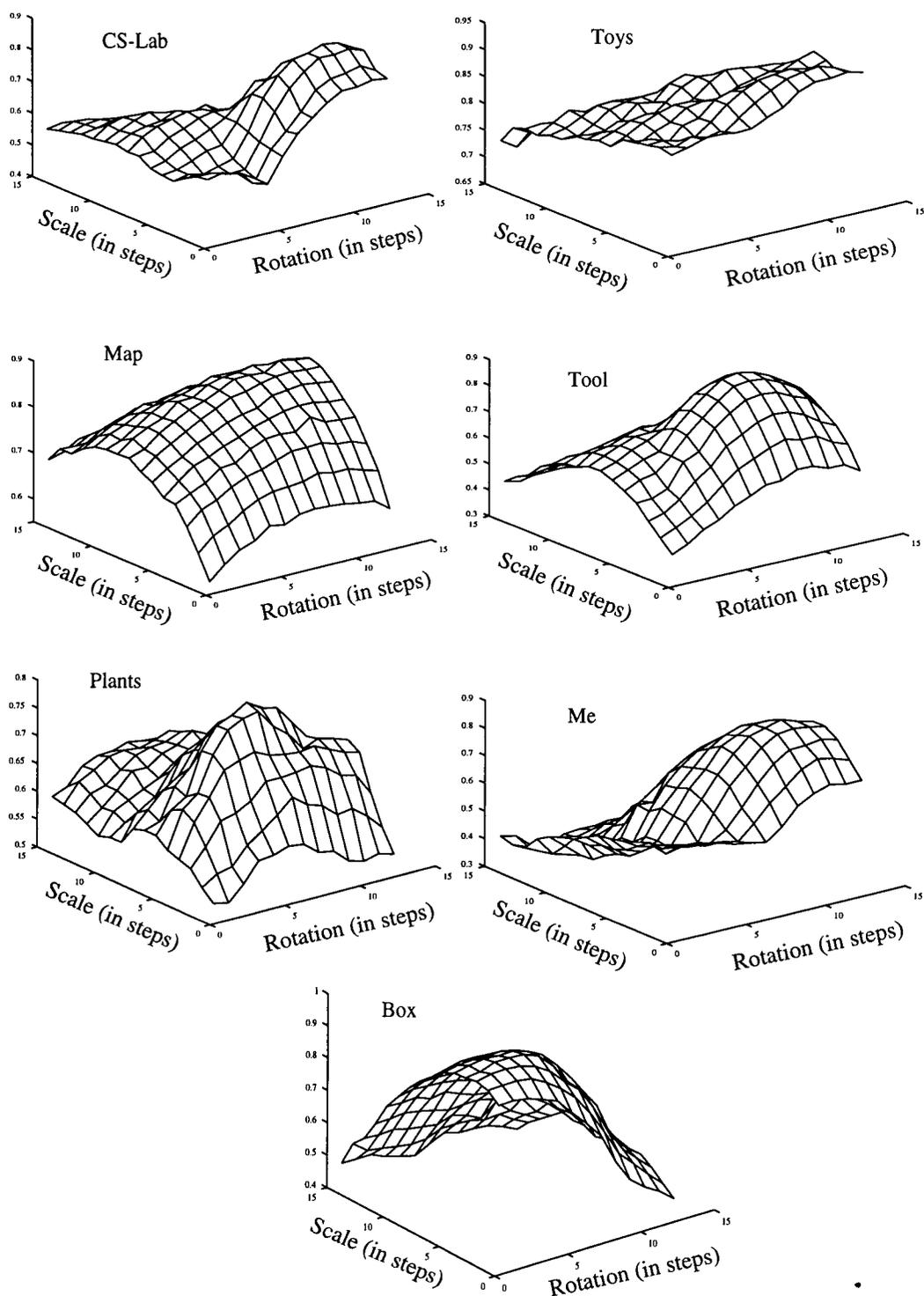
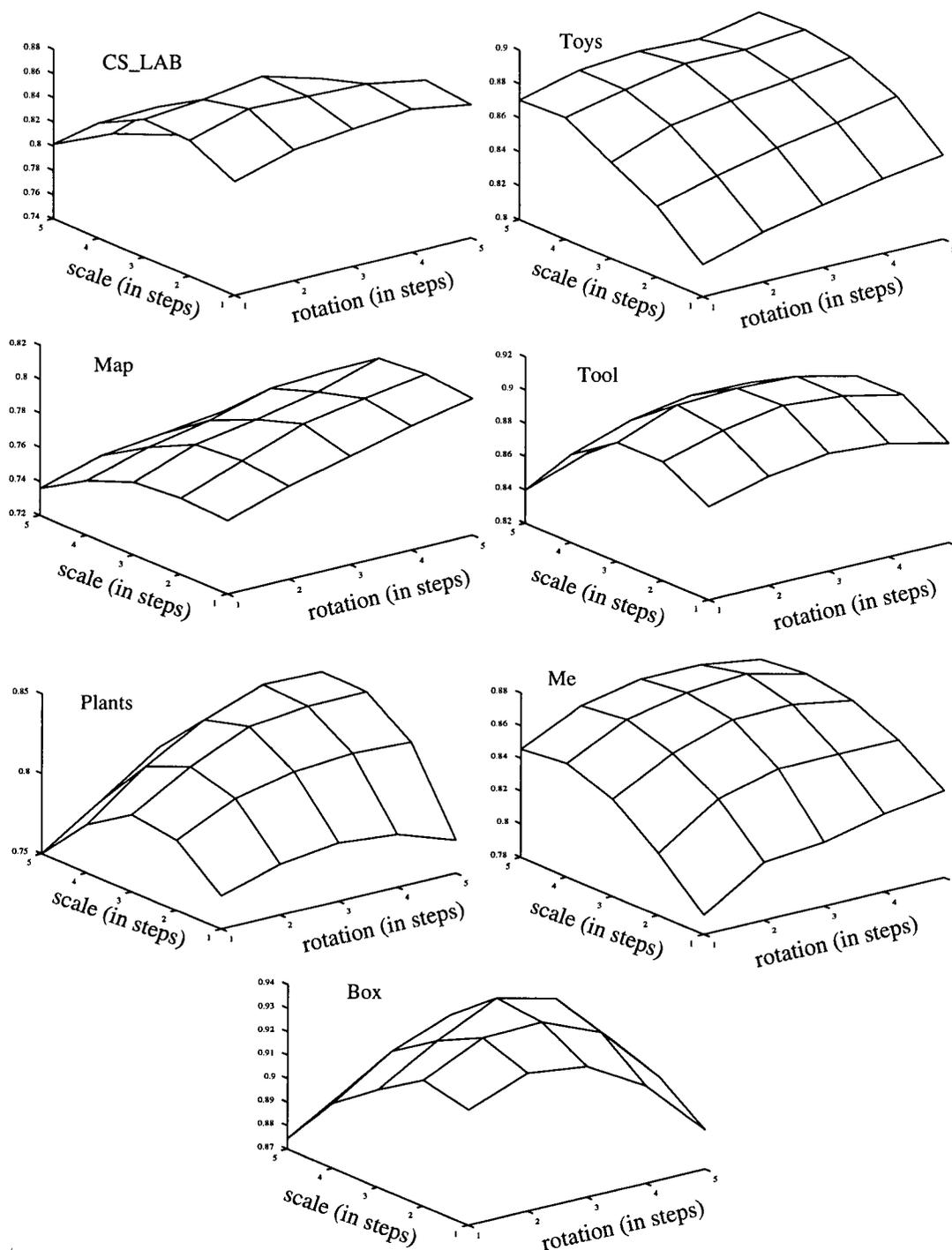


Figure 4.12 Correlation Surface at Coarsest Level with the RS Step Size Chosen for  $P_s > 90\%$



**Figure 4.13** Correlation Surface at Second Level with the RS Step Size Chosen for

Fig. 4.13 shows the correlation surface at the second coarsest level. These surface give us ideas about the variety of the characteristics of the images processed and how the surface shape changes

over the resolution level, allowing finer level of estimation.

These surfaces contain a lot of information about the sensitivity, and the noise level, etc., and can be used as a reference as to whether certain approach or step size is proper or not. Since all the calculations have to be done to obtain this surface, they can only be used as a way to evaluate how reliable the results are, not as a way to guide the selection. For example, we can tell, surfaces like CS-lab and Tool are well shaped to give good estimation, while surfaces like Map and Toys can be vulnerable to noise. These are reflected in the results.

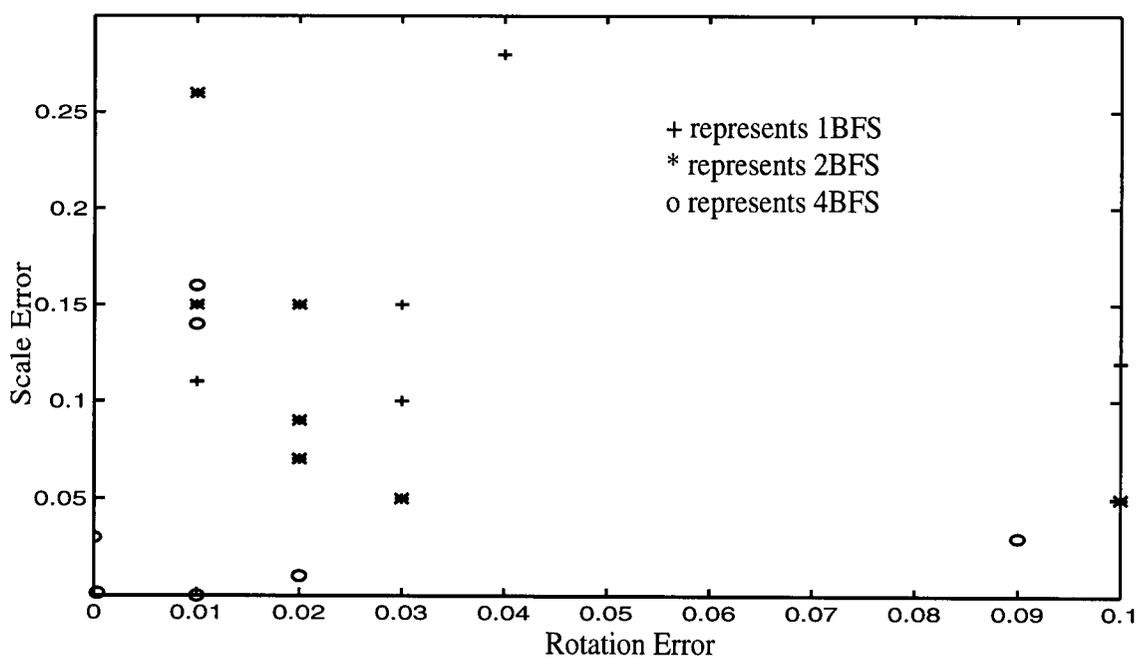


Figure 4.14 Comparison of the RS Estimation Error After Different Levels of Brute Force Search (BFS)

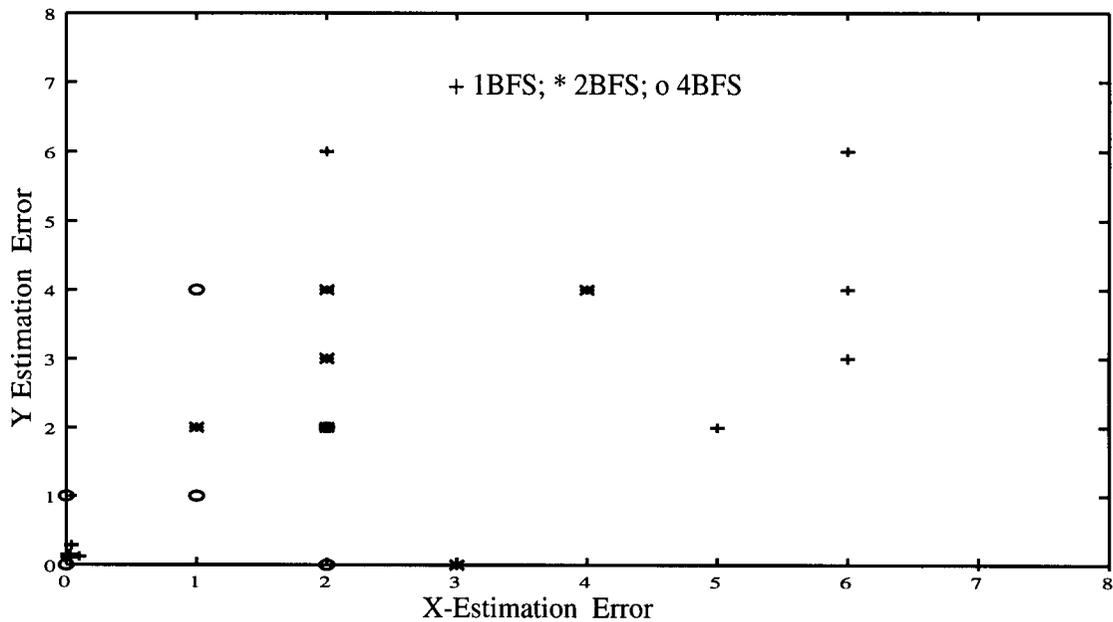


Figure 4.15 Comparison of the X/Y Estimation Error After Different Levels of Brute Force Search (BFS)

Estimation errors after one, two and four levels of brute force search are given in Fig. 4.14–4.16 to give an idea how the estimation errors decrease as the resolution level increases. The outliers that are located further away from the origin are not desirable and should be reduced as much as possible. There are some overlapping points, nonetheless, we can see the outliers decrease for higher level BFS. For details, see the tables in the Appendix D. This picture mainly shows the range of the estimation errors. From the figure, we can see that a first level brute force search estimation is not accurate enough to give acceptable registration result for image fusion, while two levels brute force search is close to the acceptable range. Fourth level is used for high performance system, at the cost of computation time. The final fused image is shown in Fig. 4.17. Note that only part of the fused image are given to show more clearly the transition band between the high resolution and low resolution transition. The central part marked by cross is the high resolution part.

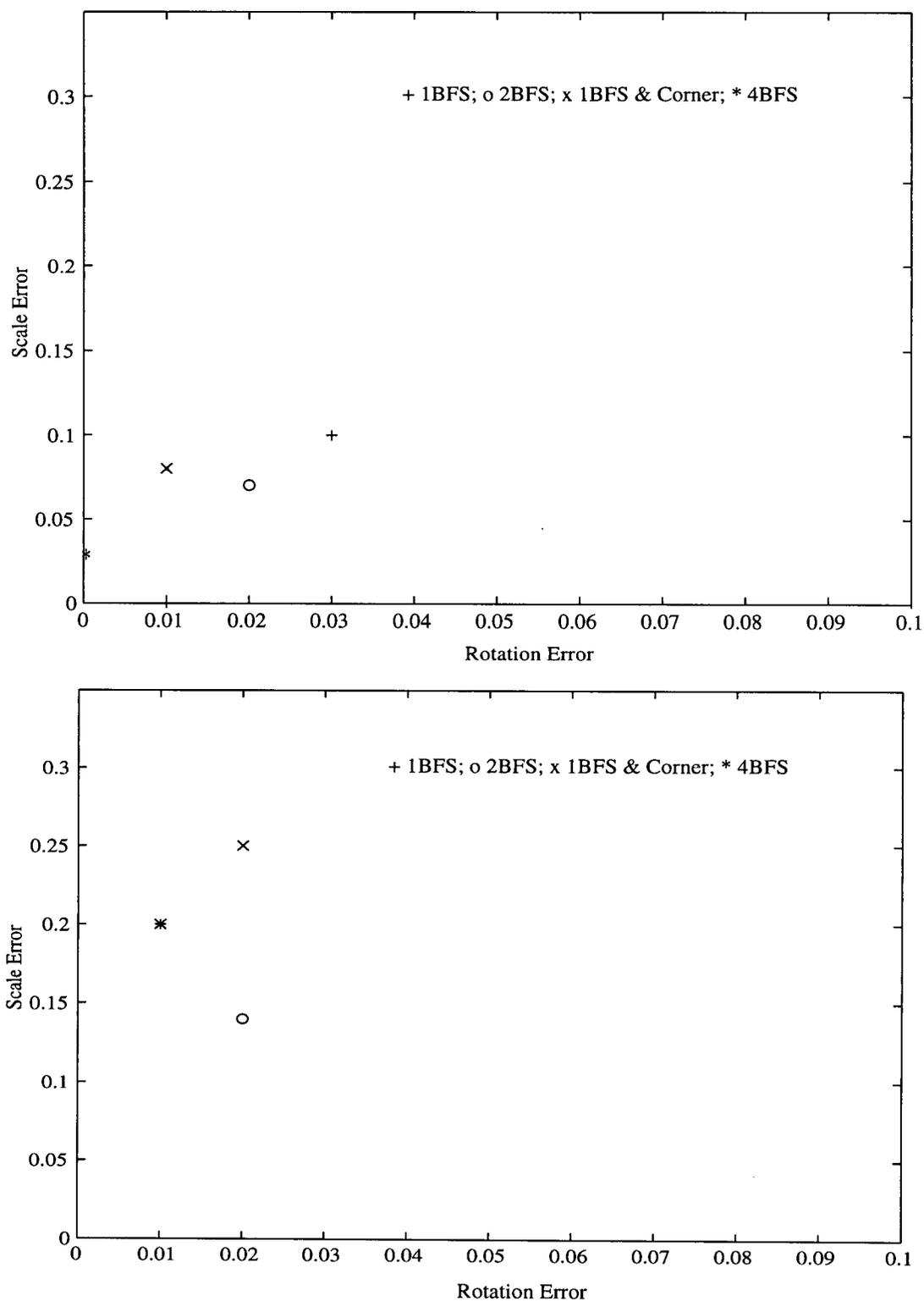
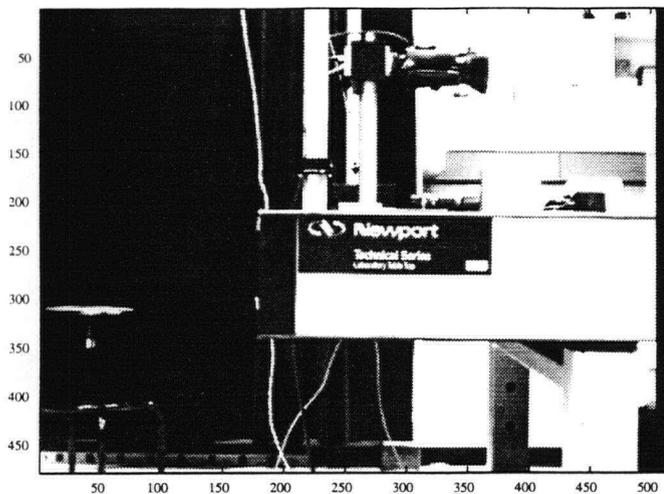
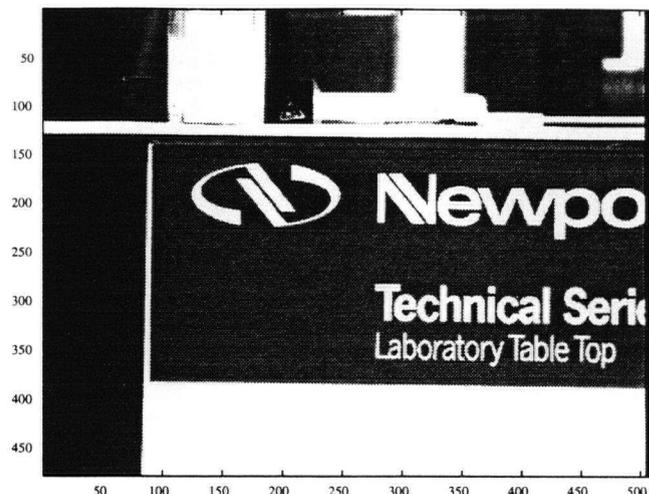


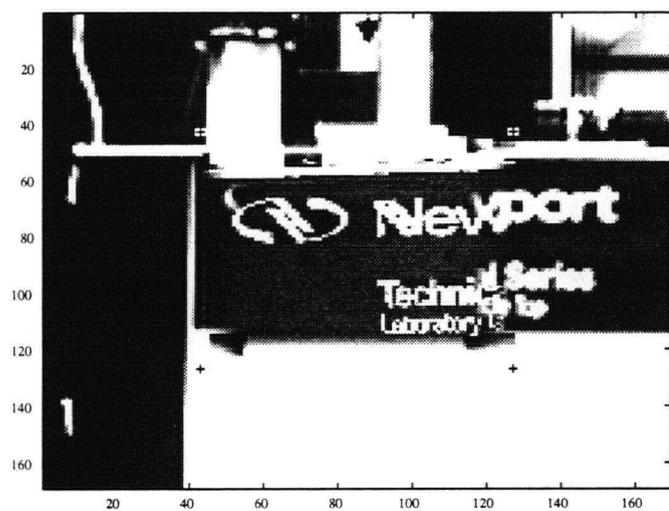
Figure 4.16 Comparison of Estimation Errors for CS-Lab (top) and Map (bottom) Pictures.



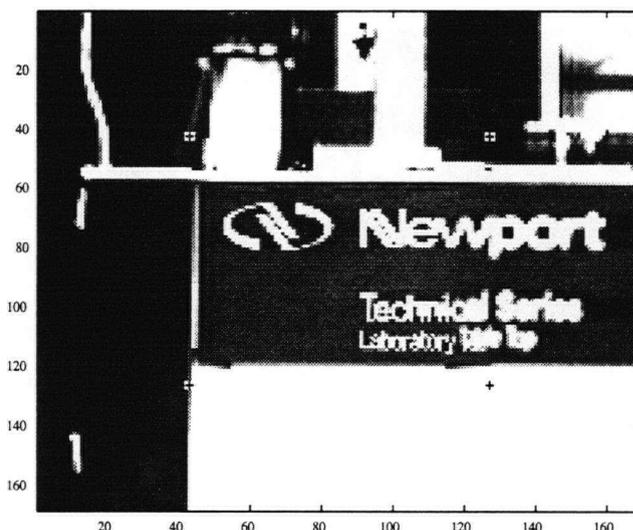
Background



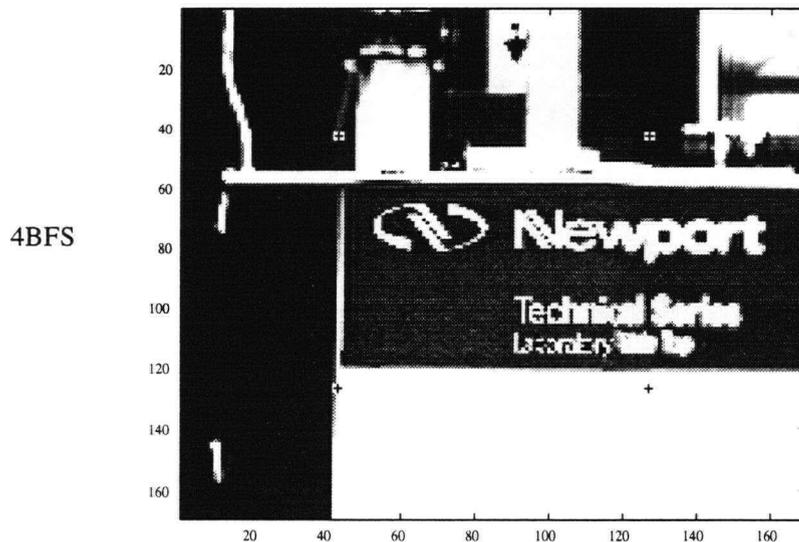
Insert



1BFS



2BFS



4BFS

Figure 4.17 The Fused Image After Different Levels of BFS

## **Chapter 5 Using Corners as Control Points to Recover RST Changes**

Corners are defined as two dimensional discontinuity points, as measured using gradients, higher order derivatives or their combinations. Control points are the sets of corresponding points in two images to be registered. They are used to solve the transform matrix to obtain RST values. This chapter will discuss the approach of using corners as control points.

In general, it is difficult to predict the exact performance of an approach using corners as control points, especially in applications where feature contents are unknown and unpredictable, such as in tele-operation, since this approach is highly feature dependant. Regardless of that, still there are other factors affecting the performance that can be analyzed, evaluated and selected to obtain the best performance in practice and will be studied in detail in this chapter.

### **5.1 Introduction**

The basic steps in using corners to recover changes are as follows: (1) corner detection; (2) corresponding corner matching; (3) change estimation. The change can be rotation/scale/translation (RST), motion, or distortion, etc. The success of the approach highly depends on the feature contents, and the free parameters chosen. In many applications, such as tele-operation, feature contents are unknown or unpredictable. This means only the free parameters can be adjusted in the implementation to obtain a better performance. Little literature is available to understand how these parameters are related to the performance, and they are usually chosen based on trial and error. This is the issue to be discussed in the following sections.

From the performance point of view, we have accuracy, reliability and computational cost to be concerned with, while the free parameters in the approach consist of the detector, the number of points, the location and distribution, the search range and the matching scheme. These parameters have different impacts on the performance. Some only affect one or two performance index, others affect all. This is illustrated in Fig. 5.1, where the related factors are connected with solid lines. The disconnected factors do not affect each other.

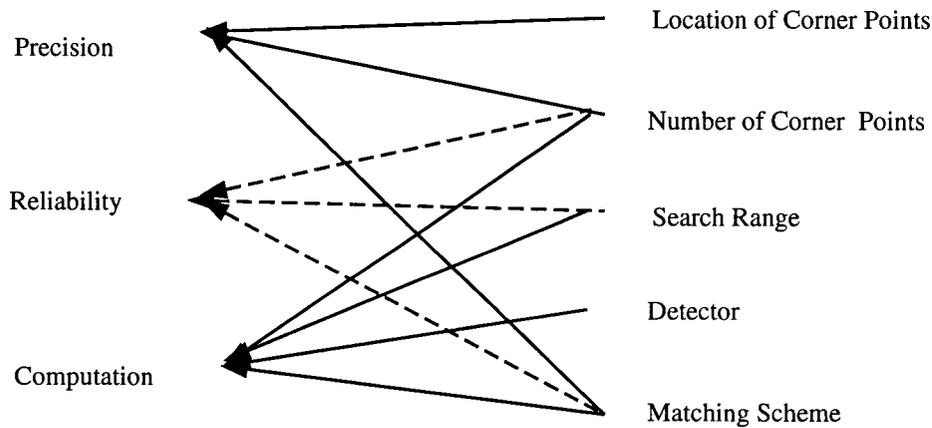


Figure 5.1 The Relation Between Performance Index and other Factors

Essentially, all corner detectors are ways of measuring the cornerness, defined as the product of gradient magnitude (a measure of 'edgeness') and the rate change of gradient direction with gradient magnitude, i.e.,  $C = (I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y) / (I_x^2 + I_y^2)^{3/2}$ . Threshold is used to select corner points from the cornerness map, see [Noble, 88]. In our analysis and experiments, Plessey corner detector is used for its simplicity, (only first order of derivatives are used), and it is assumed that corners are properly detected. Cornerness measurement is In the following section, we will try to establish analytical relations between connections in Fig. 5.1.

## 5.2 Estimation Accuracy

Accuracy is dependant on the sensitivity of the function to the variables to be estimated. That is, if the function values change significantly while its variables changes, that gives high sensitivity. Then, the function can detect smaller variable changes and gives higher estimation precision. In this case, the function is the similarity measure which is correlation, and the variables are RST. Next, we will analyze how the sensitivity varies with the free parameters.

**Corner Location:** Given two images  $I$  and  $I'$  with scale difference of  $S$ , a point  $A$  in  $I$  matches to point  $A'=D*S$  in  $I'$ , where  $D$  is the distance of  $A$  from the image centre. Given the scale change  $S'$  instead of  $S$ , the mapping will lead to a different point  $A''=D*S'$  in  $I'$ . The change between  $A'$  and  $A''$  is  $\Delta A = D * (S - S') = D * \Delta S$ . This is shown in Fig. 5.2., we can see the larger the distance  $D$ , the larger the location change  $\Delta A$ . This change is the base that scale change can

be detected. The reason is larger location changes usually leads to larger correlation changes, and correlation value is used to detect the scale change. This leads to the fact that the distant corners give higher scale estimation accuracy. The same is true for rotation estimation, though translational estimation accuracy is not affected. This is shown in Fig. 5.3, where point A->(A) is less sensitive to R change than point B->(B).

