A thesis submitted for the award of the degree of Doctor of Philosophy

Geometric techniques in multiple view point correspondences

By

D. D. A. P. Ariyawansa

Department of Electrical, Electronic, and Information Engineering
City University
London

April, 1999

Acknowledgements

i

Declaration

"I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement"

Abstract

Estimation of the size, shape, and location of an object(s) using images acquired from multiple view points is required in many close range measurement and computer vision tasks. The first step in this process is the extraction of features in the images. In the second step, the features in images that correspond to the same *3-D* points are identified. This identification process is termed the establishment of correspondences.

In this thesis a geometric technique, based on image rectification principles, for the establishment of multiple view point correspondences is developed. The technique uses parallel epipolar lines in rectified image space hence, it is termed the *PEL* algorithm. This technique can establish faster image point correspondences than the usual epipolar line or *3-D* space intersection techniques. This is due to avoiding the non-linear search. An initialisation process, which is not required by the epipolar line or the *3-D* space intersection methods, is required for the *PEL* algorithm to pre-estimate rectification and search band window parameters. Hence, the speed performance of the technique is better when the number of points is larger than a certain value which is depending upon the software implementation and the speed of the computer used. The technique has been rigorously tested with both simulated and real-image data under various situations.

Another aspect of the work conducted for this thesis contributed to the development of a real-time multiple camera *3-D* measurement system. In conventional frame grabber based multiple camera systems sequential image data acquisition is a time consuming task. In order to achieve frame rate functionality, dedicated hardware based on a *Digital Signal Processor* for real-time video signal processing and data communications were developed. The dedicated hardware plus a *CCD* camera is termed an intelligent camera. In a multiple camera system each intelligent camera processes images simultaneously and the processed data are transferred to a central computing unit for further processing. This system provided the context for the development of the *PEL* algorithm. The real-time performances of the system was analysed and the *PEL* algorithm was tested using the data acquired from multiple intelligent camera system.

Contents

Chapter 4   Multiple view point correspondences using rectification --- *PEL* algorithm

Chapter 5   Simulation tests

Chapter 6   Development of a real-time close-range photogrammetric system

Chapter 7 Experiments with real-image data

Chapter 8   Achievements, suggestions for further work, and summary

Chapter 1

Introduction

"The correspondence problem"

In close-range photogrammetric and computer vision tasks, the estimation of *3-D* co-ordinates (precisely or coarsely) using *2-D* image information is often required. A vast number of applications that require range information may be imagined. It may be an optical triangulation system for pipe inspection (Gooch e*t al*, 1996), a space robot tracking and locating various objects (Shapiro and Rosenfeld, 1992), or a measurement system guiding a robot for assembly of aircraft wing components (Clarke *et al*, 1998). Active methods of acquiring range information may be used in some applications where a single camera and a point laser source may be sufficient. An image point formed by the reflected laser beam may be conveniently identified by setting appropriate illumination levels. Hence, the correspondence between the laser source and its image is automatically known. Depth information can be computed given the distance between the laser source and the camera. In passive methods, images of the objects are acquired from two or more view points. Active illumination may be used to obtain images of objects with high contrast. Techniques are required to identify features in different views that correspond to the same object feature hence, the "correspondence problem".

1.1 Introduction to correspondence techniques

The correspondence problem has been researched extensively by computer vision researchers. Hundreds of research papers have appeared in various conferences and journals over the years. Some research concentrated developing geometric techniques (Longuet-Higgins, 1981; Faugeras, 1993). Certain others developed methods involving both geometric and radiometric techniques (Kanade *et al*, 1999). In photogrammetric community less emphasis has been placed on this topic and few researchers addressed this topic alone (Gruen and Baltsavias, 1988). In many algorithms geometric techniques are used to complement radiometric techniques. In certain applications, only the geometric techniques are used to establish correspondences (Mass, 1992). The technique applied to a particular problem depends

on the scene content, the imaging geometry, and the required precision of range estimations. After an extensive literature survey, the author could not find an universal algorithm that is capable of performing equally well in every situation.

1.2 Introduction to geometric techniques

A typical situation is a pair of images acquired from two view points. The features of interest may be extracted in one image. In order to find the corresponding features, the intensity information in a small area (window) surrounding the extracted feature in one image can be matched with the other image until a matching area is found. This is a computationally expensive search procedure. If the relative or absolute orientation of images are known, the geometric relationship between two images (*i.e.* epipolar geometry) can be used to constrain the search and to establish reliable correspondences. Hence, the geometric information is as important as the radiometric information. In a multiple view configuration, the role of geometric information become more important as far as the computational efficiency and reliability of correspondences are concerned.