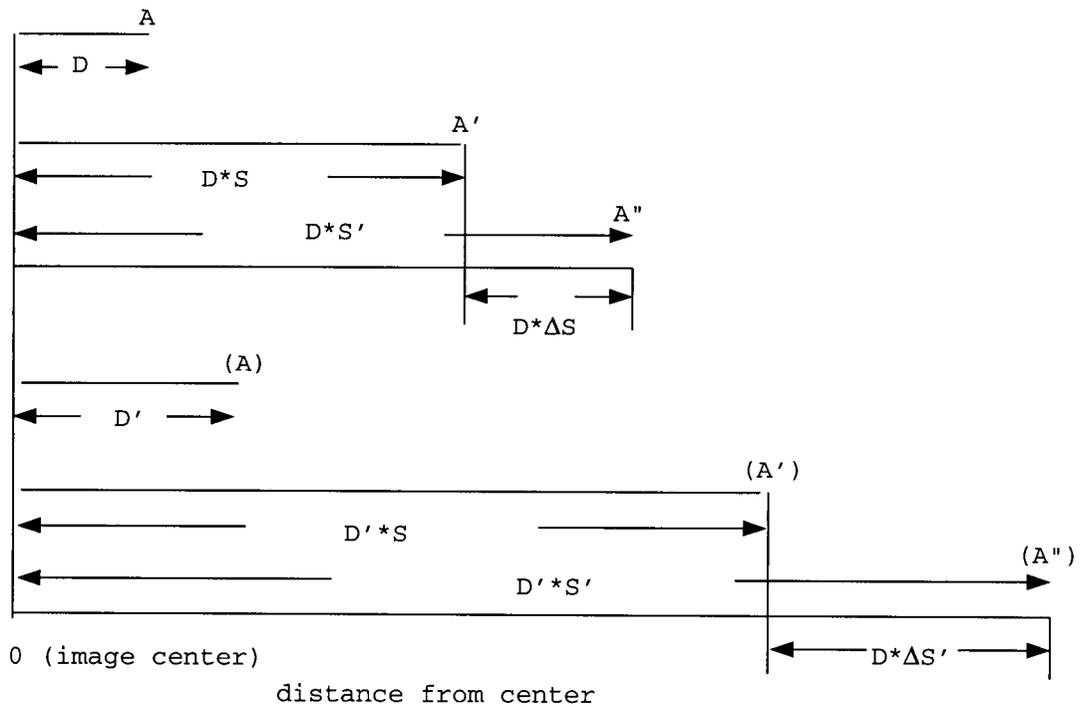
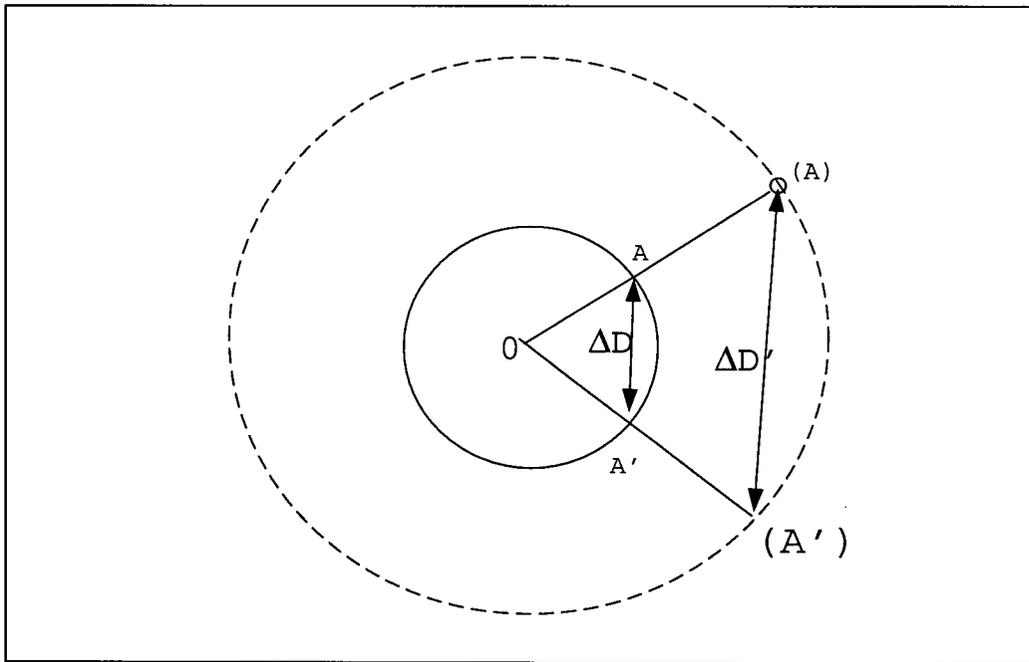
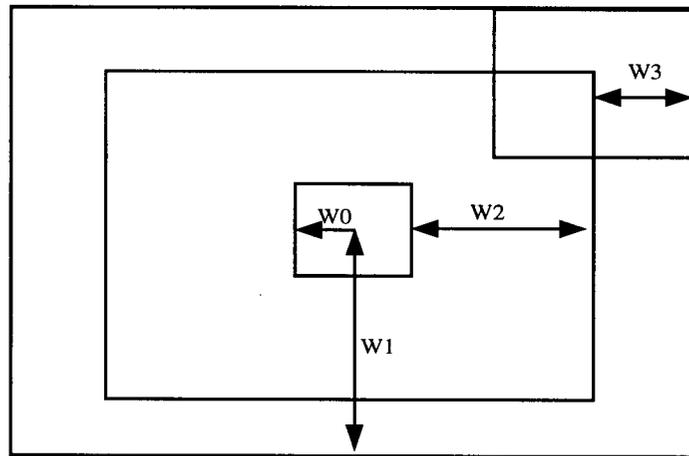


Figure 5.2 The Sensitivity to Scale vs the Corner Distance



**Figure 5.3** The Sensitivity to Rotation vs. Corner Distance

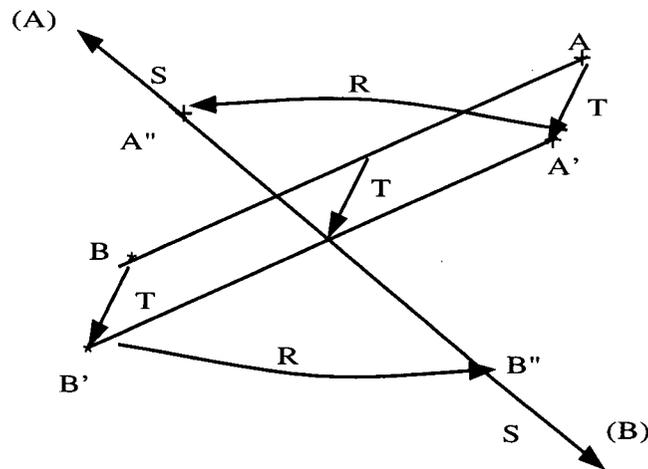
Since points have to be one pixel apart to be distinguished, i.e.,  $\Delta S * D > 1(\text{pixel})$ , so the point distance should be  $D > 1/\Delta S$  pixels to be used as a control point. This means no corner detection within the window of  $W_0 = 1/\Delta S$ . On the other hand, corners are bounded by the similarity measure window in matching. This window should be large enough to give reliable similarity measure (details in next section). The relation is shown in Fig. 5.4.



W0: The Region without Location Change      W1: The Image Size;  
W2: Corner Detection Region;      W3: Corner Matching Window

**Figure 5.4** The Location of Corners Bounded by the Windows

**Number of Points:** The minimum number of points needed to recover RST values are two corresponding points. This is shown in Fig. 5.5. Using Two corresponding points  $A \rightarrow (A)$ , and  $B \rightarrow (B)$ , we can obtain  $T$  by shifting the line  $AB$  to  $A'B'$  to align the center of  $A'B'$  to the center of line  $(A)(B)$ . The shift is the  $T$  value.  $R$  can be obtained by rotating  $A'B'$  about the center to  $A''B''$  to be parallel with  $(A)(B)$ . The rotation angle is  $R$ . Finally,  $S$  can be obtained by scaling  $A''B''$  to the same length of  $(A)(B)$ . The scaling amount is equivalent to  $S$ . The order can be different without affecting the results.



**Figure 5.5** Estimation of RST ( $T \rightarrow R \rightarrow S$ ) Using Two Control Points  $A \rightarrow A' \rightarrow A'' \rightarrow (A)$  and  $B \rightarrow B' \rightarrow B'' \rightarrow (B)$

**MSE Estimation:** Given  $n$  matching points between two images,  $(x'_i, y'_i)$  and  $(x_i, y_i)$ ,  $i=\{1,2,\dots,n\}$ , and their relations:

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = S \begin{bmatrix} \cos R & -\sin R \\ \sin R & \cos R \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (5.1)$$

To reduce the effect of inaccurate corresponding points on the estimation of parameter values, an averaging scheme like the least squares error criterion is appropriate. If  $n$  corresponding points from the two images are available, then  $R, S, T(x_0, y_0)$  can be estimated by minimizing the sum of squared errors (MSE):

$$\begin{aligned} E = & \sum_{i=1}^n \left[ x'_i - s(x_i \cos \theta - y_i \sin \theta) - x_0 \right]^2 \\ & + \sum_{i=1}^n \left[ y'_i - s(x_i \sin \theta + y_i \cos \theta) - y_0 \right]^2 \end{aligned} \quad (5.2)$$

or equivalently, we have:

$$\begin{bmatrix} x'_i & y'_i \end{bmatrix} = \begin{bmatrix} S \cos R & S \sin R & x_0 & y_0 \end{bmatrix} \begin{bmatrix} x_i & y_i \\ y_i & -x_i \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (5.3)$$

To make this problem easier, we replace  $S \cos R$  by  $a$  and  $S \sin R$  by  $b$ , we have:

$$\begin{bmatrix} \sum_i x'_i & \sum_i y'_i \end{bmatrix} = \begin{bmatrix} a & b & x_0 & y_0 \end{bmatrix} \begin{bmatrix} \sum_i x_i & \sum_i y_i \\ \sum_i y_i & -\sum_i x_i \\ n & 0 \\ 0 & n \end{bmatrix} \quad (5.4)$$

again, assume:

$$X' = \begin{bmatrix} \sum_i x'_i & \sum_i y'_i \end{bmatrix} \quad (5.5)$$

$$P = \begin{bmatrix} a & b & x_0 & y_0 \end{bmatrix} \quad (5.6)$$

$$X = \begin{bmatrix} \sum_i x_i & \sum_i y_i \\ \sum_i y_i & -\sum_i x_i \\ n & 0 \\ 0 & n \end{bmatrix} \quad (5.7)$$

then, we have,

$$X' = P * X \quad (5.8)$$

$$P = X' X^T (X X^T)^{-1} \quad (5.9)$$

by which we can find the transformation parameters as:

$$S = \sqrt{(P^2(1) + P^2(2))} \quad (5.10)$$

$$R = \arctan(P(2)/P(1)) \quad (5.11)$$

$$T = [x_0, y_0] = [P(3), P(4)] \quad (5.12)$$

The estimation precision is higher when more matching points are given and MSE is used to reduce the noise level by averaging. In this case, the variance will decrease by  $\sqrt{n}$  where  $n$  is the number of corners, and the estimation precision will increase by  $\sqrt{n}$ , given that all the matching points are at the same level of S/N ratio.

**Matching Scheme:** As mentioned earlier, there are basically two matching schemes, one is the correlation-based, another is the attributes and rigidity based. The performance of the first is studied in Chapter 4, the latter is more complicated and will be discussed in Chapter 6. In this chapter, we use the correlation window. Hence, the estimation accuracy is the same as shown in Section 4.2.

### 5.3 Reliability of the Estimation Results

**Number of Corners:** Though a larger number of points increases the estimation precision, it increases the chance of false or weak matching as well, which means lower reliability. Absolute thresholding can be used to avoid this. Experiments show the number of points should not go more than a few higher than necessary, which is two in RST estimation as shown in Fig. 5.5.

**Search range:** The search range of the corner are related to the range of RST changes, as shown in Fig. 5.6. As we have:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = S \begin{bmatrix} \cos R & -\sin R \\ \sin R & \cos R \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (5.13)$$

or

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} S * \cos R * x - S * \sin R * y + x_0 \\ S * \sin R * x + S * \cos R * y + y_0 \end{bmatrix} \quad (5.14)$$

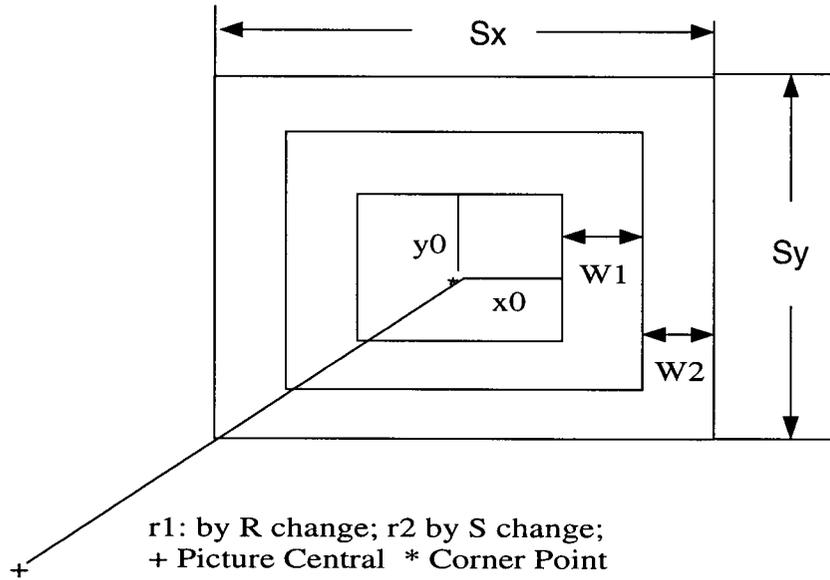


Figure 5.6 The Search Range of Corner Matching vs RST Change

Step by step, the search range by T is:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (5.15)$$

as shown in Fig. 5.6.

The search range by R is:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} x * \cos R - y * \sin R \\ y * \cos R + x * \sin R \end{bmatrix} \quad (5.16)$$

shown W1 in Fig. 5.6, note the search range is not necessarily rectangular, but in our experiment, the range is rounded to a rectangular window for simplicity.

The search range by  $\Delta S$  is:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \Delta S \begin{bmatrix} x \\ y \end{bmatrix} \quad (5.17)$$

Shown in Fig. 5.6. as region W2, again, rounded to a rectangular window.

We can see, the larger the change in RST, the larger the search window. And the larger the search range, the higher the chance of false matching, which leads to lower reliability. This has proven to be a very significant factor in our experiments.

**Matching Scheme:** Since correlation window is used for matching, the estimation reliable is the same as shown in section 4.3.

#### 5.4 Computational Cost

The cost of computation varies with the feature detector, number of points chosen, the search range and matching scheme. They are summarized in Fig. 5.7.

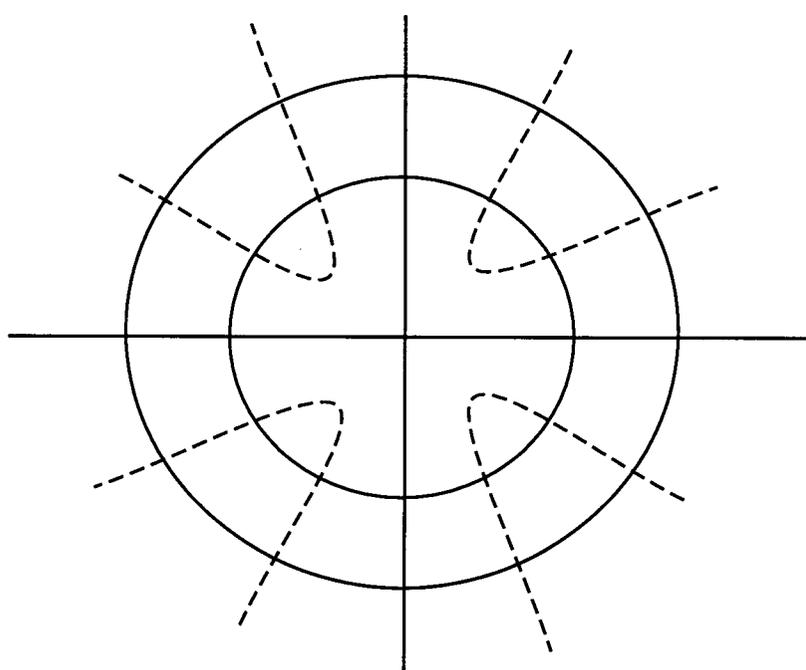
Methods (Image $L \times L$ , Correlation Window Size $W \times W$ , Search Range $T_x$ & $T_y$ , Filter Size $F \times F$ )	Mults Needed (includes divs and squares)	Typical Numbers for $L=480$ ; $W=8$ ; $T_x=10$ ; $T_y=10$ ; $F=3$ ; Corner-No=6
Plessey Corner Map	$7W^2 * T_x * T_y$	44.8k
Corner Matching with Correlation	$(3W^2 + 3) * T_x * T_y * \text{Corner-No.}$	117k

Figure 5.7 Computational Cost vs Factors

#### 5.5 Other Factors

As mentioned earlier, feature contents play the most significant role in the performance. Though it is very hard to represent feature contents, variance is usually used as a measure of feature density to avoid featureless regions. Another way of enhancing the feature contents is to apply some kind of preprocessing. Laplacian filtering is introduced to enhance the high frequency features. Similarly, a corner map can be used to enhance 2D features. The difference lies in the different frequency components being enhanced, as shown in Fig. 5.8. Laplacian filtering simply enhance all the high frequency features, including 1D features, as shown by the enclosed area of the solid lines, the inner line is the cut off frequency of the low frequency features (featureless region), while the outer line

is the cut-off of the high frequency features (smoothing out the noise). While, the corner map is to preserve only 2D features, i.e., both directional derivatives have to be above the threshold to be in the corner map, shown as enclosed area by dotted lines in Fig. 5.8. Note the graphic is an approximate representation to show the principal difference. Since we are trying to recover RS values, enhancing the 2D feature (so that sensitive to both RS equally) makes more sense. The disadvantage of this process is that the high pass filter tends to increase the noise level, and decrease the reliability of the correlation values.



Area enclosed by solid line represents the area enhanced by Laplacian Filtering

Area enclosed by dotted line represent the area enhanced by corner map

**Figure 5.8** Comparison of the Frequency Domain Area Enhanced by Laplacian & Corner Map

## 5.6 Experiments

To test the algorithms proposed, 7 set of images are taken with different zoom, different rotation, different space shift, and the following steps are applied to these images:

- 1) Plessey corner detection over AOI to obtain  $N$  corners;
- 2) Search over WFOV within  $(S_x, S_y)$  region as given in Fig. 5.4 to obtain the matching point for each corner, the corresponding correlation values are  $f_{corr}(i)$ , where  $i=1, \dots, N$ .

3) Select top  $Mf_{corr}(i)$  values as the control points to recover RST values using the transformation matrix and MSE as discussed in Section 3.1 and 5.2.

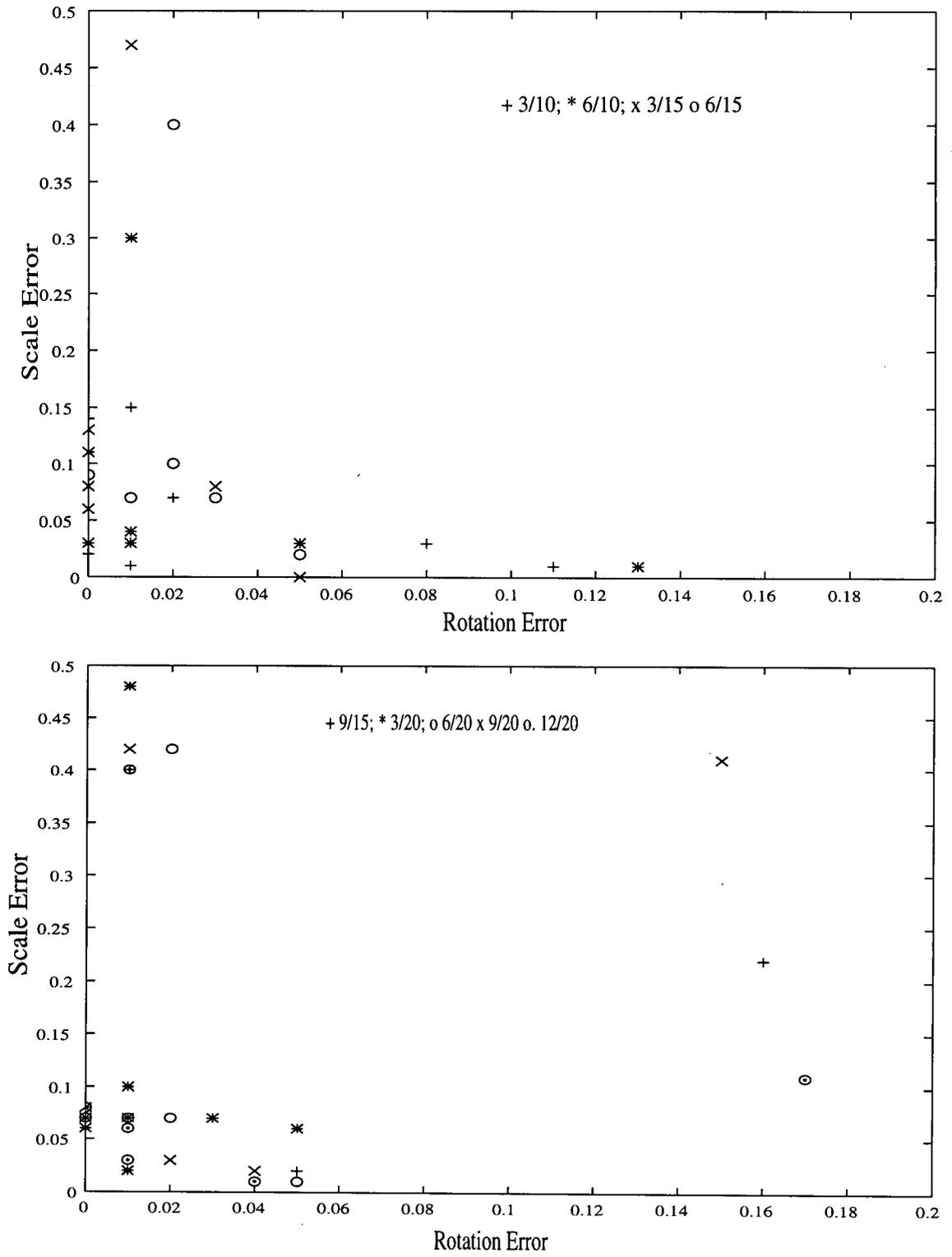


Figure 5.9 The Effects of Corner Numbers on the Estimation Error

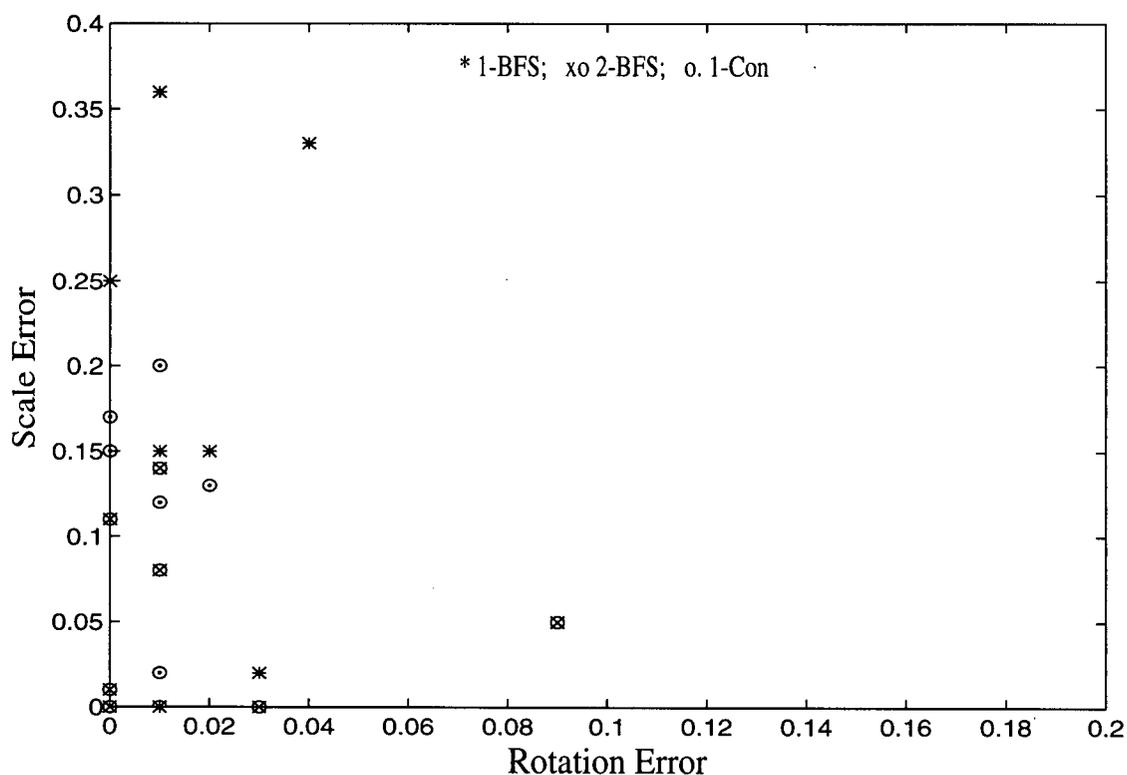


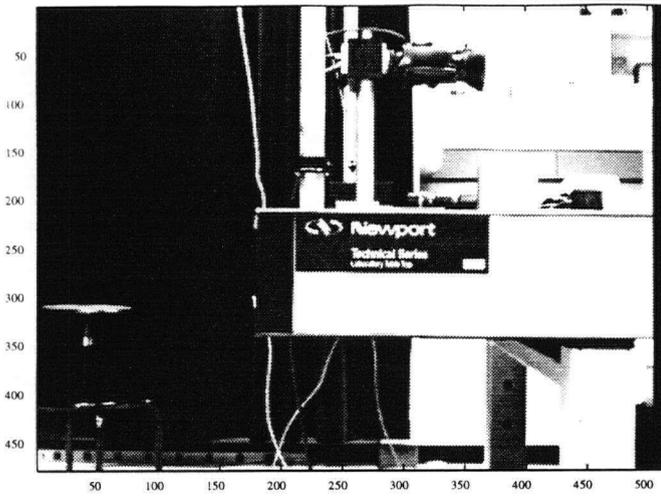
Figure 5.10 Comparison of the Estimation Error of Different Methods

Fig. 5.9 shows the effect of corner number on the estimation error. Again, the outliers are those located further away from the origin and represents points with higher estimation errors and are not desirable. The number given in Fig. 5.9 represent  $M/N$  as discussed earlier. The larger number is the number of original corners detected from the image  $N$ , the smaller one is the number of matching pairs selected from the original corners based on the top correlation value,  $M$ . The figure shows that corner numbers between 3 and 6 give good estimation. A higher number doesn't improve and might even degrade the results. For example, 12/20 estimation does not give better accuracy than 3/20 overall. The reason for this is some images have with very limited number of significant corners, and if large number of corners is required, the threshold has to be lowered to the level where weak corners or noise are detected and matched as control points.

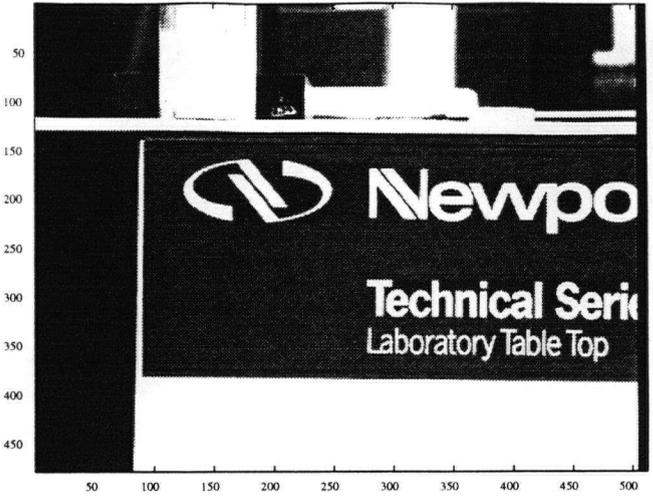
Fig. 5.10 shows the estimation error after one level brute force search, two level brute force search and one level brute force search followed by corner-based search. Note, the second level brute force search gives the same level of estimation error as corner-based matching, but corner

matching is computationally more efficient. But there are cases where corner-based estimation will fail completely, this is unavoidable due to a lack of feature variety in some images. Multi-level brute force search seems more reliable in this case.

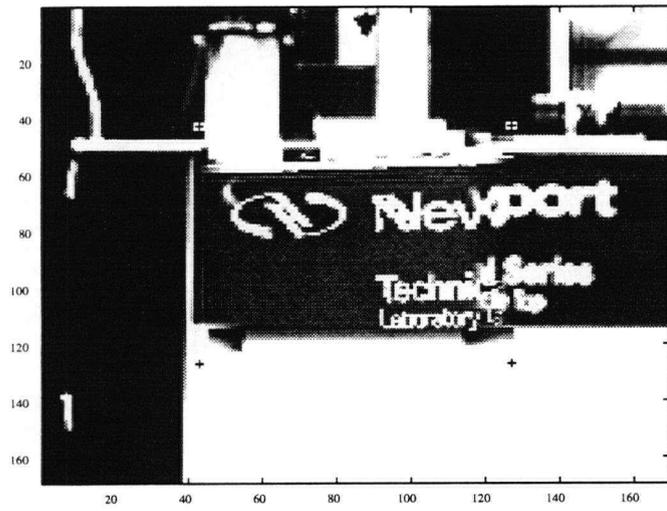
The final fused image is shown in Fig. 5.11. Note that only part of the fused image are given to show more clearly the transition band between the high resolution and low resolution transition. The central part marked by cross is the high resolution part.



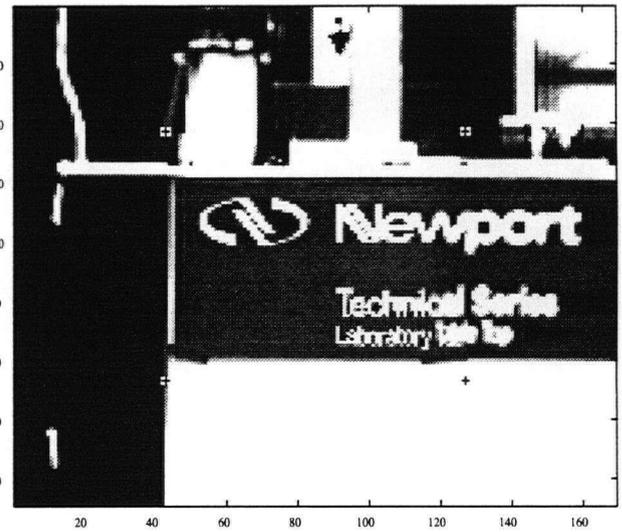
Background



Insert



1BFS



2BFS

Corner Match

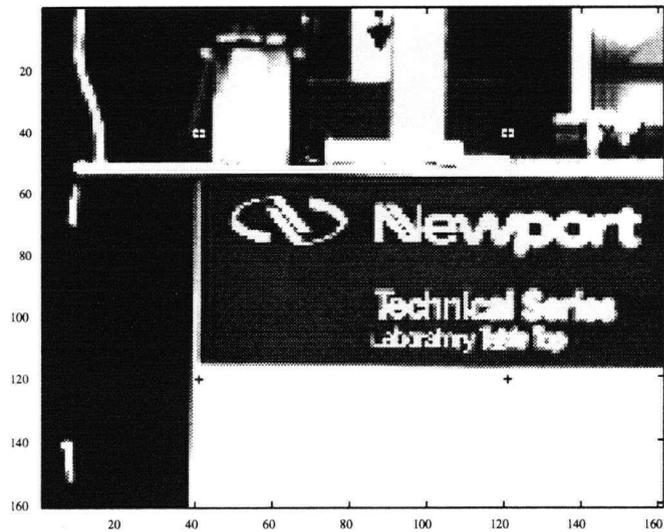


Figure 5.11 The Fused Images After Each Process

## Chapter 6 RST Invariant Corner Detection

In the previous chapter, we determined that corners are used in fine RST estimation. The question is why not use corners to recover any RST values?

Corner matching using correlation without any knowledge of RST values requires a large range of search space, i.e., must correlate over large RST space for each corner. This is more computationally expensive than a brute force search, and can have few benefits.

There are other ways of corner matching, such as using corner attribute and geometric information. But not all of them are valid depending on the changes involved, like geometric information is hard to apply in matching where both R and S changes are involved. Only those attributes that are invariant to the changes at hand can be used as consistent features to be matched over pictures.

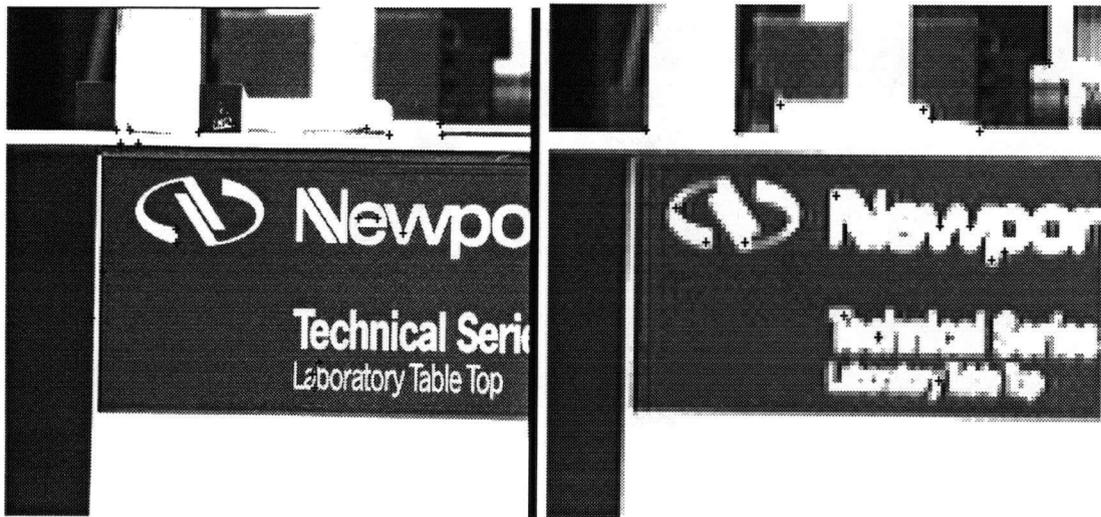
No matter what kind of matching scheme is used, the success of matching results highly depends on the corner set in two pictures. This is even more important when scale change is involved, since scale difference can significantly change the corner set detected, and making it even more difficult to have a reliable matching results. This chapter will study this phenomena, starting from the study of corner behavior over scale and then detection of the scale consistent corners. This is one step further towards the approach of using corners to recover total RST changes, which proved to be very difficult.

### 6.1 Introduction

As mentioned earlier, significant scale changes will result in significant differences in the corner map, which means the corner sets in two pictures with significant scale change is very different, making the corner matching very difficult. So, first, we need to study how the corners behave over scale change so that only scale consistent corners are selected to give more corresponding points between two pictures with a scale difference.

Ideally, a corner detector should be designed to produce a corner map proportional to their consistency. Unfortunately, almost all corner detectors fail to give consistent results over changes in smoothing scale, which makes the use of corners as control points unreliable for multi-resolution applications. To solve this, either adaptive smoothing or corner detection at multiple scales are

applied. As will be discussed in this chapter, we have reasons to believe that scale-invariant corners can be identified based on one fine scale corner map. A new concept of significant values associated with a corner's detectability and displacement over scale is introduced. Analytical results are obtained on smoothing effects derived from a cylindrical tube approximation. These results are shown to be representative for general smoothing functions as well. Experiments are carried out to prove that the significance value can be used to predict the corner behavior over space scale, and to identify the scale-invariant corners based only on one fine scale corner map. The advantage of this approach is better localization, reliability, and computational efficiency.



**Figure 6.1** Comparison of The Corner Sets for Different Resolution Images

As shown in Fig. 6.1, the two images are taken with different zooms,  $f=12.5$  and  $f=50$ , so the resolution ratio is 4:1. The points in the picture are the corners detected using Plessey corner detector [Harris88], threshold for strongest 20 corner responses. We can see from the picture that the corner maps are very different with only 7 matching pairs out of 20 which makes 35% corresponding points. If we apply any matching algorithm on this pair of corner maps, the chance of mismatch is high. The reason for this inconsistent result in most corner detectors is that their design is only based on two corner attributes, the amplitude of the corner edge, and the curvature of the corner edge. But, a corner has more than just two attributes and these attributes are strongly correlated. Neglecting this fact may contribute strongly to the inconsistency of the results. A solution to this

problem without going through multi-resolution or adaptive filtering is examined in this chapter. A significance value based corner selection procedure is introduced. Experiments are carried out to show the improvement over traditional methods.

## 6.2 Modeling a Corner

There are mainly two ways to approach multi-scale corner detection: 1. Apply multi-scale edge/corner detectors and interpret the results [Lu&Jain, 89], etc.; 2. Apply adaptive scale smoothing to obtain a multi-scale effect on the edge/corner detection [Saint-Marc&Chen, 91], etc. The problem with the first is that the matching procedure among the different scale edge/corner map is non-trivial and might introduce some errors along the way. The second approach needs a reliable way to determine the right scale for smoothing which is a difficult issue in itself. Nonetheless, the above research has led to some interesting findings on the scale-space behavior of the edge/corners, and these are part of the basis of our new approach. So, we will briefly summarize these results before we present the approach. Many of these studies are based on 1-D features such as edges, so there will be a bias in using these results to study the corner behavior, we will introduce only those that are relevant to corner behavior.

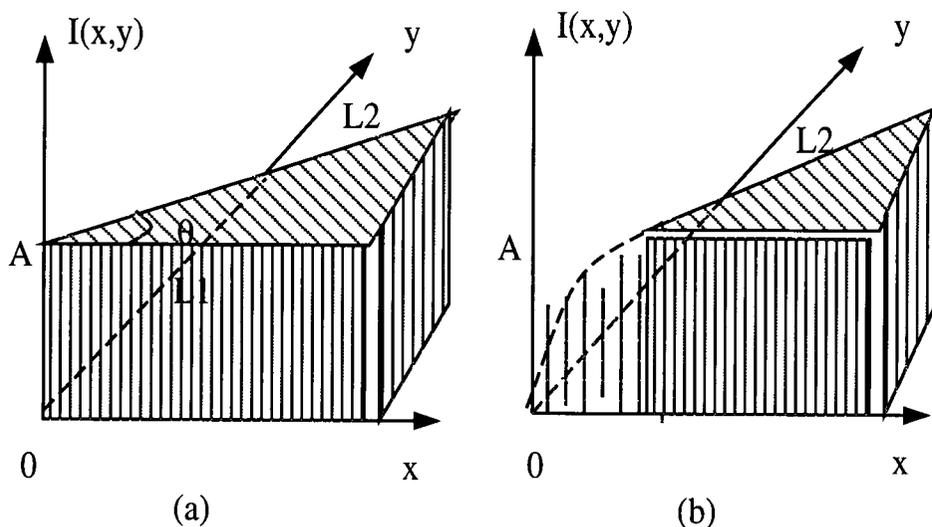


Figure 6.2 Corner Model

In [Antonio Guiducci, 88], a corner is represented with the amplitude  $A$  of the edge, its aperture angle  $\theta$  and a parameter  $\sigma$  which is a measure of the smoothness of the corner (A non-ideal corner with radius  $\sigma$  can be approximately represented by an ideal corner, i.e., an infinitely sharp corner, smoothed by a Gaussian with variance  $\sigma$ ), as shown in Fig. 6.2. The definition of the attributes are as follows:

**Corner Arm Orientation:** the direction of the lines forming the corner;

**Corner Angle:** the angle between the two corner arms, with the high intensity level part designated as the corner area;

**Corner Depth:** the gray level difference between the corner area and the rest area;

**Corner Surface Smoothness:** The Variance in the corner area;

**Corner Arm Length:** the length of the corner arms within a small direction change.

For more complicated corners, their representations need more parameters and will not be covered in this paper. All the following studies are based on the above simple corner model.

### 6.3 Smoothing Effects and Definition of Significance Value

#### Review of Multi-scale Corner Detection and Corner Behavior

As studied in [Anothai Rattarangsi and Roland T. Chin, 92], a) depending on the location and sharpness of the corners, and their interaction with adjacent corners, the line patterns of the scale space map may merge, disappear, repel, or attract each other as  $\sigma$  increases; b) new additional line patterns will not be introduced by smoothing; c) different corner models have different scale space characteristics, basically, some persist, some merge and some will disappear over scales, also some will be dislocated over the scale; d) isolated corners are persistent, and non-isolated corners are mainly affected by the immediate neighboring features. D-isolated corners will be detected by Gaussian smoothing function with a variance of  $\sigma < D$ .

In [Fredrik Bergholm, 87], the study shows (a) a corner with a step edge forming an angle  $\beta$  is rounded off with a displacement vector  $\bar{v}$  along the bisector satisfying:  $\bar{v} = C(\beta) * \sigma * \bar{e}$ ;  $C(\beta) < 2$  if  $\beta > \pi/8$ ,  $C(\beta)$  is a constant depending on  $\beta$ , and  $\bar{e}$  is the unit vector along the bisector. (b) a double edge profile has two edge points that start to glide apart when the scale parameter  $\sigma$  is greater

than half the width ( $h/2$ ) of the double edge. When blurring is further increased, i.e.,  $\sigma > h/2$ , the speed with which they move apart approaches  $\Delta\sigma$ .

[Yi Lu, Ramesh C. Jain, 89], a) false edges disappear as scale parameter changes from large to fine scale. b) only false edges can disappear from large to fine scales. c) edges missing at larger scales can always be recovered at smaller scales.

The study in [Valdis Berzins, 84] shows that: 1) Marr and Hildreth edge detection locates infinite straight edges with linear illuminations exactly; 2) the contour found by the edge detector goes through the corner point exactly for corners with infinite arm length, the effect of the edge detector is to round out the corner without displacing it at the corner point; 3) the displacement of the contours is due entirely to errors introduced by approximating the directional derivatives by the Laplacian, and not to the Gaussian filtering for corners with infinite arm length; 4) the displacements along the corner edge are largest for small angles (acute corners), and that the smaller the angle, the further the displacement extends from the corner; 5) for finite arm length edges, the zero-crossing contour no longer passes through the true corners. The displacements near corners are dominated by errors in the Laplacian approximation to the directional derivative, while displacements around small objects are dominated by the effects of the Gaussian filter.

Other studies more or less fit into these results. In summary, we can conclude:

- Corners may merge, disappear, repel, or attract each other as  $\sigma$  increases;
- New additional line patterns will not be introduced by smoothing;
- D-isolated corners will be detected by Gaussian smoothing function with a variance of  $\sigma < D$ ;
- Isolated corners are persistent, and non-isolated corners are mainly affected by the immediate neighboring features;
- Marr and Hildreth edge detection locates infinite straight edges or corner points with infinite arm length exactly;
- The displacements along the corner edge and the range of displacement along the corner edge are largest for small angles (acute corners);

- For finite corner edges, the zero-crossing contour no longer passes through the true corners;
- False edges disappear as scale parameter changes from large to fine scale.

From these studies, we can conclude that the fine scale features includes all the features that can be found at coarser scale (given the same cornerness threshold), and gives better location. How these features change over the scale are related to the corner attributes and the corner adjacency.

### The Smoothing Effects on an Isolated Corners

#### The Smoothing Function

An ideal corner with one edge along the x axis and an angle  $\theta$  can be modeled by the following 2-D step function

$$I(x, y) = AU(y)U(x \tan \theta - y) \quad (6.1)$$

where

$$U(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6.2)$$

and A is the amplitude of the corner edge depth.

A 2-D Gaussian filter  $G(x, y)$  can be represented as follows,

$$G(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x^2 + y^2)/2\sigma^2) \quad (6.3)$$

where  $\sigma$  is referred to as the scale.

If we convolve this 2-D Gaussian function with the ideal corner model, we get the following smoothed image  $S(x,y)$ :

$$S(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(p, q) * I(x - p, y - q) dpdq \quad (6.4)$$

Only numerical solutions can be obtained, and it is hard to derive the relation as to how the corner attributes changes over the smoothing scale, especially how these attributes affect each other

in this change. To simplify the analysis, we use a pillbox to approximate the Gaussian function,  $G_c$ , as follows,

$$G_c = \begin{cases} 1 & x^2 + y^2 < \sigma \\ 0 & \text{otherwise} \end{cases} \quad (6.5)$$

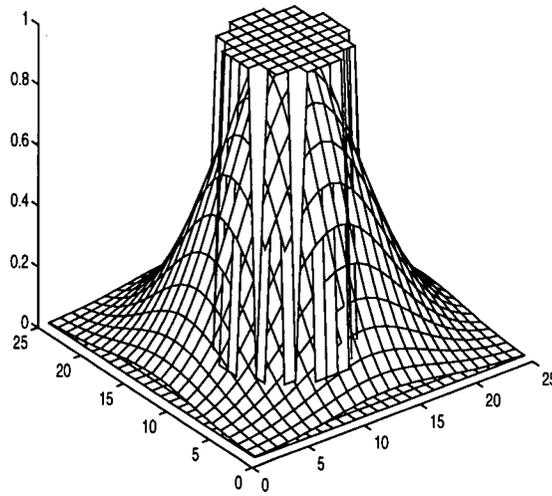


Figure 6.3 Gaussian and Cylindrical tube as Smoothing Function

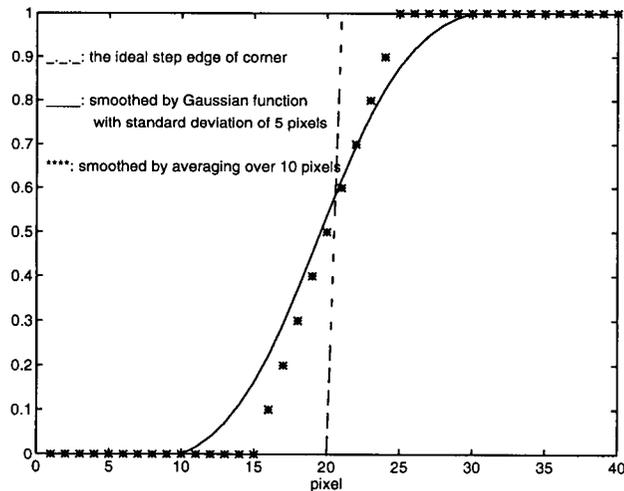
The advantage is with this simplification, we are able to derive the analytical results as follow:

$$S(x, y) = \left\{ \begin{array}{l} A \\ ifU\left(x - \frac{\sigma}{\tan\theta/2}\right) * U(y - \sigma) * \\ U\left(\left(\left(x - \frac{\sigma}{\tan\theta/2}\right) \tan\theta - \frac{\sigma}{\tan\theta/2}\right) - y\right) \\ 0 \\ ifU(r - \sigma) * U(\phi - 90 - \theta) * U(270 - \phi) \\ 0 < S < A \end{array} \right\} \quad \begin{array}{l} \\ \\ \\ \\ \text{otherwise} \end{array} \quad (6.6)$$

It is hard to say how good this approximation is, except that this approximation can be representative for all the smoothing effect in general.

**The Smoothing Effects:**

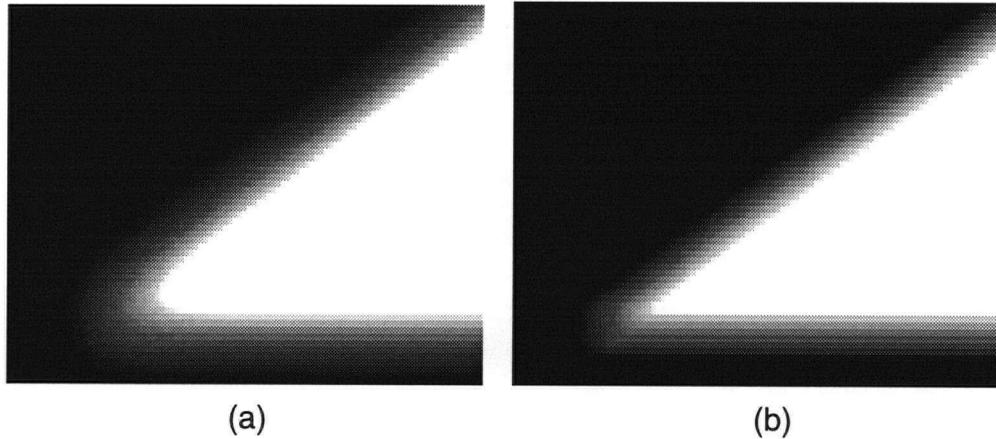
The key to understand the scale-space behavior of the corners is in the smoothing function. To understand what a smoothing process does to a corner, we use the simplest version — a cylindrical tube — to derive the analytical results. A popular smoothing function is the Gaussian function. The problem with using a Gaussian function is that numerical results can be obtained for the displacement value only over the smoothing scale, but it is hard to obtain the relation to the corner attributes. Note that the energy of Gaussian function lies heavily in the range of  $[-\sigma, \sigma]$ , where  $\sigma$  is referred to as the scale. To simplify the analysis, we use a cylindrical tube with radius  $\sigma$  as our smoothing function. Fig. 6.3 gives the comparison of the cylindrical tube and the Gaussian smoothing function. It is a reasonable approximation and it can be shown the results obtained are good representation for other general smoothing function as well.



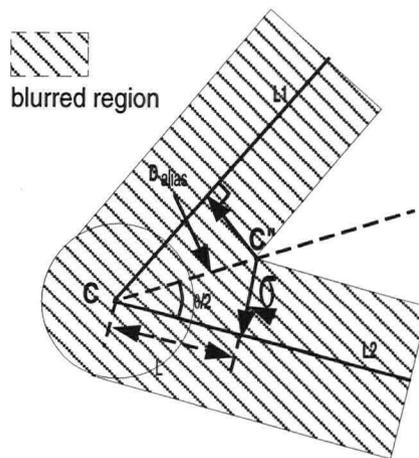
**Figure 6.4** The Approximation of The Gaussian Smoothing Function

Smoothing can be regarded as low pass filtering in the Fourier domain, while in the space domain, it can be taken as a weighted sum of neighboring pixel intensities. It is hard to represent corner attributes in the Fourier domain since after being transformed to Fourier domain, all the spatial features are mixed up, and it is not easy to relate the Fourier representation to the corner characteristics. In the space domain, it is relatively easy to relate the weighted sum function with the corner attributes. The effect of the smoothing function on each corner attribute is analyzed.

Take an ideal corner (i.e., an indefinitely sharp wedge as defined earlier), and see what happens when we apply the smoothing function with scale  $\sigma$ , Fig. 6.5 gives two smoothing scale effect, and we can see the corner sharpness changes to different extent accordingly.



**Figure 6.5** The Matlab Simulation Results for Smoothing Effects (a)  $\sigma=4$ ; (b)  $\sigma=2$



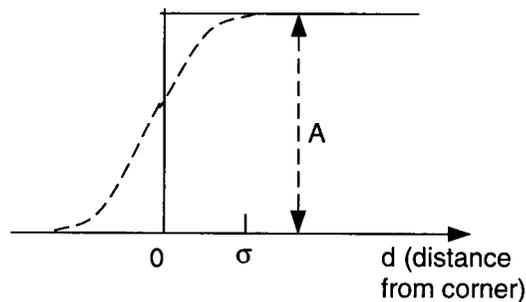
**Figure 6.6** The Top View of the Smoothing Effects on the Corner

**Corner Sharpness:** An ideal corner has infinite curvature or zero radius. After applying a smoothing function with scale  $\sigma$ , the radius is approximately equal to the smoothing scale  $\sigma$ , or say the curvature is equal to  $1/\sigma$ , shown in Fig. 6.6. If the original corner is not ideal, i.e., with non-zero radius of  $\sigma_0$ , then after smoothing, the overall radius will be  $\sigma_0 + \sigma$ , since the smoothing

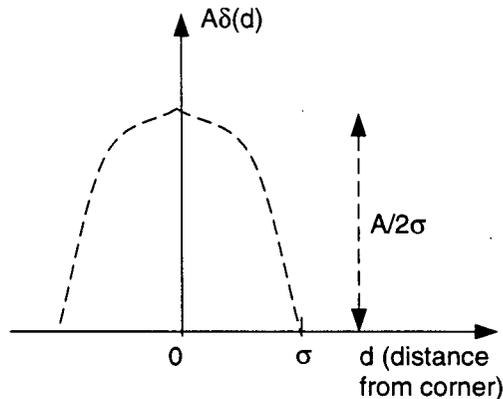
is a linear process. Note, that for Gaussian smoothing, this will be  $\sqrt{\sigma_0^2 + \sigma^2}$ . The difference shows that there will be some bias in the representation.

**Corner Point Displacement:** displacement is defined as the distance between true corner point C and the detected corner point C', as shown in Fig. 6.6. The maximum displacement can be derived as  $D_{alias} = \sigma / \sin(\theta/2)$ . This is consistent with the round off effect analyzed in [Fredrik Bergholm, 87] using a Gaussian smoothing function, where the numerical result shows that the displacement reaches the minimum at 180 degrees and maximum at 0 degrees.

**Step Edge Amplitude:** Edge depth is usually measured by the gradient, and smoothing gives decreased gradient value. As shown in Fig. 6.7 & 6.8, the gradient is infinite for the step edge, and  $A/2\sigma$  for the smoothed one. Fortunately, all the edges are smoothed by the same scale, the gradient still represents the relative amplitude, though the threshold chosen should be smaller for a smoothed picture accordingly.



**Figure 6.7** The Step Edge Before (Solid Line) /After (Dotted Line) Smoothing



**Figure 6.8** The Gradient of Step Edge Before (solid line) /After (dotted line) Smoothing

**Arm length L1 & L2:** As shown in Fig. 6.7, for the corner to be detected after smoothing, the arm length has to be longer than  $L = \sigma / \tan(\theta/2)$ , i.e., the arm length has to extend beyond the blurred region. On the other hand, the arm length also needs to be longer than the smoothing scale to be detected. As a result, when  $\min(L1, L2) > \max\{\sigma / \sin(\theta/2), \sigma / \tan(\theta/2)\}$ , there exists a corner; otherwise no corner.

In the following section, we will combine all the factors together, with the correlation among them being considered as well.

### The Detectability and The Displacement of an Isolated Corner Over Smoothing Scale

In the literatures reviewed in section 6.2, the studies are carried out from one perspective or the other, such as angle, corner edge, or isolation. In this paper, we will introduce a measure to consider all the factors and their correlation, define a significance value associated with the persistency of the corner over smoothing scale. This value will represent the level or extent to which scale the corners will exist. Also, knowing that a definite corner or ideal corner does not exist in a real image, it is the cornerness or the likelihood that marks a point with its probability as a corner candidate. When a point is detected as a corner point, it makes more sense to say that this point has a certain likelihood or certain significance level as a corner candidate.

From the study of the smoothing effect, we are able to relate the detectability of the corners with their attribute in the following relation (the function is normalized to equal to zero at the turning point of detection). In the equation, we choose to use the exponential to represent the significant

contribution of the length of the corner arm. As a result, the significance value associated with the detectability is formed as:

$$SD = (G_{edge} - G_{threshold}) * \left( \exp\left(\frac{\min(L1, L2)}{L}\right)^{-1} - 1 \right) \quad (6.7)$$

or,

$$SD = (G_{edge} - G_{threshold}) * \left( \exp\left(\frac{\min(L1, L2)}{\max\{\sigma, (\sigma / \tan(\theta/2))\}}\right)^{-1} - 1 \right), \quad (6.8)$$

where G stands for the gradient value at the corner, L is the minimum length requirement as shown in Fig. 6.7. Here, the significance of detectability SD is normalized so that SD positive represents the detectable corners at scale  $\sigma$  with the value proportional to the confidence level; SD negative represents the blurred or undetectable corners at smoothing scale  $\sigma$ ; S=0 represent the turning point.