The epipolar geometry is a key projective geometric relationship used in most correspondence algorithms. Considering a stereo pair, if the relative or the absolute orientation of cameras are known the epipolar constraint can be used to find correspondences or potential correspondences (Haralick and Shapiro, 1993; Atkinson, 1996; Mass, 1992). Figure 1.1 illustrates a convergent axes stereo pair where projections from a *3-D* point $A_i(X_i, Y_i, Z_i)$ on images I_1 and I_2 are at points $p_{1i}(x_{1i}, y_{1i})$ and $p_{2i}(x_{2i}, y_{2i})$. The plane formed by points A_i, O_1, and O_2, termed epipolar plane, intersects image planes along epipolar lines l_{1i} and l_{2i}. Hence, for a selected point on l_{1i}, the corresponding point should ideally fall on the conjugate epipolar line l_{2i}. The points $e_1(x_{e1}, y_{e1})$ and $e_2(x_{e2}, y_{e2})$ are termed epipoles and are common points for epipolar lines in each image.

In some algorithms that use radiometric information, a point is selected in one image and a window surrounding the point is determined. This intensity window is slid along the conjugate epipolar line in the other image until a match is found hence, constraining the search. The criteria for a match may be based on correlation,

minimum intensity differences, least squares or other techniques (Kanade *et al*, 1999; Baltsavias, 1991; Ruther, 1996).
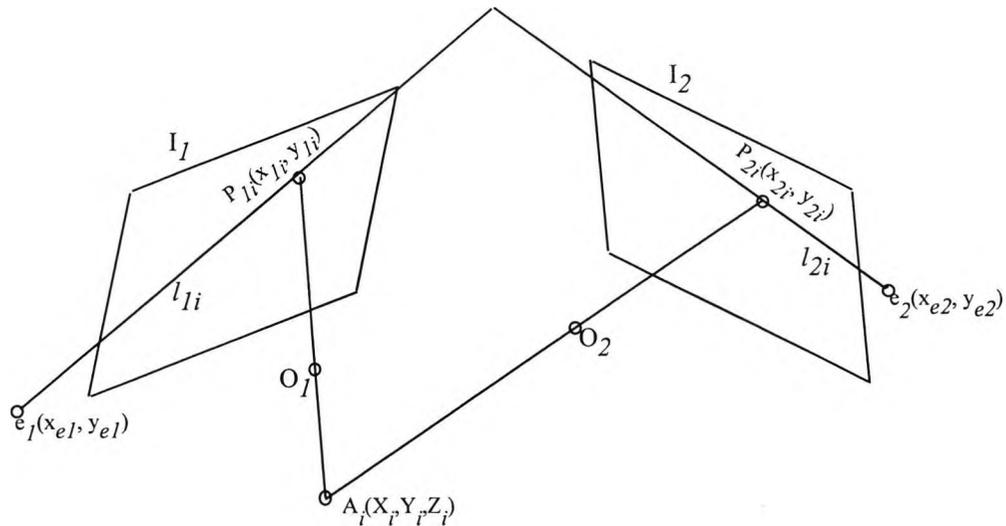
Figure 1.1 Epipolar geometry

In some computer vision applications the camera exterior orientation is unknown and has to be determined on-line using the acquired image data. In the past few years, there has been an interest in developing novel geometric techniques for the establishment of point correspondences when the exterior orientation are not known. The "essential matrix" due to Longet-Higgins, 1981 can be used to map camera co-ordinates of points in one image to the conjugate epipolar lines in another image. Hence, camera interior parameters are required to convert pixel co-ordinates to camera co-ordinates. Another development due to Luong and Faugeras, 1996 is the "fundamental matrix" which can be used to map pixel co-ordinates in one image to conjugate epipolar lines in another image. The camera interior parameters are not required when using this technique since they are included in the fundamental matrix. The use of this matrix to establish correspondences between *n* views is still under investigation (Faugeras and Robert, 1996).

In most high precision close range photogrammetric applications, either the relative or the absolute orientation of the cameras are estimated using control points. The camera interior parameters are estimated off-line. However due to the nature of the high

precision requirements artificial targets are often placed on object points of interest. The illumination and the cameras are set up so that only the target images appear in each view. These target images are located to subpixel accuracy. Hence, for the establishment of correspondences, accurately located image points are the only available information apart from the camera exterior and interior parameters. As a result the geometric techniques are the only means that can be used for the establishment of correspondences. The epipolar line *(EL)*, *3-D* space intersection *(3DSI)*, and back projection are the mostly used techniques in high precision close range applications (Fraser, 1997; Mass, 1992; Chen, 1995). The definitions of *EL* and *3DSI* algorithms used for this thesis are given in sections 3.4.1 and 3.4.2.