Detecting the existence of corners and determining their exact location are two different but related goals. So, we need to distinguish the factors that affect the detectability and the factors that affect the localization. As discussed in section 3, the maximum displacement of an isolated corner point is

$$D_{alias} = \sigma / \sin(\theta/2).$$

### The Detectability and The Displacement of a DS-Isolated Corner Over Smoothing Scale

Before we can complete our significance function, one more thing to consider is the adjacency which is another factor that has been found to affect the behavior of the corners over scale. As pointed in [Lu & Jain, 89], we can assume that the scale space behavior of a corner is mainly influenced by its immediate neighbors and that the effects from other features further away are negligible. To address this issue more properly, we have redefined the concept of D-isolation. In [Lu & Jain, 89], an edge curve is called D-isolated, if for any point on the curve, we can draw a disk centered at that edge point with radius D, and that disk will intersect no edges except those on this edge curve, D-isolated corners are usually regarded as the distance to the nearest edge/corner. In contrast to this, we use a term “DS-isolation” defined as the distance to the nearest edge or corner of

the same or higher level of significance value. This definition is chosen based on the fact that both the distance of the neighboring feature and the strength of the feature compared to the corner under concern contributes to the difference in the corner behavior over scale. For example, if a corner is adjacent to another one which is significantly weaker than the corner of interest, then the corner of interest will remain virtually unaffected by this neighboring corner. So when defining the isolation, one must take the strength of the neighboring features into account to make the concept more useful. Also, in contrast to the traditional definition of the strength of the corner which is usually defined as the edge amplitude or cornerness, we use the significance of detectability value  $SD$  as the measure of the corner strength. The edge amplitude multiplied by the edge length is used as the edge strength. Knowing that the scale behavior of corners is strongly affected by neighboring features, we choose the exponential to represent the contribution of the DS-isolation value, with the smoothing factor as a reference which matches the study in [Lu&Jain, 89], where it is proved that D-isolated points (defined in a slightly different way) can be detected by an operator with a smoothing scale less than  $D$ .

The combined significance value associated with the detectability of a DS-isolated corner can be represented by:

$$SD = \left( \exp\left(\frac{DS_{iso}}{\sigma} - 1\right) - 1 \right) * (G_{edge} - G_{threshold}) * \left( \exp\left(\frac{\min(L1, L2)}{\max\{\sigma, (\sigma / \tan(\theta/2))\}} - 1\right) - 1 \right), \quad (6.9)$$

where  $SD$  positive represents the detectable corners at scale  $\sigma$  with the value proportional to the confidence level;  $SD$  negative represents the blurred or undetectable corners at smoothing scale  $\sigma$ ;  $D=0$  represents the turning point.

Correspondingly, the combined displacement can be represented as

$$D_{alias} = \sigma / \sin(\theta/2) + \frac{SD_0}{SD_n} * \left( \sigma - \frac{DS_{iso}}{2} \right) * u\left( \sigma - \frac{DS_{iso}}{2} \right), \quad (6.10)$$

where  $SD_0$  and  $SD_n$  represent for the detectability value of the corner of interest and the nearest neighbor corner/edge (assuming they are isolated). The first term originates from the decrease of the corner sharpness. To represent the worst case, the maximum displacement is used as the corner displacement value. The second term originates from the merging or repelling trend of two

neighboring corners, as pointed out in [Fredrik Bergholm, 87], though the term is generalized by taking the strength of the corner into consideration. Theoretically, the sum should be a sum of vectors, since two displacements might be in different directions, here we simplify the problem by the worst case where both are in the same direction. So the worst case result becomes the scale sum.

In the following section, we will carry out some experiments to see how well our equations represent reality.

## Experiments

### In Predicting the Corner Scale Behavior

In this section, we will give some tests to prove that the significance value can be used to predict the scale behavior of the corners. To compare with the results obtained by traditional methods, such as multi-scale corner detection, we choose the test images used in [Anothai Rattarangi & Roland T. Chin, 92], where corners detected in different scales are represented in a sparse tree. In our experiments, the corners are marked as a, b, c,..., shown in Fig. 6.9, ranked in order of their detectability value, and compared to the corresponding sparse tree in [Anothai Rattarangi & Roland T. Chin, 92].

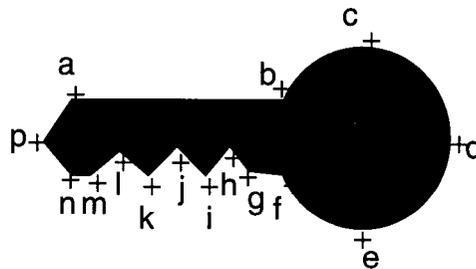


Figure 6.9 Test Image

In ranking the corners, both the detectability value and the displacement value, as well as the cornerness are considered. Cornerness is defined as the curvature, which can be obtained by

$$C = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{(I_x^2 + I_y^2)^{3/2}} = 1/|\nabla I|(k_{n\perp}) \quad (6.11)$$

Note that there might be a confusion as whether the corners are displaced or disappeared when two neighboring corners are in the same level of detectability, such as m and n. In this situation, the

stronger corner is recognized as the persistent one, in our case, corner n is said to be persistent rather than corner m. The corners are ranked, in the order of significance value, as b, a, f, c, d, e, h, i, j, k, l, p, g, n, m. We can tell this order is almost consistent with the corresponding scale space tree given in [Anothai Rattarangsi & Roland T. Chin, 92], with the exception of corner a, which is failed to be detected at coarse scale in the tree representation. But intuitively, we see no reason that corner a is not detected while corners i, j, k, ... are detected. This is due to the weakness of the corner detector applied in [Anothai Rattarangsi & Roland T. Chin, 92], and suggests that a new corner detector is required for multi-scale applications which can provide more consistent results. This issue is addressed in details in the following sections where multi-resolution image registration is studied.

### Selecting Scale Invariant Corners

For the multi-resolution image registration, ideally, the corner detector should be designed to be proportional to the significance value associated with the detectability, i.e.,  $SD = \left( \exp\left(\frac{DS_{iso}}{\sigma} - 1\right) - 1 \right) * (G_{edge} - G_{threshold}) * \left( \exp\left(\frac{\min(L1, L2)}{\max\{\sigma, (\sigma/\tan(\theta/2))\}}\right) - 1 \right)$ , and inversely proportional to the displacement value, which is  $S_{alias} = \sigma / \sin(\theta/2) + \frac{SD_0}{SD_n} * \left( \sigma - \frac{DS_{iso}}{2} \right) * u\left(\sigma - \frac{DS_{iso}}{2}\right)$ , where L's are the corner edge length,  $\theta$  is the corner aperture, G is the corner amplitude,  $\sigma$  is the smoothing scale, DS is the D-significantly-isolated as defined in section 4 & 5. In addition, cornerness is also considered as a proportional factor, which is defined as the rate of change of gradient direction along an edge, multiplied by the gradient magnitude,

$$C = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{(I_x^2 + I_y^2)^{3/2}} = 1/|\nabla I|(k_{n\perp}) \quad (6.12)$$

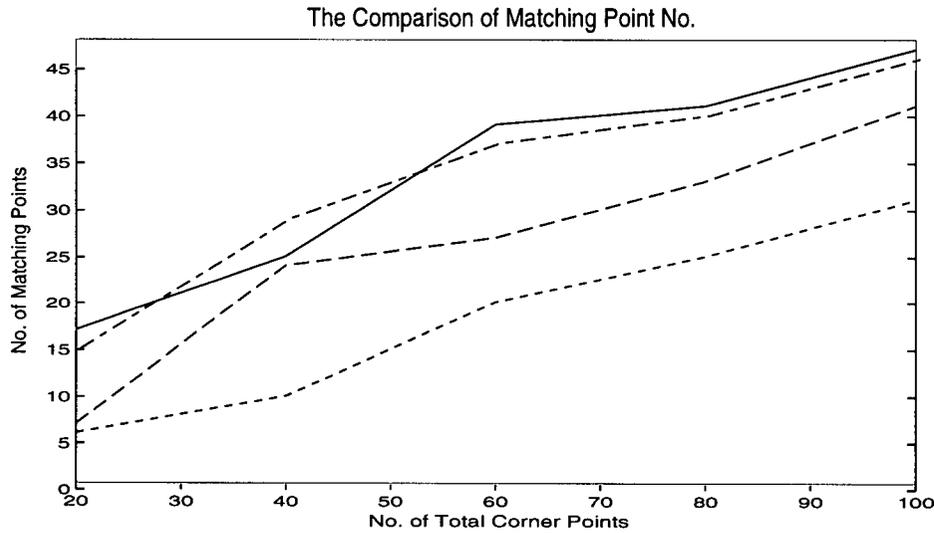
To be proportional to cornerness and detectability and inverse proportional to the displacement, in corner attributes, implies being proportional to corner amplitude (or depth), corner arm length L1 & L2, and inversely proportional to the smoothing scale  $\sigma$ . As to the aperture, it should be proportional to  $\cos(\theta/2)$  due to cornerness,  $\sin(\theta/2)$  due to detectability, and  $\tan(\theta/2)$  due to displacement. This results in a conflict situation. Since from the cornerness point of view, which has been used as a criterion so far,  $\cos\theta/2$  should be used (the smaller the corner, the higher the curvature), while from the persistent/displacement point of view, the  $\sin\theta/2$  should be used in the criterion (the smaller the angle, the larger the displacement, and less detectable as well). This conflict in detecting corners is one main reason behind the inconsistent result of the traditional corner detection where only

cornerness is used as the criterion. To compromise among these factors, we propose to use  $\sin\theta$  as the criterion in designing a corner detector.

From the above analysis, we can see that the key to consistent corner detection is to form a measure that considers all these factors and is less expensive computationally. This can be used to evaluate the corner detectors or as a guideline in designing new corner detector.

An alternative will be to apply one corner detector to obtain corner points, estimate the corner attributes and evaluate the corners' significance value. To calculate the significance values, we start with corner arms orientations. They are obtained using the gradient orientation histogram as introduced in [Rosin, 96]. Then the angle is obtained as the difference between the two corner arms' orientations with high gray level area designated as corner area, and the depth is obtained as the mean difference between corner area and non-corner area in the window centered at the corner. The surface smoothness is added to the significance value to represent the noise level of the corner, so that corners in the noisy area is discarded as well. The surface smoothness is defined as the variance of the corner area. On the other hand, the corner arm lengths are not used since it is complicated to calculate, the same with the D-isolation value. In computer graphics where the geometry of objects are known, and view point is changing with zooming, these attributes and significance values can be obtained and used as scale consistency measure.

We first applied the selection process on Fig. 6.1. There is a general improvement in the percentage of matching points, as shown in Fig. 6.10. As we can see, more improvement are obtained in applying the selection process to image after smoothing. The result can be close by applying the right amount of smoothing scale which is not always known.



**Figure 6.10** Solid: partial smoothing+selection; Dashdot: full smoothing; Dashed: Selection; Dot: no selection.

## 6.4 Summary

In this chapter, we addressed the inconsistency problem in the corner detection, and studied the possibility of solving this problem. The concept of significance value of corners is introduced, as well as the DS-isolation and detectability. Analytical formulas to calculate these values are obtained using a simplified smoothing function, which proved to be representative for general smoothing function as well. An implementation measure is proposed and experimental results are given to show that some improvements are obtained in increasing the percentage of matching points for multi-resolution registration.

The advantages of this approach: 1) only one scale of corner detection is required; 2) corners are extracted on the finest scale, which gives least localization error; 3) the spurious corners and false corners detected on the fine scale can be filtered out by thresholding the significance value, since very small significance values are assigned to these corners by our definition.

By no means have we covered all the factors and aspects in detecting the scale consistent corners, and the reliability and feasibility of our approach still needs to be studied. Problems remaining: (1) the significance values are derived based on the linear approximations; (2) significance value equations are only proportional representation. An elaborate form can be obtained by more experiments; (3)

the adjacency problem is far more complicated than a simple equation can represent. And in our experiments, this is not addressed due the complexity of the computation.

There are many issues that need further study in this area, this chapter studied the feasibility of RST estimation using corner matching directly without any brute force search, and found there are two issues need to be studied beforehand. One is to increase the percentage of corresponding corners across two resolution images, another one is the reliability of matching schemes. We analyzed the source of the first problem, and proposed a measure to improve the results. The latter issue is still an open issue and needs further study. Consequently, this approach was not adopted in our final approach to recover RST values.

## Chapter 7 Image Fusion and Proposed Approach

Once the images are registered, the next step is to fuse them into one image. In hardware, they can be projected onto one screen. In this chapter, a software fusion is discussed and implemented. This involves recovering possible relative distortion, re-sampling, normalization and transition band smoothing. Details are given in the following sections.

Also in this chapter is a complete software implementation flowchart as a summary of the previous analysis and experiments. It adopts a multi-process approach where the optimal process is used based on the given initial values of the RST change range.

### 7.1 Radial Distortion Recovery

Pictures taken with cameras of different RST are subject to lens distortion to different extents, which means there is relative distortion between two pictures.

A simple radial distortion is the first order model represented as follow:

$$\begin{aligned}x' &= x k_1 r^2 \\y' &= y k_1 r^2 \\r^2 &= (x - x_c)^2 + (y - y_c)^2\end{aligned}\tag{7.1}$$

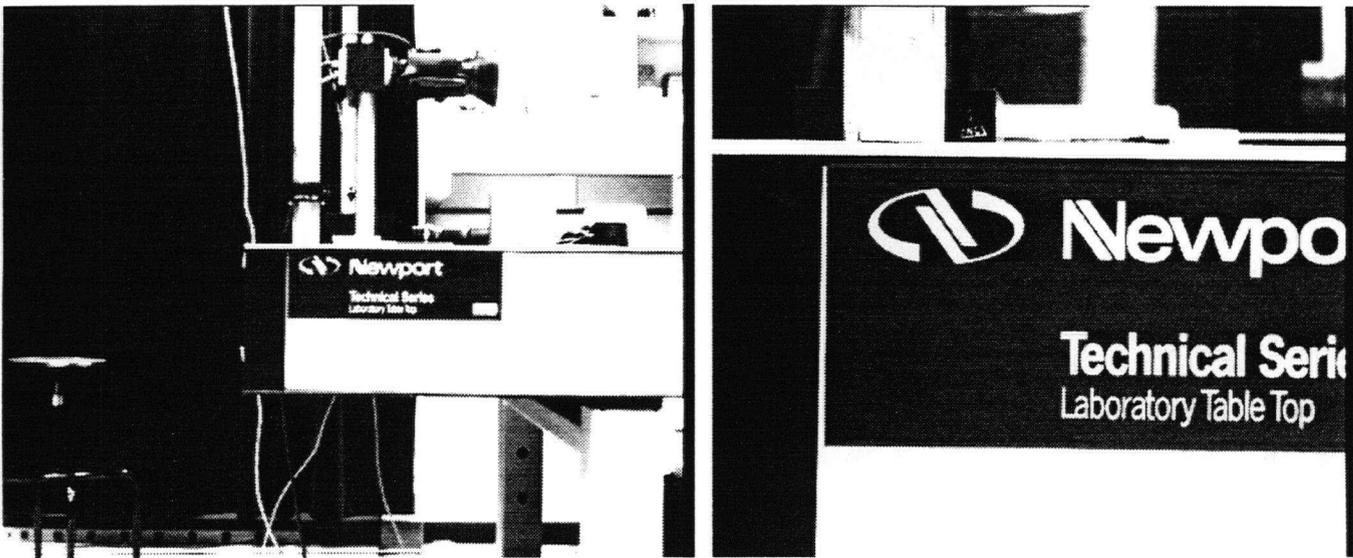
$$\begin{aligned}x' &= x k_1 r^2 \\y' &= y k_1 r^2 \\r^2 &= (x - x_c)^2 + (y - y_c)^2\end{aligned}\tag{7.2}$$

where  $X' = (x', y')$  represents the distorted location of the point  $X = (x, y)$ ,  $k_1$  is the coefficient of the radical distortion and  $(x_c, y_c)$  is the true optical centre.

In our experiment, first order radial distortion is recovered, following the corner match. The results show no significant visual differences. Therefore, this step is not included in our final implementation. Other image enhancement schemes, such as deconvolution using point-spread function, are tested, but showed no consistent improvement of the presentation, and are not adopted in our final approach.

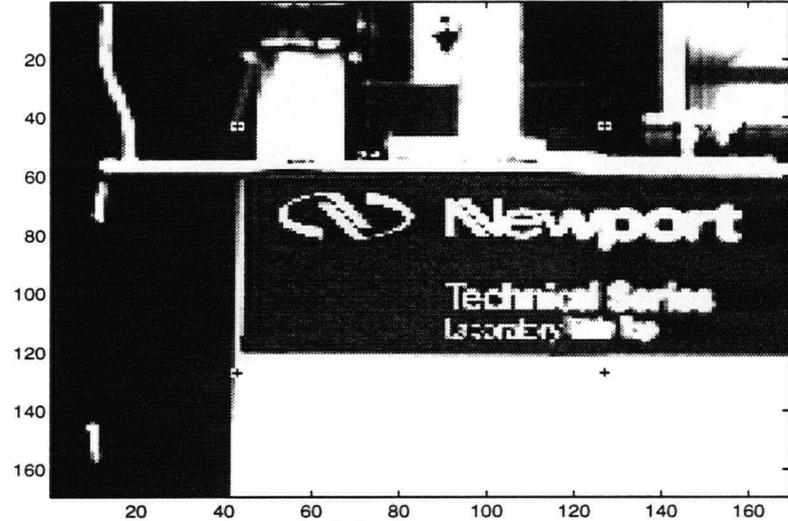
## 7.2 Up-sampling & Down-sampling

Re-sampling is needed to present two pictures on one screen, where the resolution is uniform. To have better overall resolution, up-sampling the low resolution background is needed, the disadvantage is that a much higher pixel numbers are required for the display which is not always available. Another way to fuse two pictures is to down-sample the high resolution image to match the low resolution picture. The disadvantage is the loss of the details in the insert. The result of both approaches are shown in Fig. 7.1. There is in-between sampling as well, where both up-sampling and down-sampling are done to meet in some level of middle resolution. The disadvantage is twice the computation. The resampling are done the same way as in section 4.5, with linear interpolation.



Background (Top Left)  
 Insert (Top Right)

Down-Sampling  
 (Second Row)



Up-Sampling &  
 Down-Sampling  
 (Bottom)

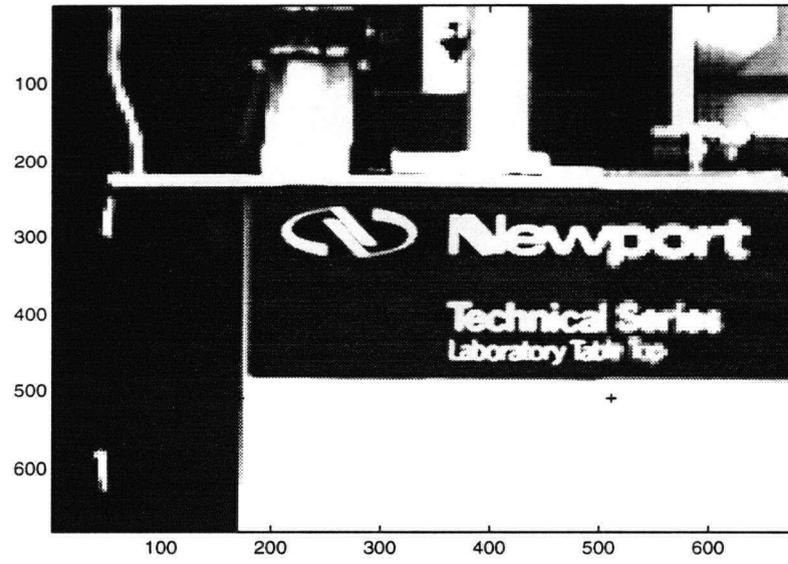


Figure 7.1 Comparison of Up-sampling and Down-Sampling Results

### 7.3 Normalization and Smoothing in Transition Band

Following the re-sampling will be the intensity level normalization of the transition band between AOI and WFOV. Both mean and variance normalization is applied to obtain better results. This is done by normalizing the WFOV image with the insert mean and variance, as

$$I_{WFOV}(i, j) = (I_{WFOV}(i, j) - M_{WFOV}) / V_{WFOV} * V_{AOI} + M_{AOI} \quad (7.3)$$

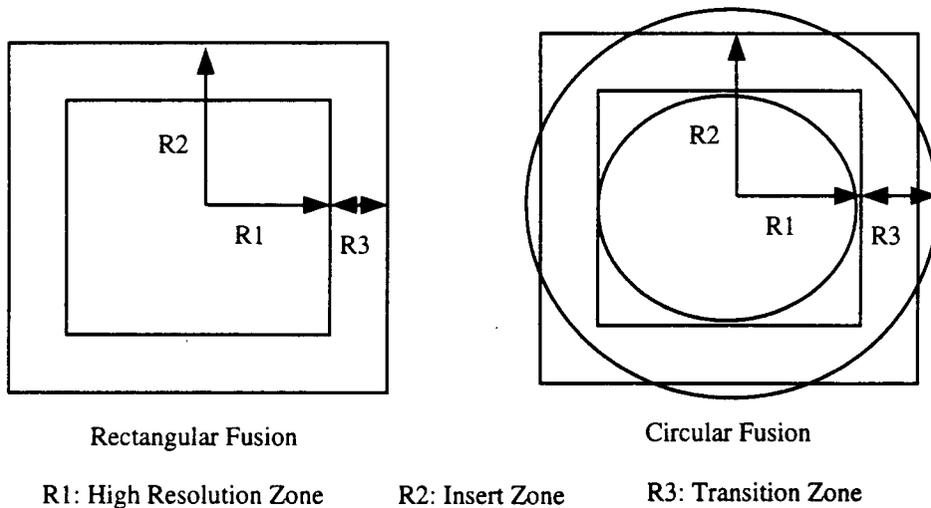
where  $M_{WFOV}$  and  $M_{AOI}$  are the means of the insert area of the WFOV and AOI, while  $V_{WFOV}$  and  $V_{AOI}$  are the variance of these two areas.

In smoothing the transition band, the principle is to keep the high resolution insert area as much as possible, while trying to smooth the sharp intensity changes around the boundary. As discussed in Chapter 3, there are several ways of doing this, and method (iii) is chosen with linear weighting proportional to the distance. That is,

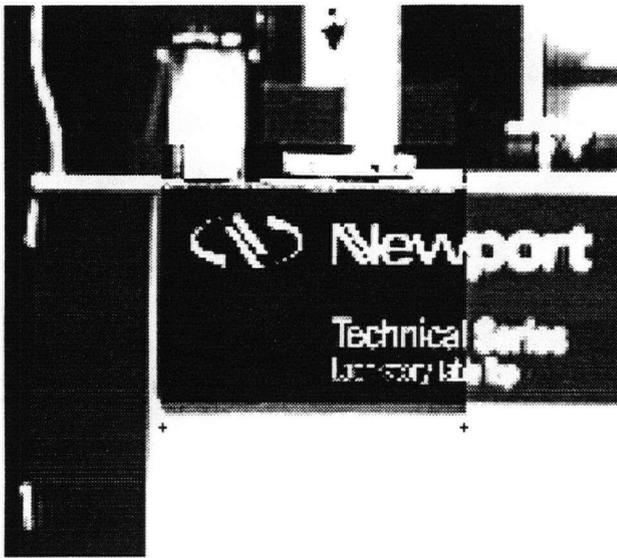
$$I_{fusion}(i, j) = I_{AOI}(i, j) * (1 - r/R3) + I_{IFOV}(i, j) * r/R3 \quad (7.4)$$

where  $r$  is the pixel distance to the image centre,  $R3$  is the width of the transition band chosen as 10% of the AOI or  $R1$ , and  $R2$  is the WFOV.

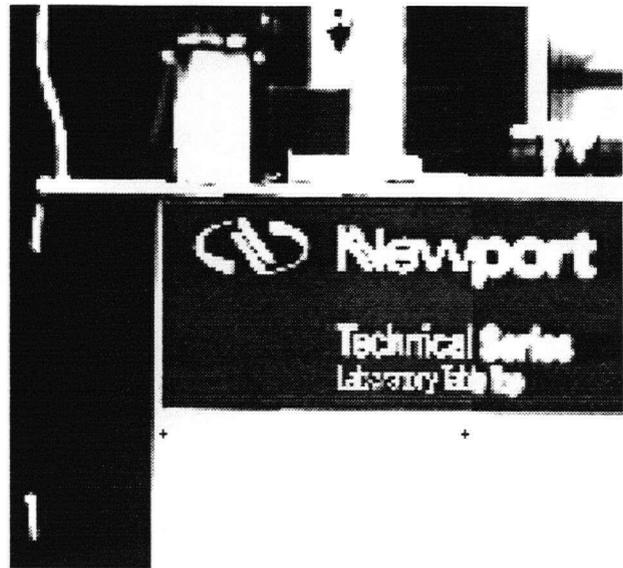
The transition band was chosen to be rectangular as shown in Fig. 7.2, since this gives larger high resolution region, and experiments show better visual effect as well. Mean and variance normalization are used to maintain intensity match between insert and background.



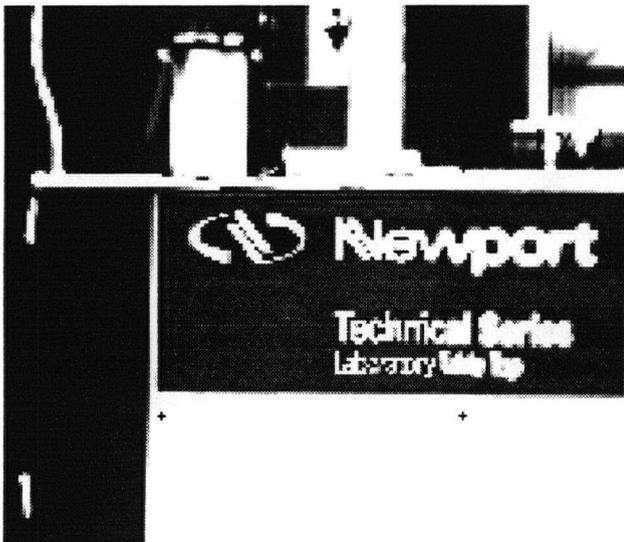
**Figure 7.2** Comparison Between Rectangular transition & Circular Transition



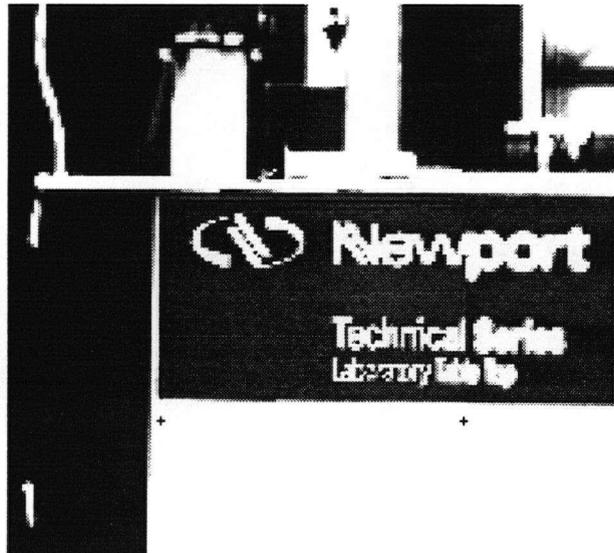
Insert Replacement



Normalized Insert



Circular Transition Fusion



Rectangular Transition Fusion

Figure 7.3 Comparison of Different Smoothing Methods

## **7.4 Proposed Approach**

From the analysis and experiments obtained in the previous chapters, we can see that a different process should be applied for different imaging conditions. For a small range of RST changes, a corner-based process worked well and efficiently, while a hierarchical BFS approach provides better performance for a larger range of RST changes, and should be used to obtain a finer estimation of RST before applying corner-based process. As a result, a multi-process approach is proposed for final implementation, as shown in Fig. 7.4. The idea is to use the optimal process for various initial values, i.e., the process that is most suitable for recovering the given range of changes. The flowchart simply combines the hierarchical BFS and corner-based approach discussed in Chapters 4 & 5. The details can be found in these two chapters.

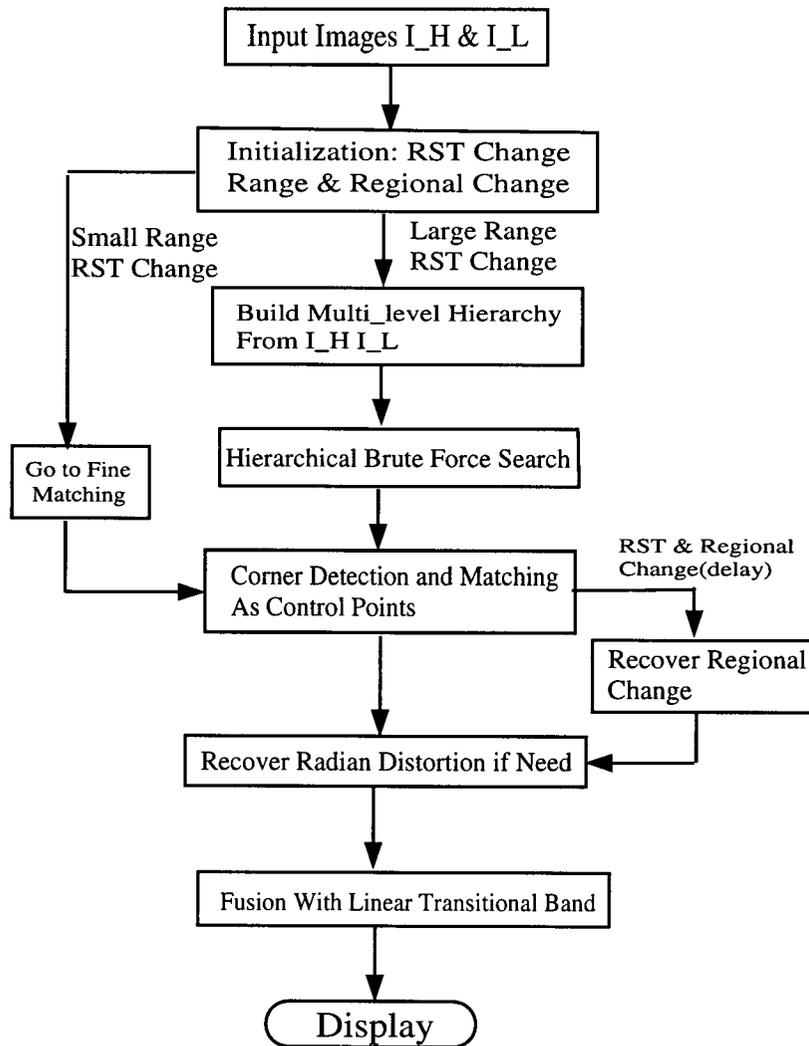
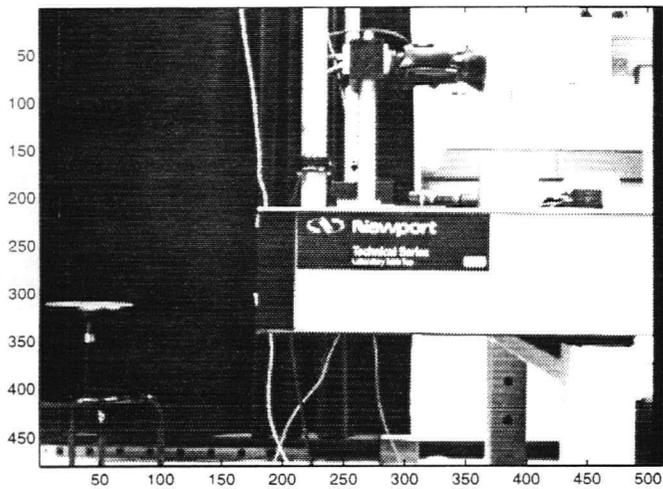
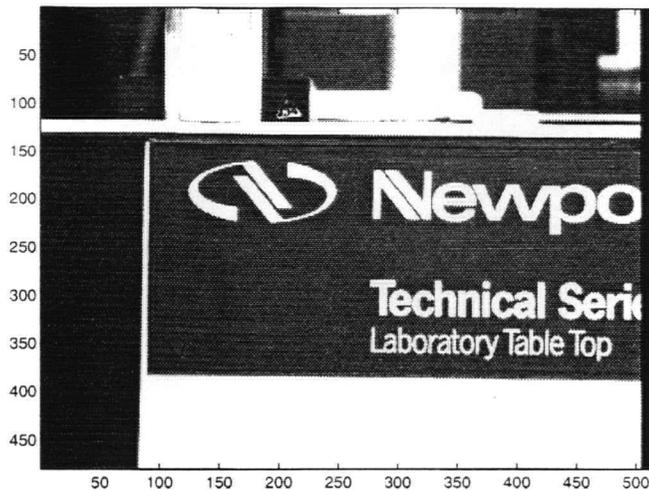


Figure 7.4 Flowchart of Multi-process Approach for RST Estimation

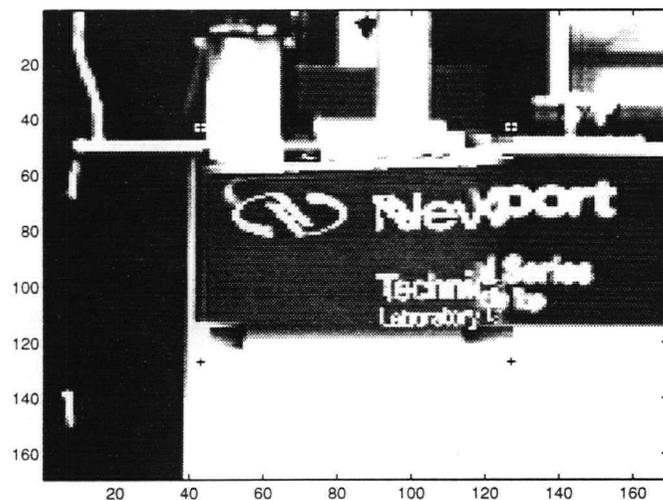
The processed images after each step of this approach are shown in Fig. 7.5–7.11. The results show consistent improvements of image merging after each additional step, and verified the effectiveness of the proposed approach. The images taken consist of variety features, and can represent typical outdoor images as well, such as the map and the plant image sets.



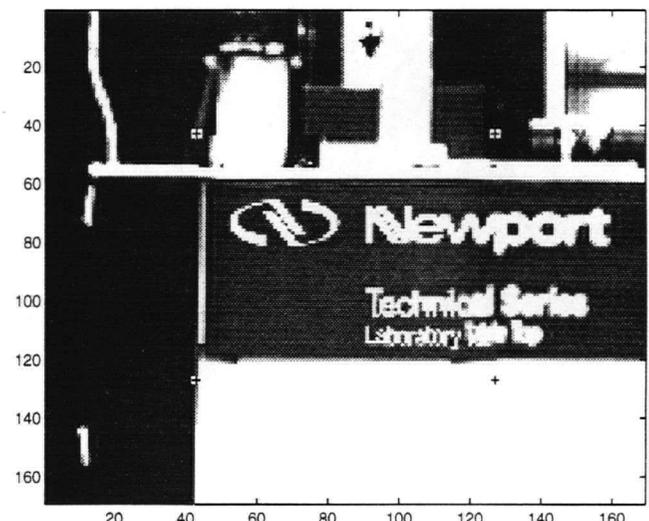
Background



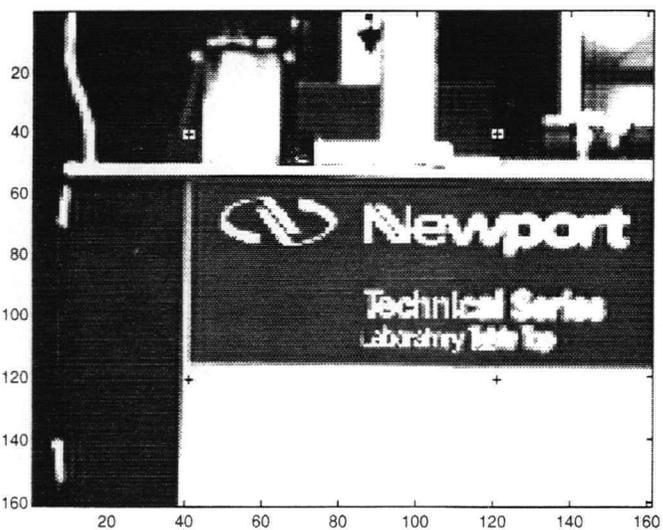
Insert



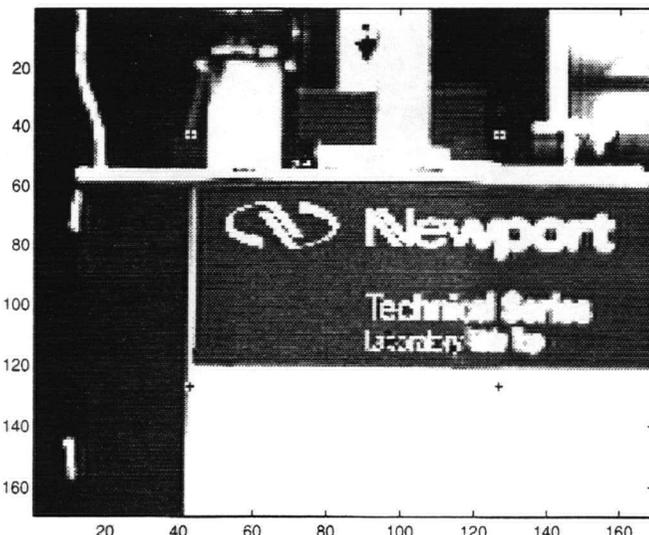
1BFS



2BFS

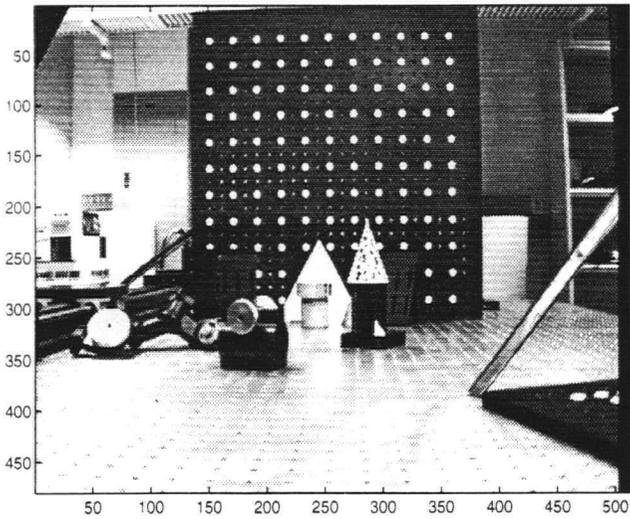


Corner Match

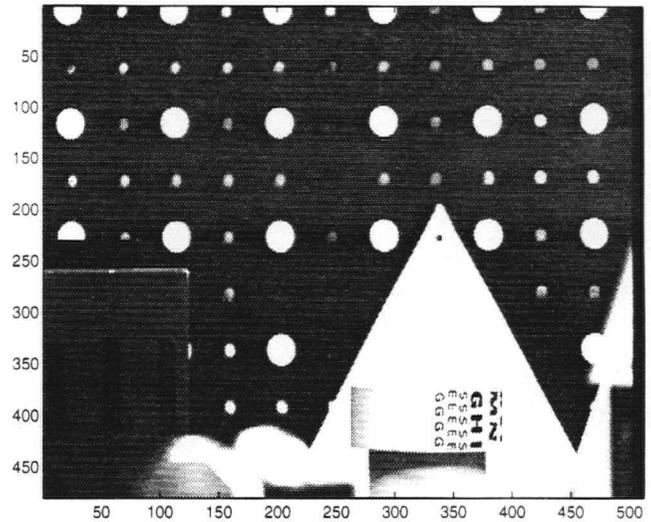


4BFS

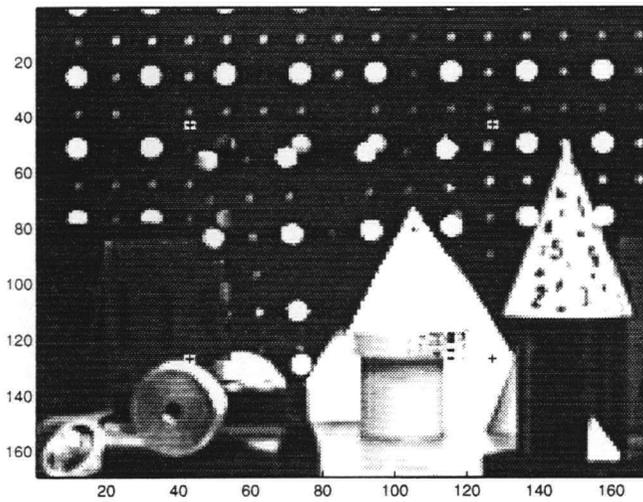
Figure 7.5 The Processed Images After Each Step in Proposed Approach



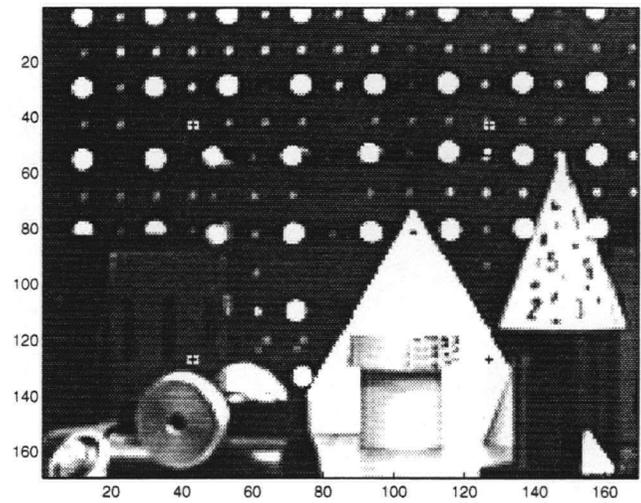
Background



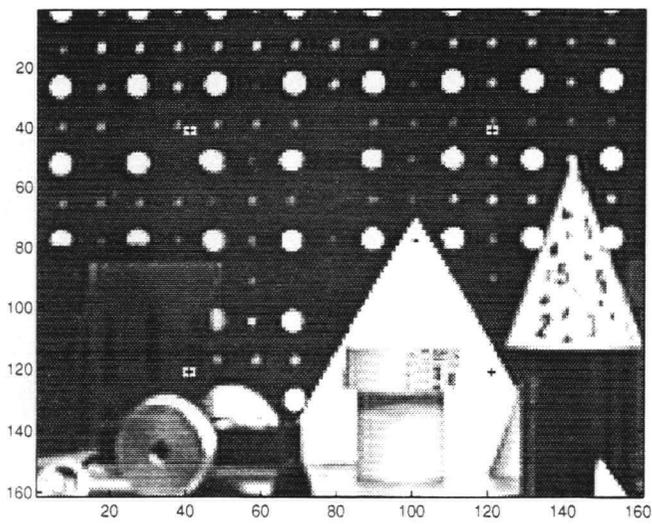
Insert



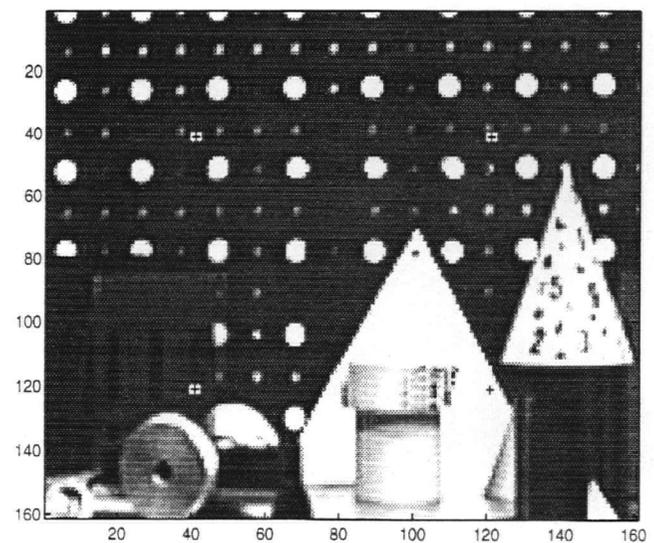
1BFS



2BFS

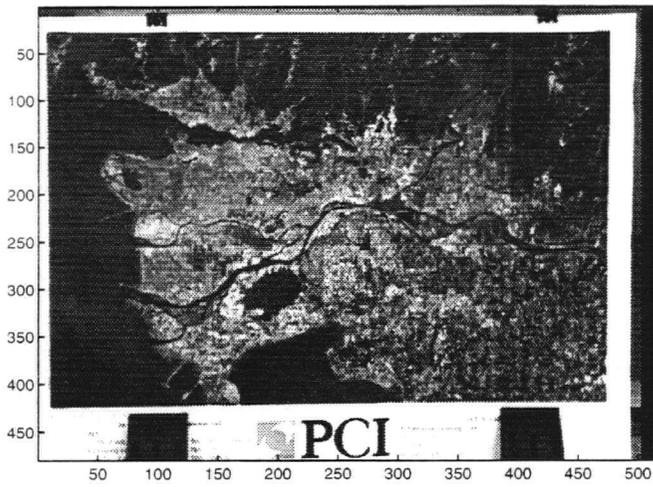


Corner Match

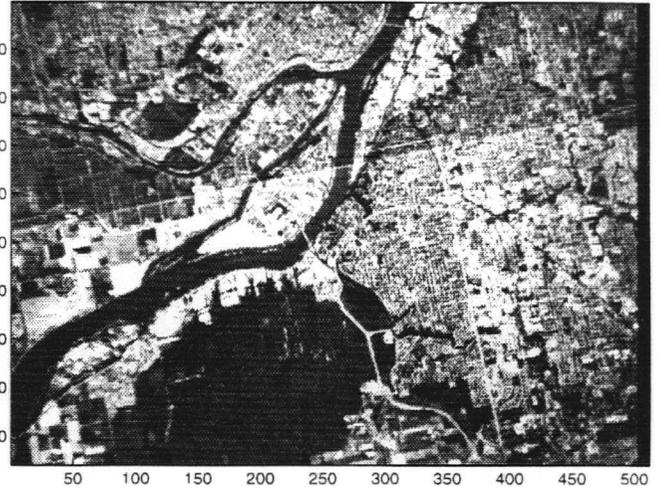


4BFS

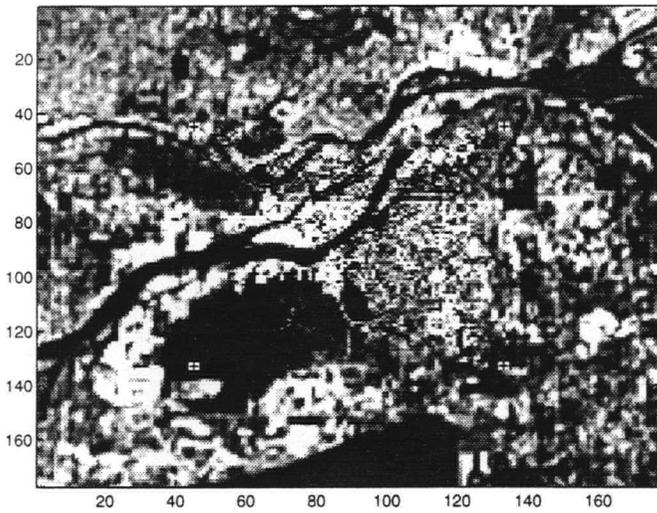
Figure 7.6 The Processed Images After Each Step in Proposed Approach



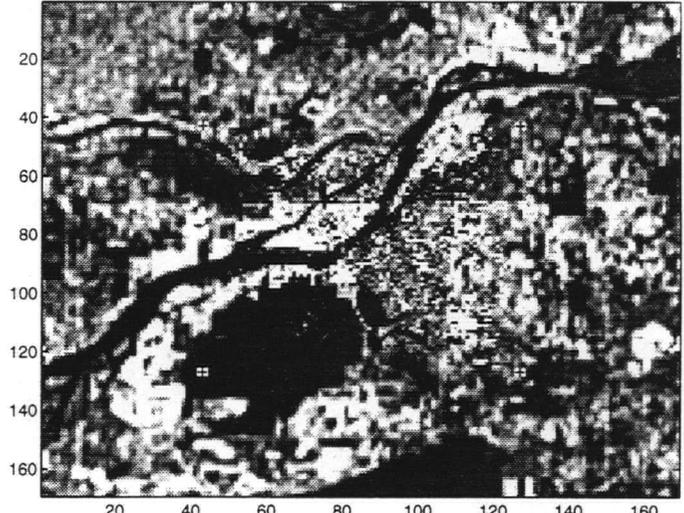
Background



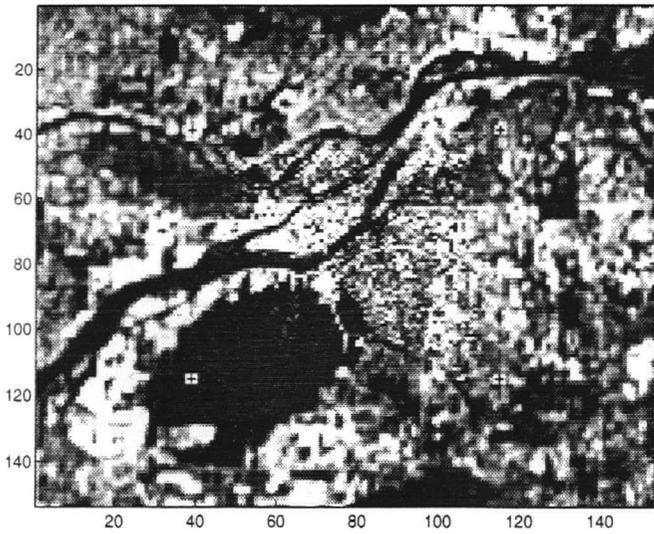
Insert



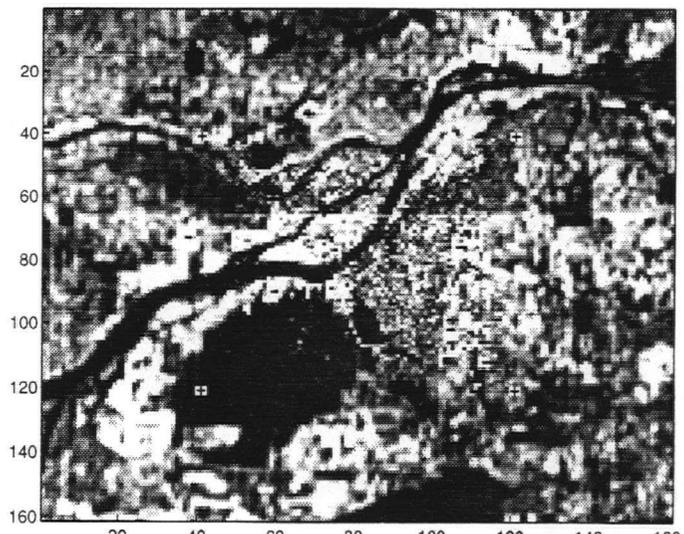
1BFS



2BFS

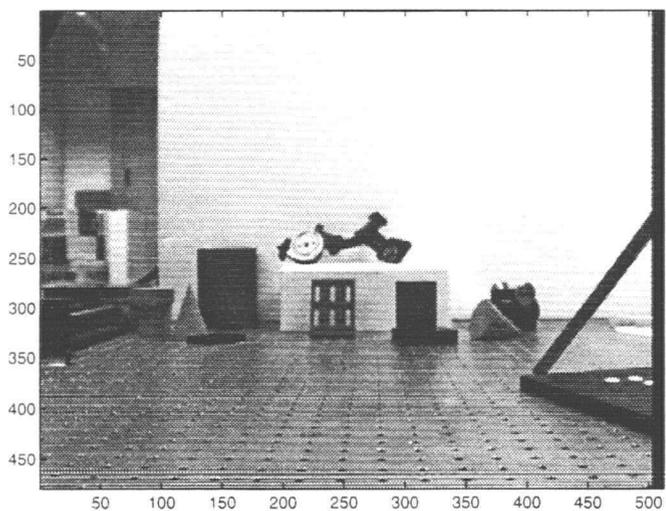


Corner Match

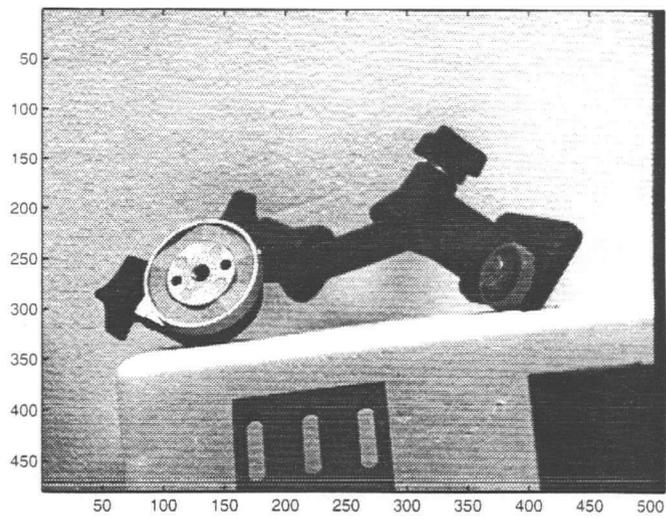


4BFS

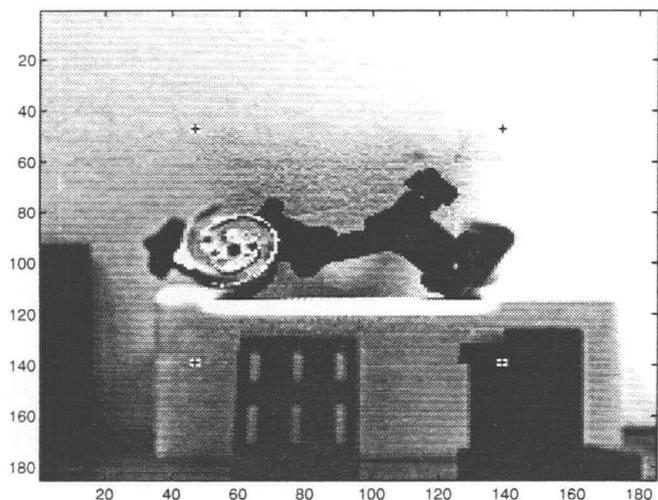
Figure 7.7 The Processed Images After Each Step in Proposed Approach