1.3 Introduction to *PEL* algorithm

In this thesis, a technique called *PEL* algorithm is developed. It has been successfully implemented for the establishment of multiple view point correspondences in high precision close-range measurement applications. The algorithm may also be used in other photogrammetric or computer vision applications.

The distinguishing feature of this technique is that it can establish faster point correspondences than the *EL* or *3DSI* methods. The algorithm exploits an special instance of epipolar geometry. If the *x*-axis of the local co-ordinate systems of an stereo pair are parallel to the base line and the image planes are co-planar, the conjugate epipolar lines become collinear. Hence, there is no computational effort required to determine conjugate epipolar lines. Figure 1.2 illustrates a pair of parallel axes images I_{r1}, I_{r2}. The projections of a *3-D* point $A_i(X_i, Y_i, Z_i)$ are at points $p_{r1i}(x_{r1i}, y_{r1i})$ and $p_{r2i}(x_{r2i}, y_{r2i})$ which fall on collinear epipolar lines l_{r1i} and l_{r2i}. Hence, $y_{r1i} = y_{r2i}$.

In practice, conjugate epipolar lines are not collinear due to errors in image point location and convergent to rectified image transformations. Hence, a search band that covers either side of the ideal collinear epipolar line is required to identify correspondences. In this algorithm, a novel method is introduced to determine this search band using the knowledge of search bands in convergent images.

The algorithm considers images taken from multiple view points as a set stereo pairs. Stereo pair correspondences are established as discussed above. Correspondences are then distributed among stereo pairs using a technique based on epipolar geometry and image point identity matching.
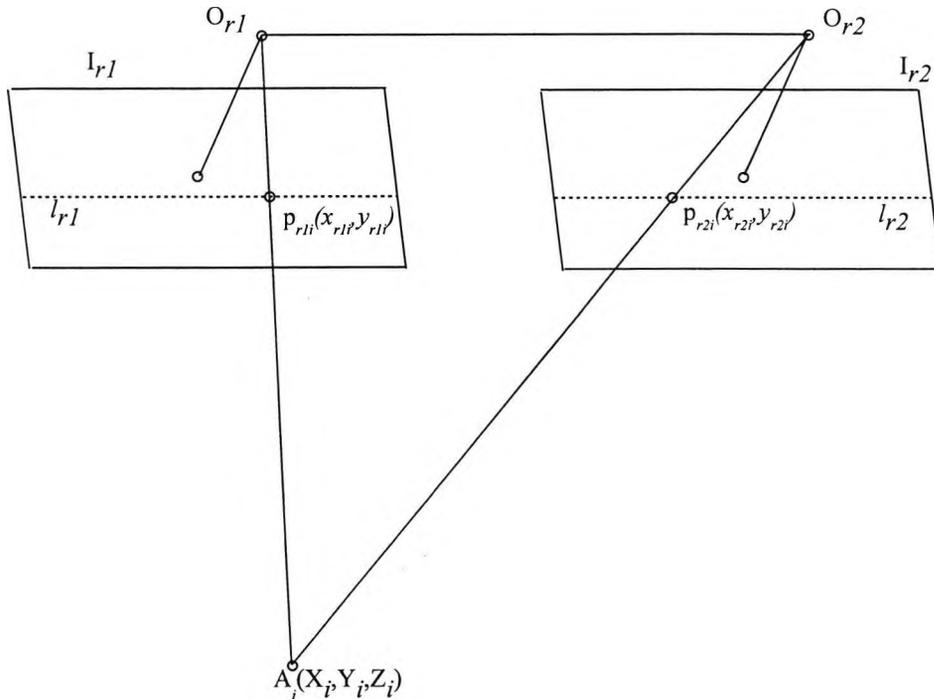
Figure 1.2 Parallel axes stereo pair and collinear epipolar lines

In close range systems, cameras are positioned in a convergent configuration to obtain an optimum network design (Mason, 1994). The *PEL* algorithm considers multiple views as a combination of stereo pairs. For each stereo pair, a pair of rectified image planes is determined and the points of interest that are located in each convergent image are transformed to rectified image planes. Provided that these transformed image points are sorted in an ascending order of the y-co-ordinate in rectified image space, an efficient search can be performed to find potential correspondences. However, the ideal geometric condition does not exist with real imaging systems. The corresponding points may not fall on collinear epipolar lines. One reason is the transformed image point location errors from the convergent images. The other is the transformation errors due to the errors in camera exterior orientation. Hence, a search band is required to identify corresponding points. A technique is introduced to

determine the search band in a rectified image pair and is specially required for the establishment of reliable multiple view point correspondences. This technique uses knowledge of the characteristics of search bands in convergent images.