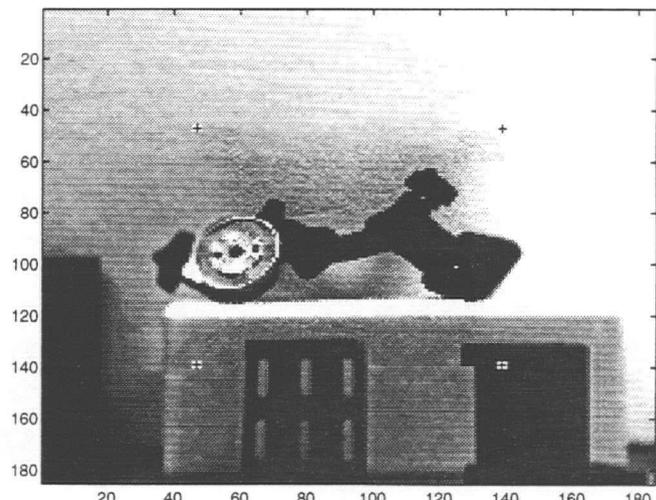
Background



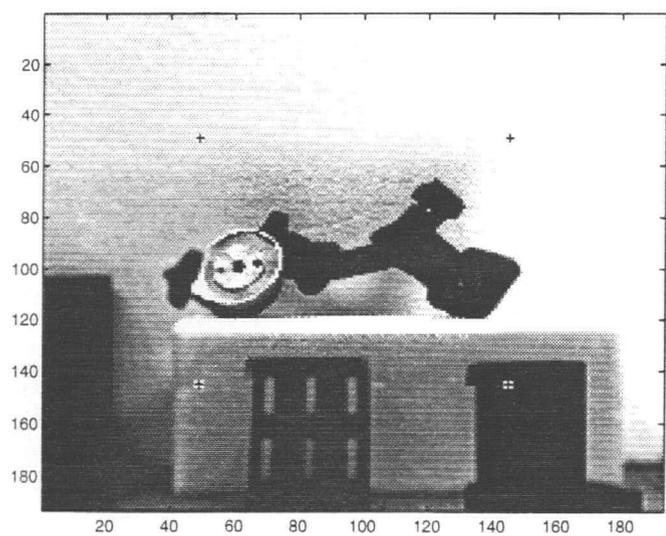
Insert



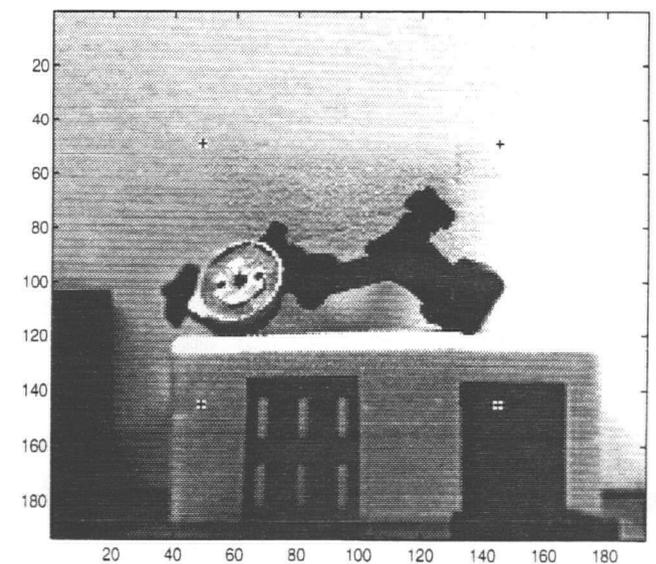
1BFS



2BFS

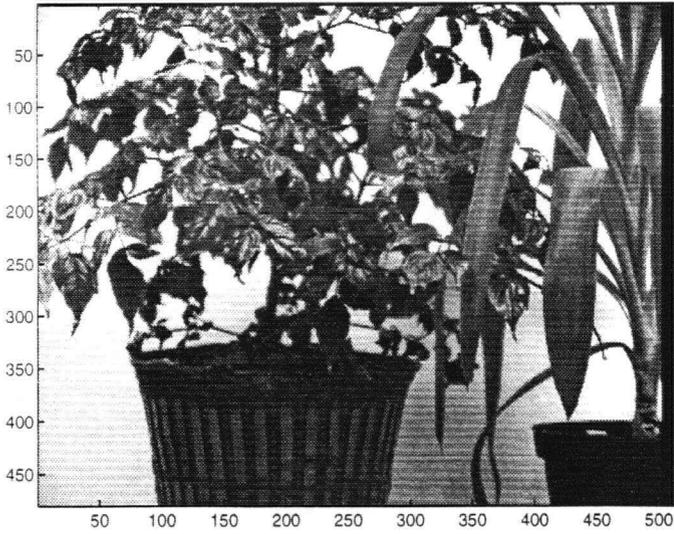


Corner Match



4BFS

Figure 7.8 The Processed Images After Each Step in Proposed Approach



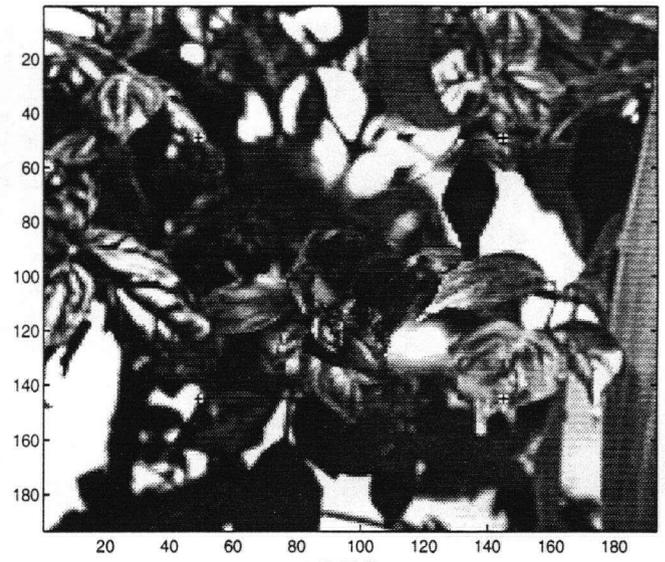
Background



Insert



1BFS



2BFS



Corner Match



4BFS

Figure 7.9 The Processed Images After Each Step in Proposed Approach

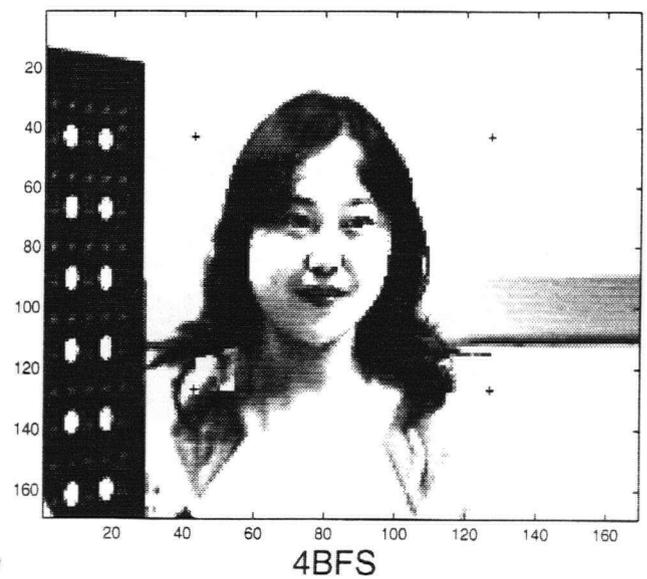
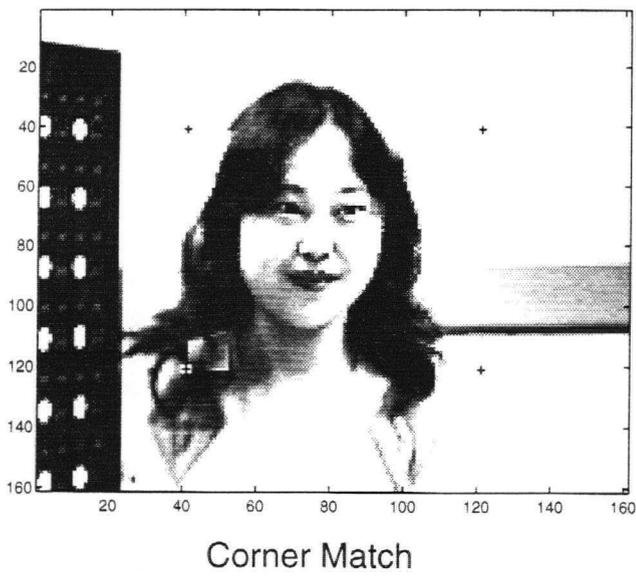
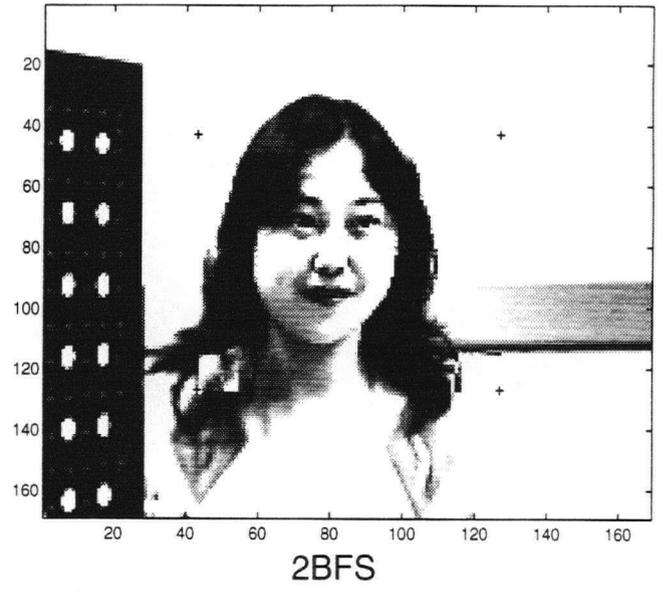
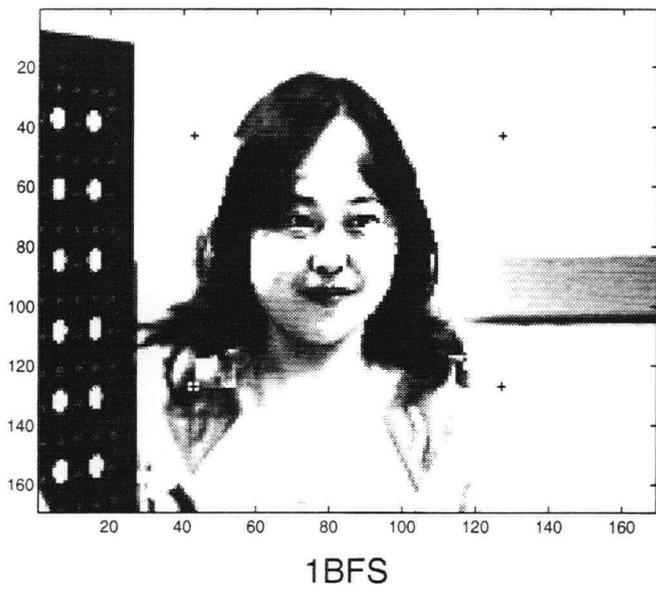
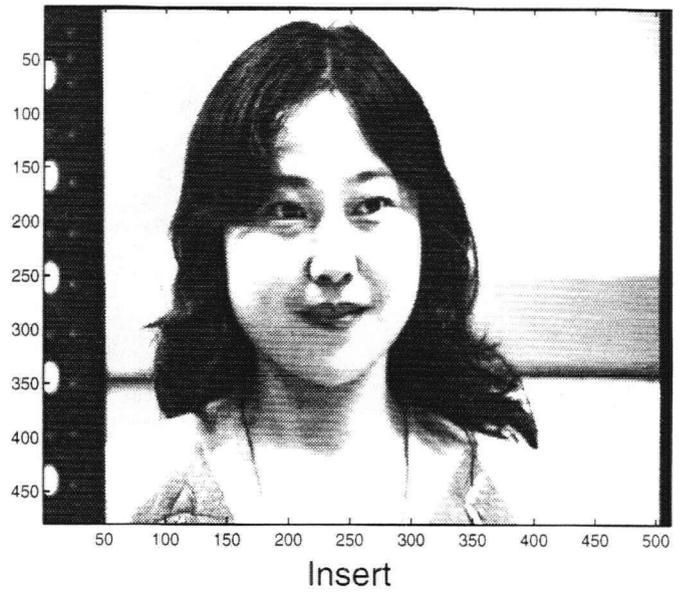
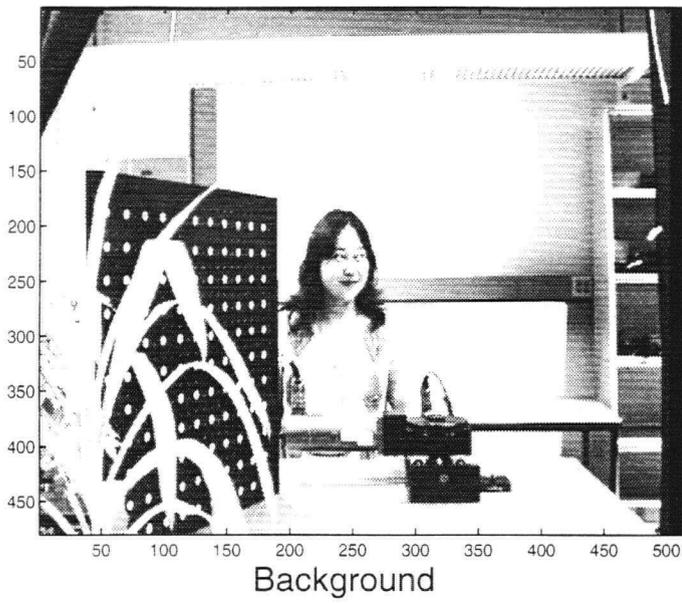
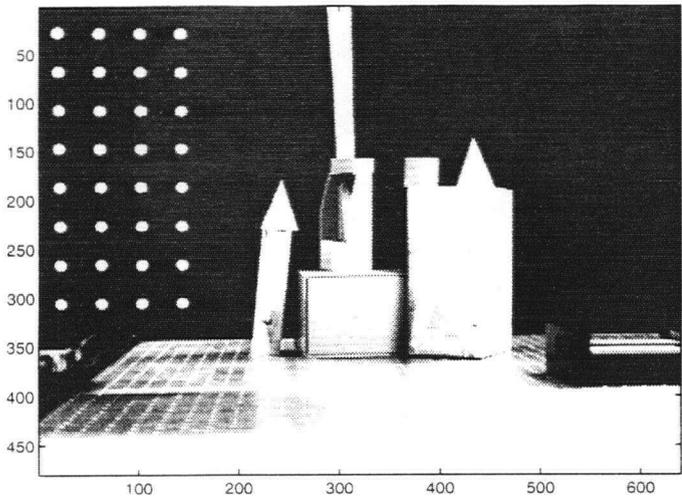
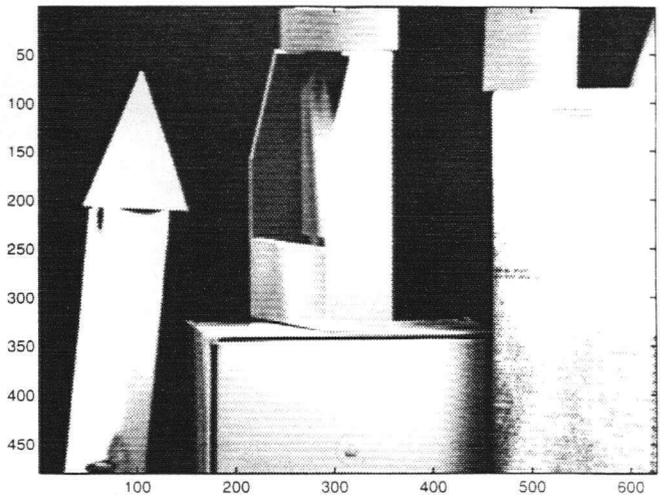


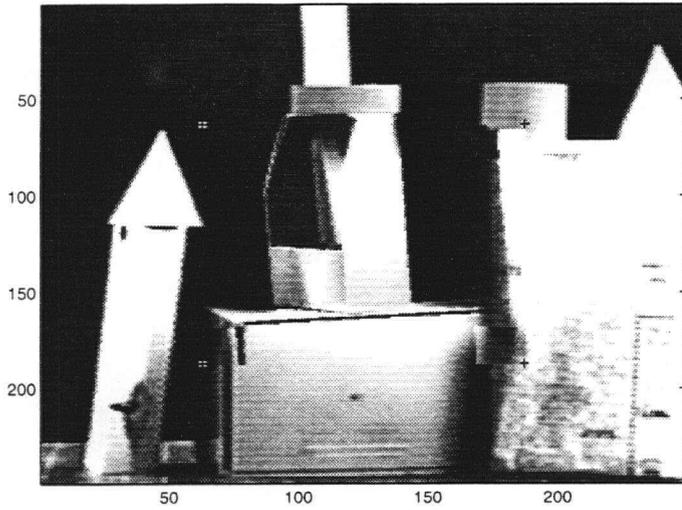
Figure 7.10 The Processed Images After Each Step in Proposed Approach



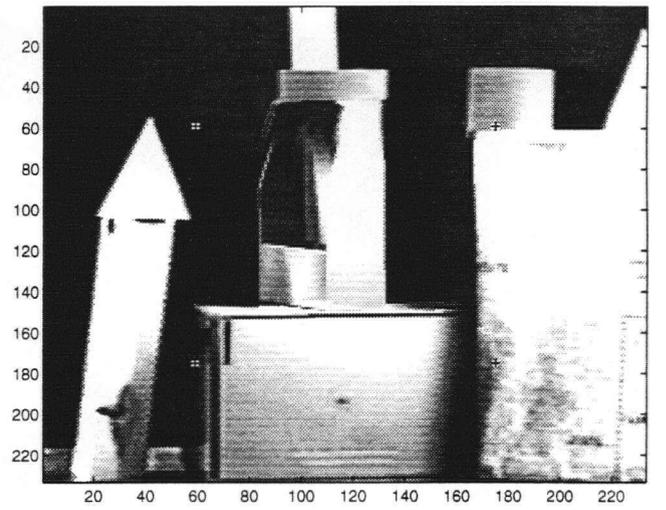
Background



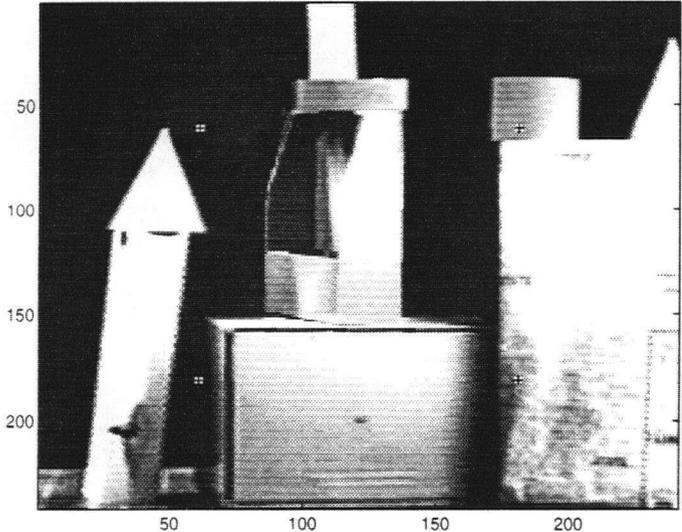
Insert



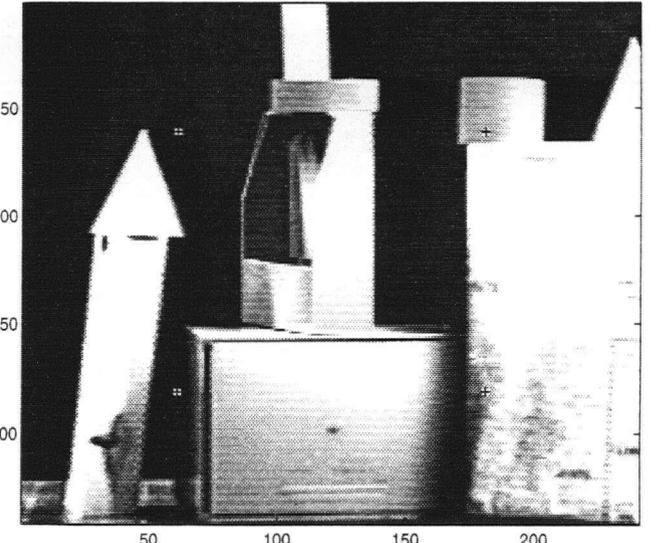
1BFS



2BFS



Corner Match



4BFS

Figure 7.11 The Processed Images After Each Step in Proposed Approach

## Chapter 8 Conclusions and Future Work

### 8.1 Conclusions

This thesis has addressed some of the issues in developing a DVC system for tele-operation applications. These include the system specifications, implementation technology, and the processing system. Examination of the current technology shows that there are many technical difficulties in matching the capabilities of the human eye. Two of the most advanced DVC systems are introduced as the state of the art system models with up-to-date performance. For industrial applications, there are demands for more economic implementations, especially WFOV imaging and display, high resolution insert tracking and projection device, better display resolution, etc. A few design ideas are proposed for improving current DVC systems.

The processing system for DVCS is a new issue since the images are taken from remote cameras instead of computer-generated as in other DVCS such as flight simulator displays. This thesis performed analysis and experiments on this topic. A final approach for image merging is proposed and implemented that gives best trade-off of the performance characteristics for various imaging conditions. In summary, the conclusions are:

- A correlation-based brute force search is reliable and accurate, but can be computationally expensive;
- A corner-based approach using correlation window in matching is efficient in recovering small range of RST changes, but may fail in featureless image pairs, and can be unreliable and computationally expensive when the RST change ranges are large;
- A corner-based approach using attributes in matching proved to be very difficult in recovering large RST changes. There are many issues to be resolved before its practical implementation;
- A hierarchical approach can reduce the computation cost of the correlation-based brute force search, by adapting a different step size and window size at different levels of

resolution, and by using low resolution estimation to reduce the search range at higher resolution levels;

- A approach which combines the hierarchical BFS and corner matching can provide the performance for various imaging conditions and is implemented and recommended as the final approach;
- The proposed calibration-free image registration of AOI and WFOV approach is designed to minimize operations for real-time implementation purposes;
- This software implementation can relax the requirements for position sensing and camera control devices.

## 8.2 Future Work

A DVCS is a complex system and is still in the developmental stage. Since none of the current DVCS can present ideal scene, there are many works to be done for improvements. As new technology and processing develops, better system will evolve. Increasing computing power will have great impact as hardware specs can be relaxed by computer processing capabilities. Some of the future work includes:

- Design of an application specific computing process, including using parallel computing for real-time implementation;
- Design of a programmable image computing board, which can replace individual processing boards such as correlation board and warping board. This will significantly improve the overall software system by reducing the volume and cost;
- Better interfaces for the ease of use;
- Hardware designs for components such as wide angle imaging, position control devices to achieve cost efficiency required for industrial applications.
- Additional features such as stereo, color and augmented information in the presentation.

## Appendix A

### Imaging System Using Intensity Array

**Abstract** This paper introduces a new concept of intensity array and the digital imaging system using this array. This system has the potential of wide angle of view, high resolution and variable resolution pattern, etc. We will present the concept first, give a simple design as an example and then analyze the possible problems that might occur in implementation.

### Section 1 Introduction

In medical imaging system, such as x-ray, r-ray, there is a high demand for an imaging system that can provide both aperture and resolution since no lens is available. Pinhole array (or called coded aperture imaging) is introduced as a mean to improve the SNR (aperture) while keeping the resolution of small size of holes. The recorded picture will consist of many overlapping images of the emitting object and, in general, bears no resemblance to the object. Computer or optical processing of the picture is required to produce the reconstructed object which should resemble the original object. In this implementation, the detector should be large enough to observe the entire shadow produced by any source within the field of view. This restricts the angle of view, the size of the source image, and the detector size, etc.

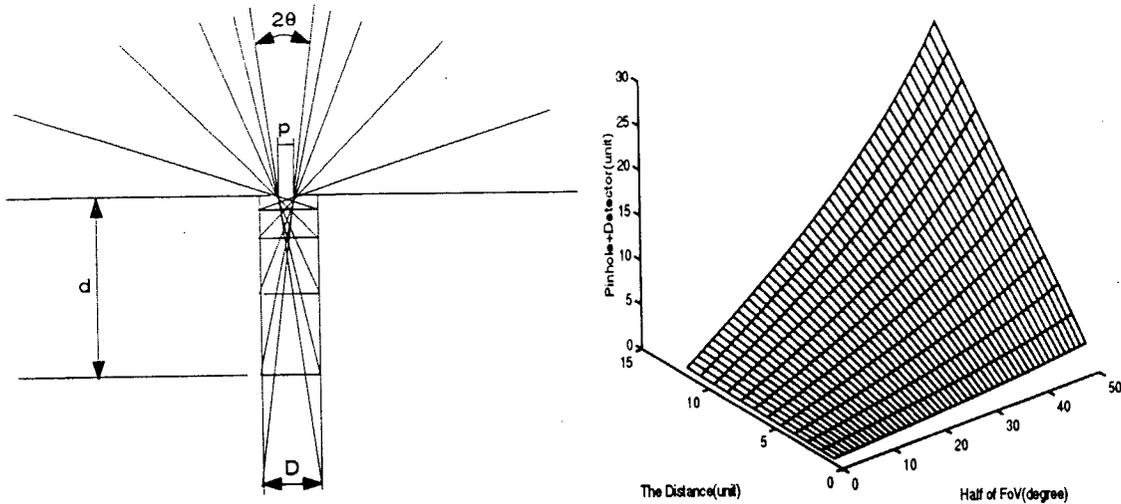
3) in visual spectrum region, light diffraction by the pinhole causes the poor resolution.

### Section 2 Concept of Intensity Array & the Principles

After studying the advantages and disadvantages of pinholes and pinhole masks, we propose a new concept of intensity element and intensity array which can be used in digital imaging system at very low cost of computation time.

We define an intensity element as a pinhole with a detector (sensing material) and an intensity array as an array of such intensity elements, see Fig.1. It is easy to obtain the relation between the field of view (FoV) and the size of the intensity element from the figure. Assume the FoV is  $2\theta$ , pinhole length is  $P$ , detector length is  $D$ , the distance between the pinhole and the detector is  $d$ , then, we have:

$$\tan \theta = \frac{P+D}{2*d}, \theta = \arctan \left( \frac{P+D}{2*d} \right)$$



**Fig.1 The Intensity Element and the Relations Between its Parameters**

As we can see from the equations, we can get intensity elements with different FoV simply by changing the structure of the pinhole. This give us the chance to set up a array consist of intensity elements with different FoV. The idea is to have each intensity element cover different spatial resolution points of the real world(FoV), and build up the relations between the intensity array and the spatial resolution point of the real world. By solving these equations, we can obtain the image of the real world from the intensity array. A distortion-free, wide FoV imaging system can be obtained if we set up the intensity array properly. By properly, we mean several principles to follow in setting the intensity array. We summarize as follows:

**The principle of singularity:** Any two intensity array elements should have different FoV or cover at least one different spatial resolution point in 2-D, and vice versa, any two spatial resolution points in real world should corresponds to at least one different intensity array element.

**The principle of resolution distribution:** The spatial resolution pattern of the real world is mainly determined by the overall FoV, the dimension of the intensity array and the distribution pattern of the intensity array.

**The principle of no-disparity:** The overall size of intensity array should be small enough that assumption of no disparity can be made. This restricts the number of intensity elements we can have, which means the spatial resolution points in real world is also restricted since the dimension of the intensity array decides the number resolution points of the real world.

Once the array is set, the transform function between spatial resolution points in real world and the intensity array is fixed and can be determined in advance. So, only matrix multiplications are needed to obtain the real world image from the intensity array, we will give an example in the following section. In case of non-uniform-resolution or non-traditional resolution pattern, interpolation and resampling will be needed as well. Since this is a variable imaging system, the optimum angular resolution pattern can be arranged according to human visual perception pattern.

### Section 3 Simple Design & its Equations

Assuming that the intensity array has a) no diffraction, b) no-interference between intensity elements, and c) each intensity element covers integer number of spatial resolution points in 2-D, in addition to the principles stated in the previous section, we can set up simple relation between real world and the intensity array with singular solution. We call this ideal case, and non-ideal case will be discussed in the next section.

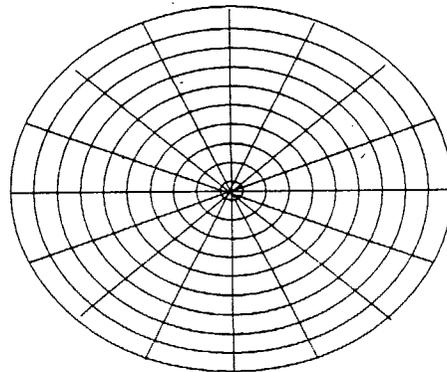


Fig. 2 One Design of Intensity Array

Fig.2 gives one example of such an intensity array.

Suppose the coverage of each pinhole is  $M$  in angular direction,  $M=1$  represents one spatial resolution coverage in angular direction. In radial direction, assume the coverage is from the current point to the centre point. For  $M=1$ , we have the following relation:

$$R = M_L * I * M_R, \text{ where}$$

$$R = \begin{bmatrix} R_{11} & 0 & \dots & 0 \\ R_{21} & R_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ R_{N1} & R_{N2} & \dots & R_{NN} \end{bmatrix}, M_L = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix},$$

$$I = \begin{bmatrix} I_{11} & 0 & \dots & 0 \\ I_{21} & I_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ I_{N1} & I_{N2} & \dots & I_{NN} \end{bmatrix}, \text{ and } M_R = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

where  $R$  is the detected detected intensity array image,  $M_L$  is the left modification matrix,  $I$  is the real world image,  $M_R$  is the right modification matrix, and finally,  $I = M_L^{-1} * R * M_R^{-1}$ .

There are many possible designs, as to which one is more practical highly depends on the manufacturing side.

## Section 4 Reality & the Possibilities

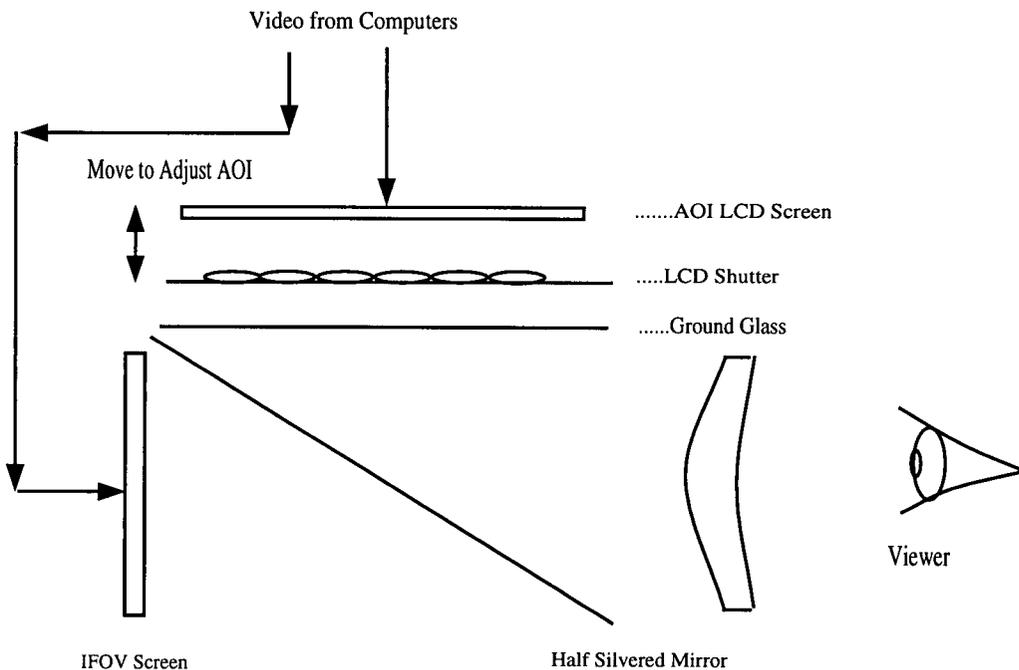
In the visual spectral region the ordinary pinhole camera is of limited usefulness, because of its poor resolution due to light diffraction by the pinhole.

In non-ideal case, first, there is a restriction introduced by the diffraction, which means the size of the pinhole has a limit, this is the main reason we emphasize our design in X-ray and  $\Gamma$ -ray camera, since with X-ray and  $\Gamma$ -rays, diffraction is usually of no consequence. Although application in visual light is possible, but the benefit wouldn't be the same. Second, to meet the assumption of non-disparity between pinholes, the array size has to be small, which means limited dimension of the intensity array which also means limited resolution. Third, some noise will be introduced by non-integer number of coverage of the resolution points of each pinhole, interference from neighboring intensity element due to non-complete absorption of the energy of the isolating material, etc.

## Appendix B

### Moving Insert Using a Microlens Array

The idea is to use an array of lenses to shift the image to the location where the insert is expected by eye-tracking signal. With the object located at a significant distance from the array (within the aperture of all the lenses), each lens forms a miniature image of the object being photographed. In our application, we go further than forming multiple images, we use an LCD panel to control which lens to use each time instant, i.e., one lens is valid each time while all the others are closed (no passage of light through these lenses). In this way, we are able to move the insert around the background. See Fig. B-1.

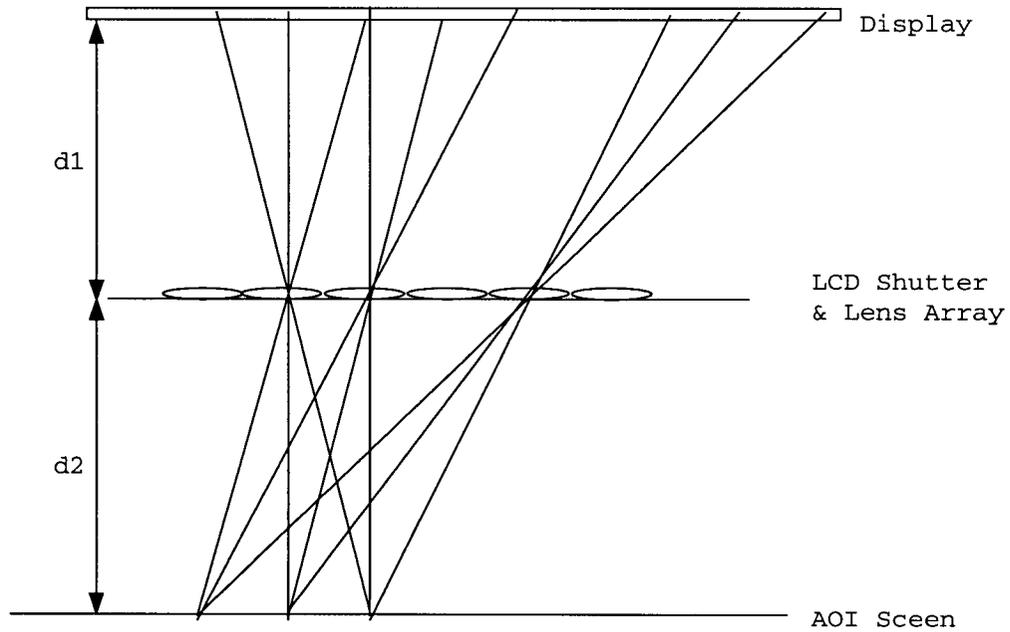


**Figure B.1** The Setup of the Optical System with Moving Insert

When designing this lens array, the dimension of the array and the size of the lens should be considered according to the screen size of both insert and background, insert size on the display, distortion tolerance and the frame size of the system, etc. On one hand, the insert is expected to perform the eye-tracking function, i.e., to move around the background. For the insert to move around the background smoothly, the larger the number of lenses the smoother the movement, i.e.,

the finer the spatial resolution of eye-tracking . On the other hand, smaller lenses means lower overall luminance and higher distortion due to diffraction. If the lens size is too small, the luminance of the insert can be too weak compared to the background, though we can increase the overall luminance of the insert, there is a limit as to how much we can adjust the luminance. Suppose we de-magnified the insert by  $N$  times, theoretically, the lens size (considering the luminance) should be the LCD panel size divided by  $N$ , if the lens size is smaller than this, we need to enhance the luminance of the insert panel by  $\text{LCD size}/N/\text{Lens size}$ . To make the full use of the lens size within limited space, the best design should be a square lens as shown. We choose the size of each lens to be between  $\text{LCD size}/N$  to  $\text{LCD size}/4N$  which make a compromise between the luminance and smoothness of the insert movement. The lens size is also restricted by light diffraction, in theory, the lens diameter should be larger than 1mm to avoid significant diffraction effect. See [Davies&Lau, 91]

Since the dimension of the lens array is very restricted, the system's ability in eye-tracking performance is also very restricted, since it gives a big step between each insert movement, as if the insert is jumping from one place to another. So in addition, we introduce an electronic way to move the insert between these gaps. Instead of using the whole screen to present the insert, we use a part of the screen to present one frame of insert picture each time, and move the insert around to the full use of the screen. Combining the lens array and electronic movement, we are able to move the insert around the whole display smoothly. From Fig. B-2, we can calculate the parameters of the display system.



**Figure B.2** The Microlens Projection

The gap is

$$\frac{N * D_{lens}}{D'_{gap}} = \frac{d_2}{d_1 + d_2} = R \quad (B.1)$$

$$D'_{gap} = \frac{N * D_{lens}}{R}$$

$$D_{gap} = D_{lens} / R$$

where  $D_{lens}$  is the size of lens,  $D_{gap}$  is the distance of neighboring lens' projection centres,  $d_1$  is the distance between display and lens array,  $d_2$  is the distance between lens array and insert screen,  $R$  is the ratio of these distance,  $D_{move}$  is the moving range of the insert screen,  $D_{LCD1}$  is the LCD length and  $D_{LCD2}$  is the LCD width.

To cover the gap, the centre of the image should be able to move, to get  $D_{move}$  as follow:

$$\frac{N * D_{lens} / R + D_{lens} / 2R - N * D_{lens}}{D_{move} + N * D_{lens}} = \frac{d_1}{d_2} = \frac{1 - R}{R} \quad (B.2)$$

$$D_{move} = \frac{D_{lens} / 2}{1 - R}$$

While trying to move the insert smoothly, we also want to keep as many pixels as possible. The following table shows their relations. The percentage of pixels being used is as follow:

$$P = \frac{(D_{LCD1} - D_{move})(D_{LCD2} - D_{move})}{D_{LCD1} * D_{LCD2}} =$$

$$1 - \frac{D_{move} * (D_{LCD1} + D_{LCD2})}{D_{LCD1} * D_{LCD2}} + \frac{D_{move}^2}{D_{LCD1} * D_{LCD2}} \quad (B.3)$$

$d_2$	R	focal length of lens	$D_{move}$	Pixel Usage Percentage
$d_1$	1/2	$d_1/2$	$D_{lens}$	72.25%
$d_1/2$	1/3	$d_1/3$	$3/4 D_{lens}$	76.2%
$d_1/3$	1/4	$d_1/4$	$2/3 D_{lens}$	81.1%

**Figure B.3** Relation of The System Parameters

## Appendix C

### Pixel Effect Removal for LCD Screen

In a high quality display system like a helmet-mounted display (HMD), the gap between each pixel can be very annoying since the human eye has an acuity of vision which gives very high resolution perception. A person can distinguish two separate points 6 microns apart on the display if the display is 50mm away from the eye, i.e., the gap size between neighboring pixels on the display should be less than 6 microns in order for the eye to ignore the gap, otherwise, the eye will recognize the gap and the whole image will look like an array of separated dots. Now, it's not difficult to get a display component like an LCD panel or fibre optic bundle with pixel gaps less than 6 microns, the problem is when displayed to the eye in some display system like HMD, there will be a magnifying lens which focuses the image to infinity and at the same time enlarges the image. As a result, the gap is magnified at the same time. So, there is a need to remove these gaps even if they are less than 6 microns. As to what size of gap is tolerable will depend on the magnifying lens and the distance between the eye and the display.

To remove the gap effect, we propose a smoothing method by interpolating the gap area with neighbouring pixel gray levels. The way to realize it is to use a vibration system to move the fibre optics bundle or the LCD panel in such a pattern that the neighboring pixels will move back and forth from the gap position and cover the gap in such a pattern that gives the overall effects of some sort of interpolation. This is justified by the fact that the human eye has limited resolution perception of stimuli in time which is called the fusion frequency. The minimum frequency of the subjective fusion is called the critical fusion frequency (CFF) as illustrated in Fig. 2.5. The maximum CFF is 50 Hz under all luminance conditions. So, we need to vibrate the fibre bundle or LCD panel at the frequency twice of that which is 100 Hz to eliminate the vibration effect itself.

Here, we will discuss different vibration patterns for different display systems, and give the results of vibration effects that's equivalent to the interpolation. There are basically two types of vibration. First, a linear movement, move along the vector which is a combination of all the necessary movement vectors, this will give the result of filling the gap with near linear interpolation,

as shown in Fig. C-1. Second, we can move the display end along a circle so that the gap will be filled by neighboring pixels.

The effect of these two different types of movements are different on different pixel shape. Basically, linear vibration will require a very accurate moving direction, otherwise it may end up with a different interpolation in one direction which may be worse than another direction. Circular vibration does not require direction correction, which means the interpolation effect is very stable, but might not be the optimum pattern and also gives a more complicated interpolation function.

In addition to the smoothing the gap by vibration, it also reduce the jitter effect caused by the sudden change of gray level between the gap and the neighboring pixels. As shown in Fig. C.1. Here we give an example of linear vibration effect on the hexagonal pixel arrays. We can see that after vibration, the gray level between neighboring pixels is smoothly transferred and filled in the gap. Although the interpolation is not exactly linear, it gives a much smoother transition between neighboring pixels, and more important is, no gap any more.

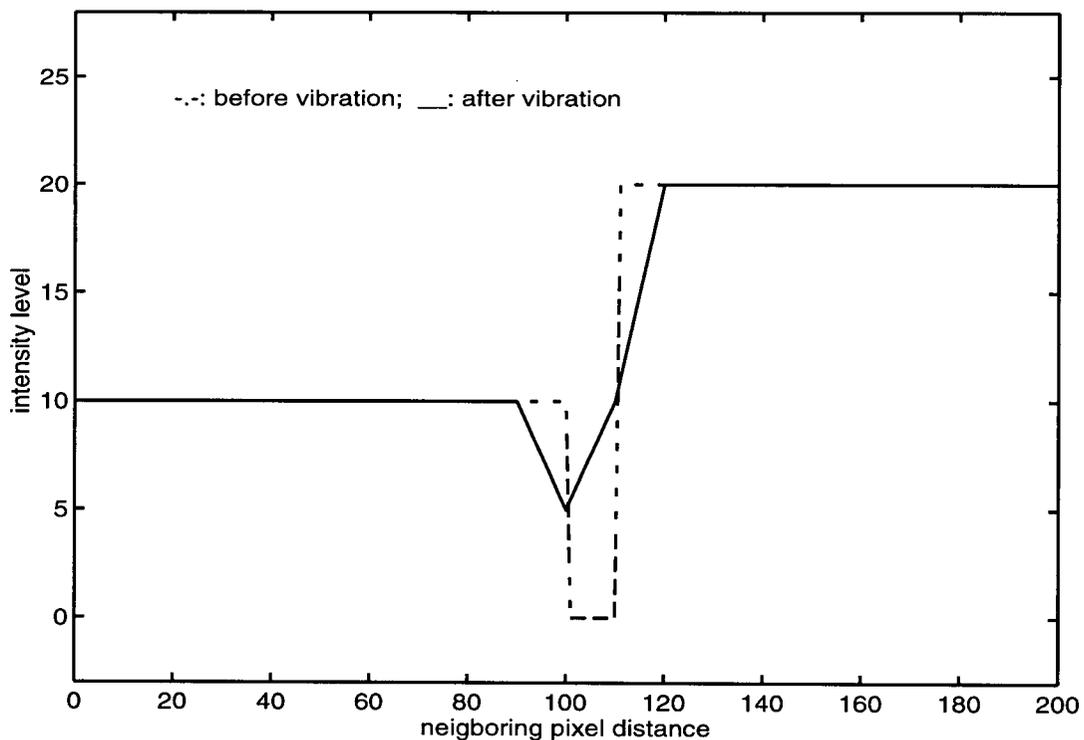


Figure C.1 The Neighboring Pixel Intensity Before/After Vibration

## Appendix D

### Additional Experimental Results

To test the algorithms proposed, 7 set of images are taken with different zoom, different rotation, different space shift, and the algorithms are tested on all of them. The true values of RST are given by visual examination.

Fig. D-1 & D-2 show the comparison of the estimation error of different methods and the precision in graphic form. There is a general trend of decreased estimation error after each additional step, at the cost of the computational time. But, there are also steps where no significant improvement occurred, though computational cost is high. The optimization approach and corner map correlation do not give a stable improvement in general, i.e., the improvement is not consistent over different pictures.

Fig. D-3 shows the number of corner points and its effects on the estimation error. After studying this relation, 5/10 is selected in the final version. That is, 10 corner points are selected by using a Plessey corner detector, and matched to locate to the background, and according to the correlation value, the top five corner pairs with highest correlation values are chosen to give the RST estimation with MSE methods. Intuitively, we would think that the larger the number of corners, the better the estimation results. The reality is there are images with a very limited number of significant corners, and if forced to choose a large number of corners, the threshold has to be lowered to the level where noise becomes significant. The minimum number needed to estimate RST are two corner points. So, any number between 3~5 seems to work fine and give similar results in our experiments.

To choose the final approach, we need to know what precision level can be accepted. From our experiment in fusion, we found that after two level brute force searches, the change is not significant, so this level of estimation error is regarded as acceptable in our experiments. For other applications, there might be different requirement, and the process need to be adjusted accordingly.

Fig. D-4 gives the estimation error of the proposed optimal approach. It shows that whether one level brute force search plus corner matching is used or two level brute force search is used seems to give similar results. The first approach is used because it can recover radial distortion if needed, and slightly less expensive computationally.

Fig. D-5 shows the correlation surface over RS at coarsest level, and Fig. D-6 shows the second coarsest level. These surfaces give us an idea about the variety of the characteristics of the images processed, and that's why caution must be taken in selection of step sizes and methods to suit general situations. These surfaces contain a lot of information about the sensitivity, and the noise level, etc., and can be used as guidance as to whether certain approaches or step sizes are proper or not. The problems is they are obtained after all the calculations are done, and can only be useful as a way to evaluate how reliable the results are. As one can see, surfaces like CS-lab and Tool are well shaped to give good estimation, while surfaces like Map and Toys can be vulnerable to noise. These are reflected in the results.



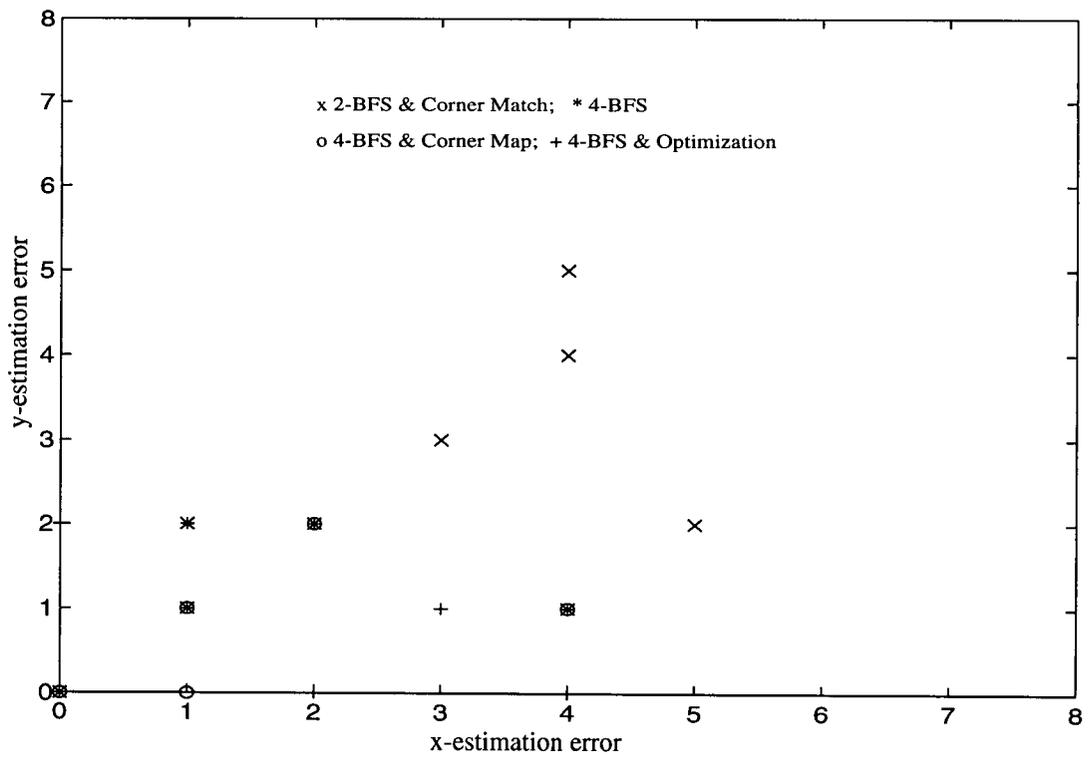
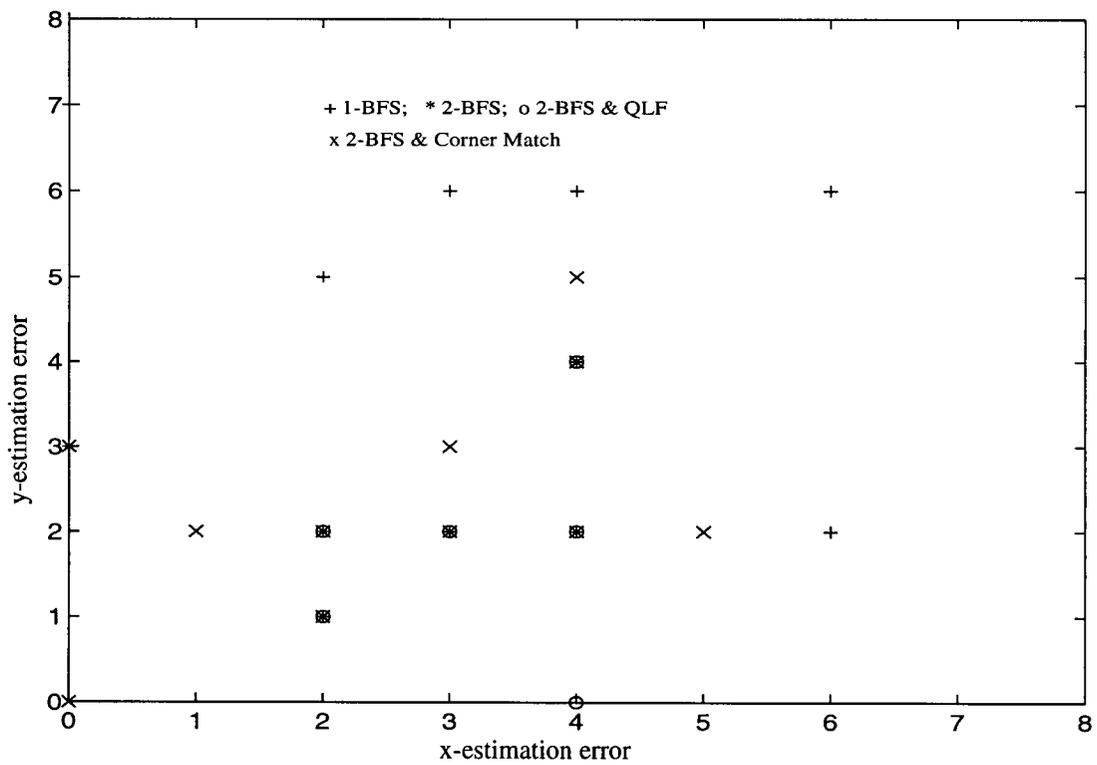
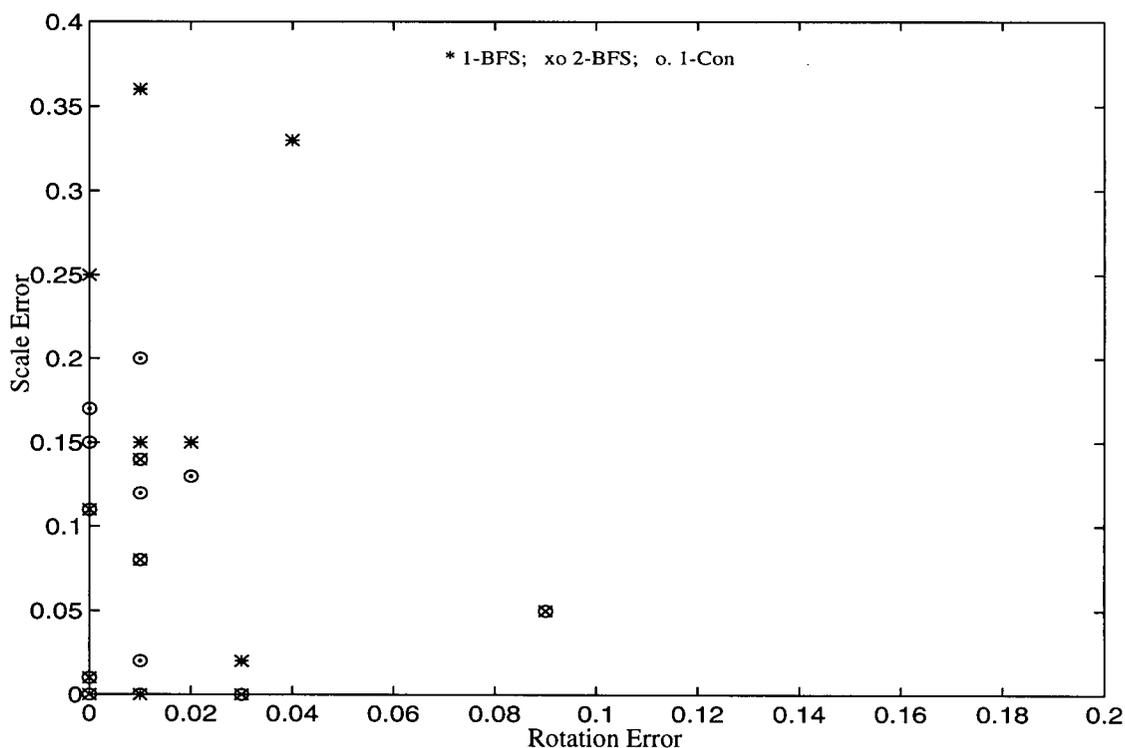


Figure D.2 X/Y Estimation Error Comparison Among Methods



**Figure D.3** Comparison of the Estimation Error of the Steps in Proposed Approach

Figure D7 and up show the errors after each step for each pair of pictures. The true values of RST are obtained by visual test, and non-maximum stands for the selection of the global maxima using non maximum suppression over the correlation surface. QLF stands for quadratic Line Fitting used for sub-step RST value estimation. The four estimation values stands for [R,S,X,Y]. As we can see, the accuracy is generally improving after each step, at extra computational cost, as given in Section 4.4.5. From these tables, we can also see each step has different effects on different pictures. As analyzed before, the success of the registration depend on the feature contents to some extend, and this is reflected in the results. Detailed examination of these results show that corner matching works better on pictures with sharp geometric features such as the Box and Toys pictures, while it fails in the Map picture where no significant features exist. The correlation based method works best in this case.