If the correspondences are to be established in the convergent image space, for a selected point in one image all the points in the other image should be checked to determine if they are corresponding. If n points are available in each image, it requires n^2 search which becomes computationally expensive when the number of points is large. Both the EL and $3DSI$ methods require n^2 search. The advantage with the PEL algorithm is that the search is approximately proportional to n. However, the algorithm is efficient only when n is greater than a certain number. This is due to the time required for initialisation stage which is dependent upon the implementation. With the current implementation, using a $200MHz$ Pentium PC the PEL algorithm in a four view configuration is faster than the EL and $3DSI$ methods when number of points is greater than about 75. Table 1.1 illustrates some results from the timing tests.

No of points	PEL (ms)	EL (ms)	3DSI (ms)
10	6.68	0.63	1.21
30	7.75	2.48	9.5
50	9.17	5.49	27.84
75	10.86	10.88	58.35
100	12.67	18.05	104.66
150	17.63	37.44	232.94
200	27.41	65.16	414.94
250	35.25	104.39	652.11
300	44.51	148.98	940.48
350	50.79	231.18	1305.78
400	61.24	351.81	1771.31

Table 1.1 Timing comparisons

1.4 The real-time system development

Usually with frame grabber based close range measurement systems, images are sequentially acquired from multiple cameras and are then processed in a PC or a workstation. The work involves extraction of target image locations in each image,

establishment of correspondences, and the *3-D* location estimation. The acquisition of images and the extraction of features is a sequential process which is time consuming. In order to overcome processing speed limitations, a real-time system was developed. In this system each camera has local processing and communication facilities which perform the target image location and the data transfer to a *PC* in real-time. Images are processed in parallel for feature extraction hence, sequential processing of images in the *PC* is avoided. Due to the transfer of processed data the amount of data transferred is minimised. A *CCD* camera with local processing and communication facilities is termed an intelligent camera which is a modular system with the processing power of a Digital Signal Processor (*DSP*) and Ethernet data communications. Various tests were carried out to assess the speed and the accuracy performances of the system. An intelligent camera can repeatedly produce about *170* target image locations every *40ms* (Clarke *et al*, 1998).

1.5 Structure of the thesis

In chapter 2 the fundamentals of close-range photogrammetry are reviewed. Discussion of random errors of measurements and the least squares technique are introduced. The camera model, collinearity equations, and camera interior parameters are discussed. The direct linear transformation for resection, the estimation of relative orientation, and the rigorous and non-rigorous intersection techniques are reviewed. The chapter concludes with an introduction to the bundle adjustment.

Chapter 3 is devoted to the discussion of multiple view point correspondence techniques. The details on the survey of image matching methods are discussed and the methods are classified. Image feature extraction methods that are used in high precision close-range applications are discussed. The geometric techniques used for establishing correspondences are discussed in detail as are point correspondence techniques used in close-range photogrammetry and their characteristics. Finally, the techniques used in uncalibrated computer vision applications are introduced.

Chapter 4 discusses the development of a technique based on image rectification for multiple view point correspondence establishment - the *PEL* algorithm. A case for development of this technique is discussed. The theoretical development of each stage

of algorithm is presented in detail. The issues involved in multiple view implementation are discussed and the speed performance of the algorithm was compared with that of the others.

Chapter 5 is devoted to the simulation tests. Description of the developed software simulation system is given. The effect of the errors in camera exterior parameters and the *2-D* image point locations on the epipolar constraint are discussed in detail with illustrations of simulation tests that were carried out with a network of four cameras. The performance of the *PEL* algorithm in multiple view configurations is illustrated in detail with simulation test results

Chapter 6 discusses the development of a real-time close-range photogrammetric measurement system. The system, based on intelligent cameras, exploits distributed processing of video data. The algorithm used for object processing is discussed. The development of dedicated hardware and software are presented in detail.

Chapter 7 presents results from practical experiments. First, the setting up of the real-time system, the camera calibration using the data acquired from the real-time system, and the *2-D* target image location accuracy performance are discussed. Real-time four view point correspondence establishment test results are presented. Finally results on the establishment of correspondences using frame grabber acquired images are illustrated.

Chapter 8 presents the achievements of the work of thesis, suggestions for further work, and the summary.

1.6 Aims and objectives of the thesis

The principal aims and objectives of the work described in this thesis can be categorised into three groups as follows:

1. Development of intelligent camera system with real-time video processing and Ethernet data communication capabilities: Each intelligent camera in a multiple view system transfers extracted target image location data to a Central Processing

Computer *(CPC)* for the establishment of correspondences and *3-D* location estimations. With respect to the speed performance, intelligent cameras were expected to be able to process over *100* target images of size *6x6* pixels (approximately) and transfer their subpixel locations to the *CPC* within a video frame period *40ms*. It was expected that *CCIR* complying *CCD* cameras were to be used. Target image location accuracy of over *1/10*th of a pixel was expected which is comparable to the accuracy performance of frame grabbers.

2. Development of an algorithm for the establishment of real-time multiple view target image correspondences: This algorithm receives the extracted target image locations from intelligent cameras as the input, establish multiple view correspondences, and produce details of the established correspondences in a suitable structure so that *3-D* location estimation algorithm can access the data efficiently. This algorithm was expected to have better speed performance than *EL* and *3DSI* techniques.

3. Integration of intelligent cameras and correspondence solution with *3-D* location estimation module to form a real-time *3-D* measurement system: This work was expected to make available target image location data flow from multiple intelligent cameras to a *CPC* and the establishment of correspondences. The *3-D* location estimation and graphical user interface were expected to be provided as a result of another project.

Chapter 2

Principles of close-range photogrammetry

Photogrammetry is a technique for obtaining information about the position, size, and shape of object(s) by measuring images instead of measuring objects directly. The term close-range photogrammetry describes the technique when the extent of the object to be measured is less than about *100* meters and the cameras are close to the object (Atkinson, 1996). In close-range applications, where high precision measurements are required, cameras are usually positioned in a convergent configuration to obtain the optimum measurement accuracy in all three dimensions (Mason, 1994; Atkinson, 1996). Redundant measurements of the object space points are made using multiple cameras to enable better results and greater statistical feedback. This chapter discusses the fundamentals of the close-range measurement process and the various algorithms that are used.

2.1 Errors in close-range measurements and least squares

When a measurement contains an error, it differs from its true value. A number of types of errors can be produced by a close-range photogrammetric measurement system. These can be categorised as blunders, constant, systematic, or random errors (Slama, 1980). Blunders are errors caused due to mistakes made during the measurement process and are unpredictable. Wrong multiple view image point correspondences is an example of a blunder and carefully planned procedures may be required to remove them. The constant errors, may be due to errors that has the same magnitude and sign. Constant errors can be detected and corrected by proper calibration of the measurement system. Systematic errors follow a rule hence, they can be estimated and measurements can be corrected. The errors due to the camera lens distortion is an example of a systematic error. Random errors do not follow any rule hence, they are dealt with according to the laws of probability.

In close-range applications, redundant measurements of object points are obtained indirectly using images acquired from multiple views. The measurements are corrected for systematic errors. The least squares technique is then used to obtain the most

probable values for the indirectly measured unknown quantities. It is assumed that image measurements contain only random errors which follow a normal distribution. It is also assumed that the measurements are mutually independent.

Finding the most probable values for the unknowns based on a set of image measurements is the same as finding the most probable set of random errors. If $(v_1, v_2, \ldots, v_i, \ldots, v_n)$ is a set of random errors of a set of image measurements where each error has a normal distribution, the density function of error v_i can be stated as (Slama, 1980),

$$f(v_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{v_i}{\sigma_i}\right)^2} \tag{2.1}$$

where σ_i is the standard deviation of error v_i. Since image measurements are mutually independent of each other, their combined or joint density function is the product of individual density functions which can be stated as,

$$f(v_1, v_2, \ldots, v_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\left(\frac{1}{\sigma_1}\right) \cdot \left(\frac{1}{\sigma_2}\right) \cdots \cdots \left(\frac{1}{\sigma_n}\right)\right) e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{v_i}{\sigma_i}\right)^2} \tag{2.2}$$

Hence, the joint probability density function can be stated as,

$$P\left((v_1 - \Delta v_1 \leq v_1 \leq v_1 + \Delta v_1), (v_2 - \Delta v_2 \leq v_2 \leq v_2 + \Delta v_2), \ldots, (v_n - \Delta v_n \leq v_n \leq v_n + \Delta v_n)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\left(\frac{1}{\sigma_1}\right) \cdot \left(\frac{1}{\sigma_2}\right) \cdots \cdots \left(\frac{1}{\sigma_n}\right)\right) e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{v_i}{\sigma_i}\right)^2} dv_1 \cdot dv_2 \ldots \ldots dv_n$$

$$\tag{2.3}$$

The joint probability will be maximum for the most probable set of errors $(v_1, v_2, ..., v_n)$. According to the Equation (2.3), the joint probability is maximum when the term $\sum_{i=1}^{n} \left(\frac{v_i}{\sigma_i}\right)^2$ is minimum. Hence,

$$\sum_{i=1}^{n} \left(\frac{v_i}{\sigma_i}\right)^2 \rightarrow minimum \qquad (2.4)$$

In order to find the most probable value for the unknown quantity, the sum of the squares of errors or residuals should be minimised and this is termed least squares estimation.