Images/RST	CS-Lab (R,S,X,Y)
------------	------------------

**Figure D.4** Estimation Error of Methods on the Lab Picture Pairs (Continued) . . .

True Values (obtained by visual test)	(0.00, 4.00, 241, 237)
Estimation Error after 1-BFS with non-max	(0.03, 0.10, 7, 3)
Estimation Error after 2-BFS with non-max	(0.02, 0.07, 3, 1)
Estimation Error after 2-BFS with QLF	(0.01, 0.00, 3, 1)
Estimation Error after 1-BFS & Corner Match	(0.01, 0.08, 0, 0)
Estimation Error after 4-BFS with non-max	(0.00, 0.03, 1, 1)
Estimation Error after 4-BFS & Corner Map	(0.01, 0.00, 1, 1)
Estimation Error after 4-BFS with non-max & Optimization	(0.00, 0.13, 1, 1)

**Figure D.4** Estimation Error of Methods on the Lab Picture Pairs

Images/RST	Toys (R,S,X,Y)
True Values (obtained by visual test)	(0.01, 4.18, 243, 210)
Estimation Error after 1-BFS with non-max	(0.04, 0.28, 5, 2)
Estimation Error after 2-BFS with non-max	(0.01, 0.26, 1, 2)
Estimation Error after 2-BFS with QLF	(0.01, 0.20, 1, 2)
Estimation Error after 1-BFS & Corner Match	(0.01, 0.03, 0, 1)
Estimation Error after 4-BFS with non-max	(0.01, 0.14, 0, 1)
Estimation Error after 4-BFS & Corner Map	(0.01, 0.13, 0, 1)
Estimation Error after 4-BFS with non-max & Optimization	(0.01, 0.07, 2, 1)

**Figure D.5** Comparison of Estimation Error of Methods on the Toys Picture Pairs

Images/RST	Map (R,S,X,Y)
True Values (obtained by visual test)	(0.25, 4.21, 270, 202)
Estimation Error after 1-BFS with non-max	(0.01, 0.15, 2, 6)
Estimation Error after 2-BFS with non-max	(0.02, 0.14, 2, 2)
Estimation Error after 2-BFS with QLF	(0.02, 0.06, 2, 2)
Estimation Error after 1-BFS & Corner Match	(0.02, 0.25, 1, 2)
Estimation Error after 4-BFS with non-max	(0.01, 0.20, 0, 0)
Estimation Error after 4-BFS & Corner Map	(0.00, 0.20, 0, 0)
Estimation Error after 4-BFS with non-max & Optimization	(0.13, 0.20, 52, 38)

**Figure D.6** Comparison of Estimation Error of Methods on the Map Picture Pairs

Images/RST	Tool (R,S,X,Y)
True Values (obtained by visual test)	(0.12, 3.55, 233, 240)
Estimation Error after 1-BFS with non-max	(0.01, 0.15, 7, 0)
Estimation Error after 2-BFS with non-max	(0.02, 0.15, 3, 0)
Estimation Error after 2-BFS with QLF	(0.00, 0.05, 0, 4)
Estimation Error after 1-BFS & Corner Match	(0.01, 0.01, 1, 2)
Estimation Error after 4-BFS with non-max	(0.01, 0.16, 2, 2)
Estimation Error after 4-BFS & Corner Map	(0.03, 0.19, 2, 2)
Estimation Error after 4-BFS with non-max & Optimization	(0.00, 0.15, 1, 3)

**Figure D.7** Comparison of Estimation Error of Methods on the Tool Picture Pairs

Images/RST	Plant (R,S,X,Y)
True Values (obtained by visual test)	(0.13, 3.65, 218, 245)
Estimation Error after 1-BFS with non-max	(0.03, 0.05, 6, 3)
Estimation Error after 2-BFS with non-max	(0.03, 0.13, 2, 3)
Estimation Error after 2-BFS with QLF	(0.01, 0.07, 2, 3)
Estimation Error after 1-BFS & Corner Match	(0.01, 0.14, 0, 0)
Estimation Error after 4-BFS with non-max	(0.02, 0.01, 0, 0)
Estimation Error after 4-BFS & Corner Map	(0.02, 0.00, 0, 0)
Estimation Error after 4-BFS with non-max & Optimization	(0.01, 0.19, 0, 2)

**Figure D.8** Comparison of Estimation Error of Methods on the Plant Picture Pairs

Images/RST	Me (R,S,X,Y)
True Values (obtained by visual test)	(0.10, 4.02, 242, 230)
Estimation Error after 1-BFS with non-max	(0.00, 0.12, 6, 6)
Estimation Error after 2-BFS with non-max	(0.00, 0.05, 2, 2)
Estimation Error after 2-BFS with QLF	(0.01, 0.02, 2, 2)
Estimation Error after 1-BFS & Corner Match	(0.09, 0.05, 1, 2)
Estimation Error after 4-BFS with non-max	(0.01, 0.03, 2, 0)

**Figure D.9** Comparison of Estimation Error of Methods on My Picture Pairs (Continued) . . .

Estimation Error after 4-BFS & Corner Map	(0.01, 0.06, 2, 0)
Estimation Error after 4-BFS with non-max & Optimization	(0.02, 0.11, 2, 0)

**Figure D.9** Comparison of Estimation Error of Methods on My Picture Pairs

Images/RST	Box (R,S,X,Y)
True Values (obtained by visual test)	(0.00, 2.80, 242, 236)
Estimation Error after 1-BFS with non-max	(0.03, 0.10, 6, 4)
Estimation Error after 2-BFS with non-max	(0.01, 0.15, 2, 4)
Estimation Error after 2-BFS with QLF	(0.01, 0.15, 2, 0)
Estimation Error after 1-BFS & Corner Match	(0.01, 0.00, 1, 0)
Estimation Error after 4-BFS with non-max	(0.02, 0.01, 1, 1)
Estimation Error after 4-BFS & Corner Map	(0.00, 0.01, 1, 1)
Estimation Error after 4-BFS with non-max & Optimization	(0.01, 0.00, 2, 1)

**Figure D.10** Comparison of Estimation Error of Methods on the Box Picture Pairs

Images\RST	True Values	Error after 1-BFS with non-max	Error after 2-BFS with non-max
CS-Lab	(0.00, 4.00, 240, 237)	(0.03, 0.10, 8, 0)	(0.02, 0.07, 4, 4)
Toys	(0.01, 4.18, 243, 210)	(0.04, 0.28, 5, 2)	(0.01, 0.26, 1, 2)
Map	(0.25, 4.01, 270, 202)	(0.01, 0.11, 2, 6)	(0.02, 0.09, 2, 2)
Tool	(0.11, 3.50, 233, 240)	(0.01, 0.15, 7, 0)	(0.02, 0.15, 3, 0)
Plants	(0.13, 3.55, 218, 245)	(0.03, 0.15, 6, 3)	(0.03, 0.05, 2, 3)
Me	(0.00, 4.02, 242, 230)	(0.10, 0.12, 6, 6)	(0.10, 0.05, 2, 2)
Box	(0.00, 2.80, 242, 236)	(0.03, 0.10, 6, 4)	(0.01, 0.15, 2, 4)

**Figure D.11** Comparison of Estimation Error Between 1-BFS & 2-BFS on All Pictures

Images\RST	True Values	Error after 1-BFS with non-max	Error after 1-BFS & Corner Match
CS-Lab	(0.00, 4.00, 240, 240)	(0.02, 0.07, 4, 4)	(0.01, 0.08, 0, 0)
Toys	(0.01, 4.18, 243, 210)	(0.01, 0.26, 1, 2)	(0.01, 0.03, 0, 1)
Map	(0.25, 4.01, 270, 202)	(0.02, 0.09, 2, 2)	(0.02, 0.25, 1, 2)
Tool	(0.11, 3.50, 233, 240)	(0.02, 0.15, 3, 0)	(0.01, 0.01, 1, 2)
Plants	(0.13, 3.55, 218, 245)	(0.03, 0.05, 2, 3)	(0.01, 0.14, 0, 0)
Me	(0.00, 4.02, 242, 230)	(0.10, 0.05, 2, 2)	(0.09, 0.05, 1, 2)
Box	(0.00, 2.80, 242, 236)	(0.01, 0.15, 2, 4)	(0.01, 0.00, 1, 0)

**Figure D.12** Comparison of 1-BFS & 1-BFS+Corner Match on All Pictures

Images\RST	True Values	Error after 2-BFS with non-max	Error after 2-BFS & QLF
CS-Lab	(0.00, 4.00, 240, 240)	(0.02, 0.07, 4, 4)	(0.01, 0.00, 4, 4)
Toys	(0.01, 4.18, 243, 210)	(0.01, 0.26, 1, 2)	(0.01, 0.20, 1, 2)
Map	(0.25, 4.01, 270, 202)	(0.02, 0.09, 2, 2)	(0.02, 0.06, 2, 2)
Tool	(0.11, 3.50, 233, 240)	(0.02, 0.15, 3, 0)	(0.00, 0.05, 0, 4)
Plants	(0.13, 3.55, 218, 245)	(0.03, 0.05, 2, 3)	(0.01, 0.03, 2, 3)
Me	(0.00, 4.02, 242, 230)	(0.10, 0.05, 2, 2)	(0.11, 0.02, 2, 2)
Box	(0.00, 2.80, 242, 236)	(0.01, 0.15, 2, 4)	(0.01, 0.15, 2, 4)

**Figure D.13** Comparison of Estimation Error Between 2-BFS with nonmaximum & 2-BFS with QLF on All Pictures

Images\RST	True Values	Error after 4-BFS & Nonmax	Error after 4-BFS & Corner Map
CS-Lab	(0.00, 4.00, 240, 240)	(0.00, 0.03, 1, 4)	(0.01, 0.00, 1, 4)
Toys	(0.01, 4.18, 243, 210)	(0.01, 0.14, 0, 1)	(0.01, 0.13, 0, 1)
Map	(0.25, 4.01, 270, 202)	(0.01, 0.00, 0, 0)	(0.00, 0.00, 0, 0)
Tool	(0.11, 3.50, 233, 240)	(0.01, 0.16, 2, 2)	(0.03, 0.19, 2, 2)

**Figure D.14** Comparison Between 4-BFS & 4-BFS with Corner Map on All Pictures (Continued) . . .

Plants	(0.13, 3.55, 218, 245)	(0.02, 0.01, 0, 0)	(0.02, 0.00, 0, 0)
Me	(0.00, 4.02, 242, 230)	(0.09, 0.03, 2, 0)	(0.09, 0.06, 2, 0)
Box	(0.00, 2.80, 242, 236)	(0.02, 0.01, 1, 1)	(0.00, 0.01, 1, 1)

**Figure D.14** Comparison Between 4-BFS & 4-BFS with Corner Map on All Pictures

Images\RST	True Values	Error after 4-BFS & nonmax	Error after 4-BFS & optimization
CS-Lab	(0.00, 4.00, 240, 240)	(0.00, 0.03, 1, 4)	(0.00, 0.13, 0, 4)
Toys	(0.01, 4.18, 243, 210)	(0.01, 0.14, 0, 1)	(0.01, 0.07, 2, 1)
Map	(0.25, 4.01, 270, 202)	(0.01, 0.00, 0, 0)	(0.13,0.40,52,38)
Tool	(0.11, 3.50, 233, 240)	(0.01, 0.16, 2, 2)	(0.00, 0.15, 1, 3)
Plants	(0.13, 3.55, 218, 245)	(0.02, 0.01, 0, 0)	(0.01, 0.19, 0, 2)
Me	(0.00, 4.02, 242, 230)	(0.09, 0.03, 2, 0)	(0.12, 0.11, 2, 0)
Box	(0.00, 2.80, 242, 236)	(0.02, 0.01, 1, 1)	(0.01, 0.00, 2, 1)

**Figure D.15** Comparison Between 4-BFS & 4-BFS+Optimization on All Pictures

## Bibliography

- [1] Edward H. Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, pp99–106, No. 2, Feb. 1992.
- [2] Dana H. Ballard and Lambert E. Wixson. Object recognition using steerable filters at multiple scales. *Workshop on Qualitative Vision*, 1993.
- [3] Daniel I. Barnea and Harvey F. Silverman. A class of algorithms for fast digital image registration. *IEEE Trans. on Computers*, vol., c-21, No.2, pp179–186, Feb. 1972.
- [4] Fredrik Bergholm. Edge focusing. *IEEE Trans. on PAMI*, pp726–741, No.6, Nov., 1987.
- [5] Valdis Berzins. Accuracy of laplacian edge detectors. *J. of Computer Vision, Graphics, and Image Processing*, Vol. 27, pp195–210, 1984.
- [6] Gerard Blais and Martin D Levine. Registering multiview range data to create 3d computer objects. *IEEE Trans. on PAMI*, Vol. 17, No. 8, pp820–824, 1995.
- [7] Robert M. Boynton and John J. Wisowaty. Equations for chromatic discrimination models. *J. Opt. Soc.*, Vol. 70, No. 12, pp1471–1476, Dec., 1980.
- [8] Denis R. Breglia. Helmet mounted laser projector. *Technical Report, Naval Training Equipment Center, 1989.*
- [9] R.F. Browne and R.M. Hodgson. Pyramid vision applied to automatic inspection. *Third International Conference on Image Processing and its Applications*, 1989.
- [10] Steve Bryson and Scott S. Fisher. Defining, modeling, and measuring system lag in virtual environments. *SPIE Proceedings on Stereoscopic Displays and Applications*, pp98–109, 1990.
- [11] Dick Burbidge and Paul M. Murray. Hardware improvements to the helmet mounted projector on the visual display research tool (vdr), at the naval training system center, pp52–60. *SPIE Proceedings on Helmet-mounted Displays*, 1989.
- [12] James R. Burley. A full color wide-field of view holographic helmet-mounted display for pilot/vehicle interface development and human factors studies. *SPIE Proceedings on Helmet-mounted Displays II*, pp9–15, 1990.

- [13] Peter J. Burt and Raymond J. Kolczynski. Enhanced image capture through fusion. *4th Int'l. Conf. on Computer Vision*, pp173–182, 1993.
- [14] Colin Bussiere and Dimitrios Hatzinakos. Intensity scale invariant motion estimation with rotation and spatial scaling information. *Symposium on Geoscience and Remote Sensing*, pp233–236, 1994.
- [15] John Canny. A computational approach to edge detection. *IEEE Trans. on PAMI*, vol 8, No. 6, pp679–698, Nov. 1986.
- [16] Kenneth R. Castleman. *Digital Image Processing*. Prentice Hall, Upper Saddle River, New Jersey, 1996.
- [17] E.D.E Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE Trans. on PAMI*, Vol. 9, No.5, pp700–703, Sept., 1987.
- [18] J. J. Clark. Authenticating edges produced by zero-crossing algorithms. *IEEE Trans. on PAMI*, vol. 11, No. 1, pp43–57, Jan. 1989.
- [19] Graeme R. Cole and Trevor Hine. Computation of cone contrasts for color vision research. *Behavior Research Methods, Instruments, & Computers*, 24(10), pp22–27, 1992,.
- [20] James Cooper, Svetha Venkatesh, and Leslie Kitchen. Early jump-out detectors. *IEEE Trans. PAMI*, vol. 15, No. 8, pp1377–1382, August 1993.
- [21] D. Crookes and T.J. Brown. Efficient implementation of an abstract programming model for image processing on transputers. *Second Euromicro Workshop on Parallel and Distributed Processing*, January, 1994.
- [22] Ying Cui and Peter D. Lawrence. Detecting scale-space consistent corners based on corner attributes. *IEEE Intl. Conf. on System, Man & Cyberspace*, pp3549–3555, 1995.
- [23] Ying Cui and Peter D. Lawrence. Analysis of registration of images with rotation/scale/translational changes. *Presented at IEEE ACAR Conf.*, 1997.
- [24] McCormick Davies and H W Lau. Microlens arrays in integral photography and optical metrology. *SPIE Proceedings on Miniature and Micro-Optics: Fabrication and System Applications*, Vol. 1544, pp189–198, 1991.
- [25] Rachid Deriche and Thierry Blaszk. Recovering and characterizing image features using an efficient model based approach. *Computer Vision and Pattern Recognition*, pp530–535, 1993.

- [26] Cornelia Fermuller and Walter Kropatsch. Multi-resolution shape description by corners. *Computer Vision and Pattern Recognition*, pp271–276, 1992.
- [27] J. Flussor. Invariant shape description and measure of object similarity. *4th Inter. Conf. on Image Processing and its Applications*, pp139–142, 1992.
- [28] Carl Frederick and Schwartz Eric L. Conformal image warping. *IEEE Computer Graphics & Applications Magazine*, pp54–61, March, 1990.
- [29] Pascal Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *INRIA Technical Report 1369*, 1993.
- [30] A. Goshtasby and G.C. Stockman. A region-based approach to digital image registration with subpixel accuracy. *IEEE Trans. on Geoscience and Remote Sensing*, vol. GE-24, No. 3, pp 390–399, May 1986.
- [31] E. Guelch. Results of test on image matching of isprs. *International Archives of Photogrammetry and Remote Sensing*, 27(III):254–271, 1988.
- [32] Antonio Guiducci. Corner characterization by differential geometry techniques. *Pattern Recognition Letters*, pp311–318, 1988 (8).
- [33] S. Lee Guth. Model for color and light adaptation. *J. Opt. Soc.*, pp976–993, Vol. 8, June 1991.
- [34] M.J. Hannah. Digital stereo image matching techniques. *International Archives of Photogrammetry and Remote Sensing*, 27(III), pp280–293, 1988.
- [35] Chris Harris and Mike Stephens. A combined corner and edge detector. pp147–151, 1988.
- [36] et al Hawryluk. Virtually distortion-free imaging system for large field , high resolution lithography. *US Patent No. 5,176,970*, Jan. 5, 1993.
- [37] E. Hayman, I. D Reid, and D. W. Murry. Zooming while tracking using affine transfer. *British Machine Vision Conference*, 1996.
- [38] Thomas Hebert and Richard Leahy. Map restoration of blurred, photon-limited images via an expectation maximization algorithm. *ICASSP: Acoustics, Speech & Signal Processing Conference*, pp2953–2956, 1991.

- [39] Philippe de Heering and Ezackiel; Wasiljeff Alexander Simmer, Klaus Uwe; Ochieng-Ogolla. A deconvolution algorithm for broadband synthetic aperture data processing. *IEEE Journal of Oceanic Engineering*, Vol. 19, No.1, pp73-83, January, 1994.
- [40] Eric M. Howlett. Wide angle orthostereo. *SPIE Proceedings on Helmet-mounted Displays*, pp2-8, 1989.
- [41] Eric M. Howlett. High resolution inserts in wide-angle head-mounted stereoscopic displays. *SPIE Proceedings on Stereoscopic Displays and Applications*, pp193-203, 1992.
- [42] Eric M. Howlett. Wide angle orthostereo. *SPIE Conference Proceedings*, pp210-220, Feb., 1990.
- [43] Eric M. Howlett. Wide angle color photography method and system. *U.S. Patent No. 4406532*, Sept. 27, 1983.
- [44] Paul S. Idell. Resolution limits for coherent optical imaging: signal-to-noise analysis in the spatial-frequency domain. *J. Opt. Soc.*, pp43-56, 1992.
- [45] Mitsuo Ikeda and Tetsuji Takeichi. Influence of foveal load on the functional visual field. *Perception & Psychophysics*, Vol.18(4), pp255-260, 1975.
- [46] Hans Irtel. Computing data for color-vision modeling. *Research Methods, Instruments, & Computers*, No. 24(3), pp397-401, 1992.
- [47] Qi Jin and Pinfan Yan. A new method of extracting invariants under affine-transform. *11th Int'l. Conf. on Pattern Recognition*, pp742-745, 1992.
- [48] Behrooz Kamgar-Parsi and Azriel Jones, L. Jeffrey; Rosenfeld. Registration of multiple overlapping range images: Scenes without distinctive features. *IEEE Trans. on PAMI*, Vol. 13, No. 9, pp857-870, September, 1991.
- [49] David L. Kanahale. Condor advanced visionics system. *SPIE Proceedings on Display Technology*, pp192-202, 1996.
- [50] M. G. Kaye, Judith Ineson, D.N. Jarrett, and Wickham G. Evaluation of virtual cockpit concepts during simulated missions. *SPIE Proceedings on Helmet-Mounted Displays*, pp236-245, 1990.
- [51] George Kelly. Helmet-mounted area of interests. *SPIE Proceedings on Helmet-mounted Displays III*, pp58-63, 1992.
- [52] James P. Kelly. Textbook of medical physiology. 1956.

- [53] Thomas F. Knoll. Adaptive gray scale mapping to reduce registration noise in difference images. *Computer Vision, Graphics, and Image Processing*, Vol 33, P129–137, 1986.
- [54] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, pp367–375, 1987.
- [55] Th. M. Koller, G. Gerig, G. Szekely, and D. Dettwiler. Multiscale detection of curvilinear structures in 2-d and 3-d image data. *Int'l. Conference on Machine Vision*, pp864–869, 1995.
- [56] Wen-wha Lee, David Wilson, and Lawrence Singerman. Correction of spatial distortion and registration in ophthalmic fluorescein angiography. *Engineering in Medicine & Biology*, pp508–509, 1994.
- [57] Thomas M. Lippert. Fundamental monocular/binocular hmd human factors. *SPIE Proceedings on Helmet-Mounted Displays II*, pp185–191, 1990.
- [58] Geoffrey R. Loftus. On-line eye movement recording systems. *Behavior Research Methods & Instrumentation*, Vol. 7(2), 201–204.
- [59] D. G. Lowe. Organization of smooth image curves at multiple scales. *Inter. J. of Computer Vision*, pp 119–130, 1989.
- [60] Yi Lu and Ramesh C. Jain. Behavior of edges in scale space. *IEEE Trans. on PAMI*, vol. 11, No. 4, pp450–468, April, 1989.
- [61] Stephane Mallat and Sifen Zhong. Characterization of signals from multiscale edges. *IEEE Trans. on PAMI*, vol. 14, No. 7, pp710–726, July 1992.
- [62] R. Manmatha. Image matching under affine deformations. *27th Asilomar Conference on Signals, Systems & Computers*, pp106–110, 1993.
- [63] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt. Multiresolution morphological fusion of mr and ct images of the human brain. *IEE Proceedings on Vision, Image, and Signal Processing*, pp137–142, June, 1994.
- [64] G. F. Mclean. Image warping for calibration and removal of lens distortion. *Communications, Computers and Signal Processing*, pp170–173, 1993.
- [65] James E. Melzer and W. Moffitt Kirk. Color helmet display for the tactical environment: the pilot's chromatic perspective. *SPIE Proceedings on Helmet-mounted Displays III*, pp47–51, 1992.

- [66] Charles Meyer, Gregg S. Leichtman, James A. Brunberg, Richard L. Wahl, and Leslie E. Quint. Simultaneous usage of homologous points, lines, and planes for optimal, 3-d, linear registration of multimodality imaging data. *IEEE Trans. on Medical Imaging, Vol. 14, No.1, pp1-11*, March 1993.
- [67] Jacqueline Le Moigne. Towards a parallel registration of multiple resolution remote sensing data. *Symposium on Geoscience and Remote Sensing, pp1011-1013*, 1995.
- [68] Vernon B. Mountcastle. *Medical physiology*. 1974.
- [69] Don Murray and Anup Basu. Motion tracking with active camera. *IEEE Trans. on PAMI, Vol. 16, No. 5, pp449-459*, May, 1994.
- [70] Shree K. Nayar. Catadioptric omnidirectional cameras. *Technical Report, Dept. of Computer Science, Columbia University*, October, 1996.
- [71] Scott A. Nelson and Cox J. Allen. Quantitative helmet mounted display system image quality model. *Helmet-mounted Displays, pp128-137*, 1992.
- [72] Jacek Nieweglowski, T. George Campbell, and Petri Haavisto. A novel video coding scheme based on temporal prediction using digital image warping. *IEEE Trans. on Computer Electronics, pp141-150*, Aug., 1993.
- [73] J Alison Noble. Finding corners. *Image & Vision Computing, Vol 6, pp121-128*, May, 1988.
- [74] Finbarr O'Sullivan and Maijian Qian. A regularized contrast statistic for object boundary estimation — implementation and statistical evaluation. *IEEE Trans. on PAMI, Vol. 16, No. 6, pp561-570*, June, 1994.
- [75] A. Papoulis. *Probability, random variables and stochastic processes*. New York: McGraw-Hill, 1965.
- [76] L. Peichl and H. Wassle. Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *J. Physio., pp117-141*, 1979.
- [77] Rajesh P.N. Rao and Dana H. Ballard. An active vision architecture based on iconic representations. *Technical Report, Department of Computer Science, University of Rochester*, 1996.

- [78] Anothai Rattarangsi and Roland T. Chin. Scale-based detection of corners of planar curves. *IEEE Trans. on PAMI*, Vol. 14, No. 4, pp430–449, April, 1992.
- [79] Robert K. Rebo and Amburn Phil. A helmet-mounted virtual environment display system. *SPIE Proceedings on Helmet-mounted Displays*, pp80–84, 1989.
- [80] Nader Riahi and Lawrence Peter D. A retinal image processing system. *SPIE Proceedings on Automated Inspection and High Speed Vision Architecture II*, pp17–24, Nov., 1988.
- [81] Paul L. Rosin. Augmenting corner descriptor. *Graphical Models and Image Processing*, vol 58, No.3, pp286–294, 1996.
- [82] W. L. Sachtler and Qasim Zaidi. Chromatic and luminance signals in visual memory. *J. Opt. Soc.*, Vol. 9, pp877–894, June 1992.
- [83] Philippe Saint-Marc and Jer-Sen Chen. Adaptive smoothing: A general tool for early vision. *IEEE Trans. on PAMI*, vol. 13, No.6, pp514–529, June 1991.
- [84] J. Santamaria, P. Artal, and J. Bescos. Determination of the point-spread function of human eyes using a hybrid optical-digital method. *J. Opt. Soc.*, Vol. 4, pp1109–1114, No. 6, June 1987.
- [85] C.E. Savin, M.O. Ahmad, and M.N.S. Swamy. Design of weighted order statistic filters using linearly separable stack-like architecture. *Midwest Symposium, Circuits and Systems*, pp753–756, 1994.
- [86] Cordelia Schmid and Roger Mohr. Matching by local invariants. *INRIA Report, N2644*, Aug. 1995.
- [87] Thorsten Schormann, Andreas Dabringhaus, and Zilles Karl. Statistic of deformations in histology and application to improved alignment with mri. *IEEE Trans. on Medical Imaging*, Vol. 14, No. 1, pp25–35, March, 1995.
- [88] A.M. Seibert, M; Waxman. Adaptive 3-d object recognition from multiple views. *IEEE Trans. on PAMI*, Vol. 14, Feb. 1992.
- [89] Terrence J. Sejnowski and Lehky Sidney R. Neural network models of visual processing. *Society for Neuroscience Short Course on Computational Neuroscience*, pp1–28, 1987.

- [90] Eero P. Simoncelli and William T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. *2nd Annual IEEE Intl' Conf. on Image Processing*, pp1-4, October, 1995.
- [91] Steven Skaggs, Jason Gerber, Griff Bilbro, and Michael B. Steer. Parameter extraction of microwave transistors using a hybrid gradient descent and tree annealing approach. *IEEE Trans. on Microwave Theory and Techniques*, pp726-729, 1993.
- [92] Anne H. Solberg, Anil K. Jain, and Torfinn Taxt. Multisource classification of remotely sensed data: Fusion of landsat tm and sar images. *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 32, No. 4, pp768-778, July, 1994.
- [93] L. J. Spreeuwens and F. van der Heijden. Evaluation of edge detectors using average risk. *Pattern Recognition, 11th Int'l. conf.*, pp771-774, 1992.
- [94] Bjorn Stabell and Ulf Stabell. Spectral sensitivity in the far peripheral retina. *J. Opt. Soc.*, Vol. 70, No. 8, pp959-963, August 1980.
- [95] Roy H. Steinberg, Miriam Reid, and Paula L. Lacy. The distribution of rods and cones in the retina of the cat. *J. Comp. Neur.*, 148: pp229-240, 1972.
- [96] George Stockman, Kopstein, and Sanford Benett. Matching images to models for registration and object vis clustering. *IEEE Trans. on PAMI*, Vol. 4, No. 3, pp229-241, May, 1982.
- [97] Cho-Huak Teh and Roland T Chin. On the detection of dominant points on digital curves. *IEEE Trans. on PAMI*, vol. 11, No. 8, Aug. 1989.
- [98] Jonathan Vaughan. On-line, real-time recording of eye orientation using the corneal reflection method. *Behavior Research Methods & Instrumentation*, Vol. 7(2), pp211-214, 1974.
- [99] John A. Webster. Stereoscopic full field of vision display system to produce total visual telepresence. *SPIE Proceedings on Display System Optics*, pp63-70, 1989.
- [100] Christopher D. Wickens. Three-dimensional stereoscopic display implementation: Guidelines derived from human visual capabilities. *SPIE Proceedings on Stereoscopic Displays and Applications*, pp2-10, 1990.
- [101] C. M. Wu, R. M. Owens, and M. J. Irwin. Distortion processing in image matching problems. *ICASSP: Acoustics, Speech & Signal Processing*, pp2181-2184, 1990.

- [102]Weishi Xia, Robert M. Lewitt, and Paul R. Edholm. Fourier correction for spatially variant collimator blurring in spect. *IEEE Trans. on Medical Imaging*, Vol. 14, No.1, March, 1995.
- [103]Yi-Sheng Yao and Rama Chellappa. Tracking a dynamic set of feature points. *IEEE Trans. on Image Processing*, Vol.4, No.10, pp1382–1396, 1995.
- [104]Laurence R. Young and Sheena. Methods & designs. *Behavior Research Methods & Instrumentation*, Vol. 7(5), pp397–429, 1975.
- [105]A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossing. *IEEE Trans. PAMI*, vol. 8, pp15–25, June 1986.
- [106]Qinfen Zheng and Rama Chellappa. A computational vision approach to image registration. *IEEE Trans. on Image Processing*, Vol. 2, No.3, pp311–326, July 1993.
- [107]Y.T. Zhou. Multi-sensor image fusion. *IEEE Intl. Conf. on Image Processing*, pp193–197, 1994.