If σ_0 is a constant, the following solution will still satisfy the least squares criteria,

$$\sum_{i=1}^{n} \left(\frac{\sigma_0 v_i}{\sigma_i}\right)^2 \rightarrow minimum$$

By substituting $W_i = \left(\frac{\sigma_0}{\sigma_i}\right)^2$, the least squares condition can be stated as,

$$\sum_{i=1}^{n} W_i v_i^2 \rightarrow minimum \qquad (2.5)$$

W_i is termed as the weight of the measurement and σ_0 is termed the variance of a measurement of unit weight (Slama, 1980). Hence, the least squares criteria can be stated as,

$$\sum_{i=1}^{m} W(v_i)^2 = W_1(v_1)^2 + W_2(v_2)^2 + ...+ W_n(v_n)^2 \rightarrow minimum$$

In matrix form, this condition can be stated as,

$$\phi = (v^t W v) \rightarrow \textbf{min}imum \qquad (2.6)$$

where v is a column matrix containing random errors and v^t is its transpose.

In close-range applications, measurements of the same object space quantity is obtained using multiple views. For each measurement, collinearity equations (Equations 2.14) can be written that represent the relationship between the unknowns and the measured quantities. A set of n equations thus obtained is termed the functional model and can be stated in general form as,

$$F = f(x) - b = 0 \qquad\qquad (2.7)$$

where,

$$x = (x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_m) \text{ are } m \text{ unknowns and,}$$

$$b = (b_1, b_2, b_3, \ldots\ldots\ldots\ldots, b_n) \text{ are } n \text{ measurements.}$$

The functional model based on collinearity equations is non-linear. Hence, the first step is the linearization of the functional model using Taylor's theorem. In general, a linearized functional model to the first order accuracy can be stated as,

$$f(x_0) - b_0 + A\,dx + D\,db = 0 \qquad\qquad (2.8)$$

where,

$f(x_0)$ represents a matrix consisting evaluations of $f(x)$ at initial values,

b_0 represents matrix consisting measured values,

A represents a matrix consisting partial differentials of the functional model with respect to unknowns evaluated at their initial values,

dx represents a matrix consisting corrections to the unknowns,

D represents an identity matrix and,

db represents a matrix consisting corrections to the measured values which are residuals.

In matrix form the functional model can be stated as,

$$A\Delta x = b + v \qquad\qquad (2.9)$$

where, $b = b_0 - f(x_0)$ and $v = -db$.

Substituting for v in Equation 2.6, the least squares condition can be stated as,

$$\phi = (A\Delta x - b)^t W (A\Delta x - b) \to minimum.$$

$$\phi = \left((A\Delta x)^t - b^t\right) W (A\Delta x - b)$$

$$\phi = \left((A\Delta x)^t (A\Delta x)W - (A\Delta x)^t bW - b^t W(A\Delta x) + b^t Wb\right)$$

ϕ is a function of Δx. Hence, a minimum of ϕ occurs when,

$$\frac{d\phi}{d\Delta x} = \left((A\Delta x)^t WA - b^t WA\right) = 0$$

$$\left((A\Delta x)^t WA - b^t WA\right) = 0$$

Hence, a matrix Δx containing corrections to unknowns can be estimated as,

$$\Delta x = \left(A^t WA\right)^{-1} A^t Wb \qquad\qquad (2.10)$$

The initial values of the unknowns (x_0) adjusted with the calculated corrections (Δx) can be stated as,

$$x = x_0 + \Delta x \qquad\qquad (2.11)$$

$\left(A^t WA\right)^{-1}$ is the cofactor matrix. Given an *a priori* value for the variance of a measurement of unit weight, the cofactor matrix can be used to obtain variances of the estimated unknowns.

2.2 Camera model

The camera model defines the relationship between *3-D* object points and corresponding image points. This relationship involves two right handed Cartesian co-ordinate systems termed object space *(XYZ)* and image space *(xyz)* co-ordinate systems. Figure 2.1 illustrates both co-ordinate systems and the formation of an image point *p(x,y)* of an object space point *P(X,Y,Z)* on the image plane *I*. The point *O(X,Y,Z)* is termed the perspective or optical centre of the camera. The ray which is perpendicular to image plane *I* and that passes through optical centre *O(X,Y,Z)* is termed the optical axis or the principal ray.
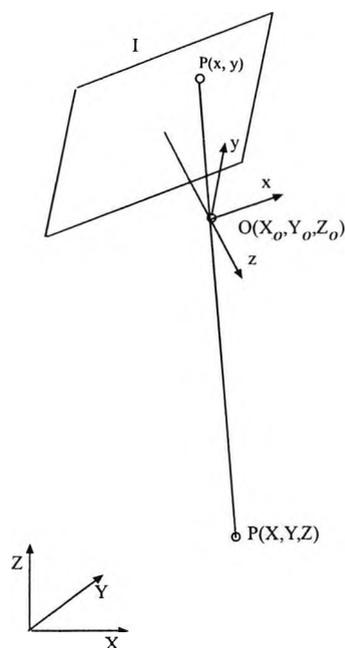
Figure 2.1 Camera model

The exterior orientation of a camera is defined by six parameters with respect to an object space co-ordinate system (X,Y,Z). Three parameters denoted by $X_0, Y_0,$ and Z_0 describe the position and the other three denoted by ω, φ, and κ describe the orientation.

The angular relationship between image and object space co-ordinate systems can be described by a *(3x3)* orthogonal rotational matrix *(M)*. Three sequential and independent angular rotations ω, φ, and κ around X, Y, and Z axes respectively can be represented by the following three matrices,

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix}$$

$$R_y = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

By multiplying these three matrices, the rotational matrix M can be formed as,

$$M = R_x R_y R_z = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \tag{2.12}$$

where,
$$m_{11} = -\cos\varphi \cos\kappa$$

$$m_{12} = \cos\omega \sin\kappa + \sin\omega \sin\varphi \cos\kappa$$

$$m_{13} = \sin\omega \sin\kappa - \cos\omega \sin\varphi \cos\kappa$$

$$m_{21} = -\cos\varphi \sin\kappa$$

$$m_{22} = \cos\omega \cos\kappa - \sin\omega \sin\varphi \sin\kappa$$

$$m_{23} = \sin\omega \cos\kappa + \cos\omega \sin\varphi \sin\kappa$$

$$m_{31} = \sin\varphi$$

$$m_{32} = -\sin\omega \cos\varphi$$

$$m_{33} = \cos\omega \cos\varphi \tag{2.13}$$

2.2.1 Collinearity condition

Referring to Figure 2.1, the object point, the optical centre of camera, and the image point are collinear in the ideal case. The mathematical equations expressing this

collinear condition are termed the collinearity equations (Wolf, 1983) which can be stated as,

$$x = -f\left[\frac{m_{11}(X-X_o)+m_{12}(Y-Y_o)+m_{13}(Z-Z_o)}{m_{31}(X-X_o)+m_{32}(Y-Y_o)+m_{33}(Z-Z_o)}\right]$$

$$y = -f\left[\frac{m_{21}(X-X_o)+m_{22}(Y-Y_o)+m_{23}(Z-Z_o)}{m_{31}(X-X_o)+m_{32}(Y-Y_o)+m_{33}(Z-Z_o)}\right] \qquad (2.14)$$

where (x,y), (X,Y,Z), and (X_o,Y_o,Z_o) are co-ordinates of the image point, object point, and perspective centre respectively. $(m_{11}, ..., m_{33})$ are nine rotational matrix elements and f is the focal length.

2.2.2 Camera interior parameters

In practice, there is a deviation from the collinear condition discussed in section 2.2.1 due to the distortion introduced by the lens system (Atkinson, 1996). For instance, because of the misalignments in the assembly of the lens components the intersection point of the optical axis and the imaging sensor does not normally coincide with the centre of the imaging sensor. Due to imperfections of the lens surface image points may shift away or towards the principal point. Furthermore, the effective focal length at close-range differs from that specified at infinity. Collectively, these systematic error parameters are termed camera interior parameters and their estimation is termed camera calibration.

2.2.2.1 Principal distance

The perpendicular distance from the optical centre of the lens system to the image plane is termed the principal distance. When the camera lens is focused at infinity, the principal distance is equal to the focal length (f). In close range applications, the camera lens needs to be refocused. Hence, the principal distance $(f+\delta f)$ needs to be determined where δf is the correction to the focal length.

2.2.2.2 Principal point shift

The location of an image point that is formed by a ray passing through the optical axis is termed the principal point. In an ideal camera, the centre of the image plane, termed the fiducial centre, coincides with the principal point. In practice, there is an offset

from fiducial centre to the principal point. Hence, this offset needs to be determined and the image co-ordinates based on the fiducial centre should be corrected.

2.2.2.3 Radial distortion

When an image point is displaced radially either closer to or away from the principal point, the distortion is termed radial. The radial distortion δr can be expressed as an odd termed polynomial function of the radial distance r,

$$\delta r = K_1 r^3 + K_2 r^5 + K_3 r^7 + \ldots\ldots\ldots \tag{2.15}$$

where K_1, K_2, K_3 are the coefficients of radial distortion and

$$r^2 = \left(x - x_p\right)^2 + \left(y - y_p\right)^2 \tag{2.16}$$

where (x,y) are the co-ordinates of an image point with respect to the fiducial centre and the (x_p, y_p) are co-ordinates of the true principal point.

The radial distortion at point (x,y) can be expressed in two components as,

$$\delta r_x = \frac{\delta r \left(x - x_p\right)}{r}, \qquad \delta r_y = \frac{\delta r \left(y - y_p\right)}{r} \tag{2.17}$$

2.2.2.4 Decentering distortion

Decentering distortion occurs due to misalignment of the components of the camera lens system. Decentering distortion at point (x,y) can be expressed as,

$$\delta x_d = \left[p_1 \left(r^2 + 2\left(x - x_p\right)^2\right) + 2 p_2 \left(x - x_p\right)\left(y - y_p\right) \right]$$

$$\delta y_d = \left[p_2 \left(r^2 + 2\left(y - y_p\right)^2\right) + 2 p_1 \left(x - x_p\right)\left(y - y_p\right) \right] \tag{2.18}$$

where δx_d and δy_d are the x and y components of the decentering distortion, r is the radial distance (equation 2.16), and P_1 and P_2 are decentering distortion coefficients.

The collinearity equations with camera interior parameters included can be stated as,

$$x - x_p + \delta x_r + \delta x_d = -\left(f + \delta f\right)\left[\frac{m_{11}(X - Xo) + m_{12}(Y - Yo) + m_{13}(Z - Zo)}{m_{31}(X - Xo) + m_{32}(Y - Yo) + m_{33}(Z - Zo)} \right]$$

$$y - y_p + \delta y_r + \delta y_d = -\left(f + \delta f\right)\left[\frac{m_{21}(X - Xo) + m_{22}(Y - Yo) + m_{23}(Z - Zo)}{m_{31}(X - Xo) + m_{32}(Y - Yo) + m_{33}(Z - Zo)}\right] \quad (2.19)$$

2.3 Camera exterior parameter estimation

The process of the estimation of location (X_0, Y_0, Z_0) and orientation $(\omega, \varphi, \kappa)$ of a camera with respect to a world co-ordinate system is known as the resection. There are number of techniques available (Wolf, 1983; Chen, 1995). Certain methods estimate these parameters directly (Zheng and Wang, 1992) which are termed the closed form solutions. Other methods estimate the parameters iteratively and require initial estimates of the parameters. The choice of the method used depends on the speed of the estimation process or the required statistical rigorousness of the estimated parameters.

The exterior orientation of one camera can also be estimated with respect to that of another camera. Thus estimated parameters represent the relative orientation.

2.3.1 Resection - direct solution

The Direct Linear Transformation *(DLT)* establishes a linear relationship between image co-ordinates and the object point co-ordinates (Abdel-Aziz and Karara, 1971). The *DLT* equations can be stated as,

$$x = \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_9 X + L_{10} Y + L_{11} Z + 1}$$

$$y = \frac{L_5 X + L_6 Y + L_7 Z + L_8}{L_9 X + L_{10} Y + L_{11} Z + 1} \quad (2.20)$$

where *(x,y)* are image point co-ordinates of an object point *(X,Y,Z)*, and $L_1, ..., L_{11}$ are *DLT* parameters. A minimum of six non-coplanar control points are required to solve for the *11* unknown *DLT* parameters. An alternative to this is the *2-D DLT* equations which only has eight unknowns (Shih, 1990; Wang, 1997). In this model control points are required to be coplanar *(Z=0)* hence, *Z* terms can be removed from Equations 2.20. The *2-D DLT* equations can be stated as,

$$x = \frac{L_1 X + L_2 Y + L_3}{L_7 X + L_8 Y + 1}$$

$$y = \frac{L_4 X + L_5 Y + L_6}{L_7 X + L_8 Y + 1} \qquad (2.21)$$

where,

$$L_1 = \frac{m_{11}}{L}, \quad L_2 = \frac{m_{12}}{L}, \quad L_3 = -\frac{(m_{11} X_L + m_{12} Y_L + m_{13} Z_L)}{L},$$

$$L_4 = \frac{m_{21}}{L}, \quad L_5 = \frac{m_{22}}{L}, \quad L_6 = -\frac{(m_{21} X_L + m_{22} Y_L + m_{23} Z_L)}{L},$$

$$L_7 = \frac{m_{31}}{L}, \quad L_8 = \frac{m_{32}}{L}, \quad L = -(m_{31} X_L + m_{32} Y_L + m_{33} Z_L) \qquad (2.22)$$

The nine rotation parameters hold the following conditions since the rotation matrix is orthogonal,

$$(m_{11}^2 + m_{21}^2 + m_{31}^2) = 1,$$

$$(m_{12}^2 + m_{22}^2 + m_{32}^2) = 1,$$

$$(m_{13}^2 + m_{23}^2 + m_{33}^2) = 1 \qquad (2.23)$$

Hence, L can be evaluated as,

$$L = \pm \frac{1}{\sqrt{L_1^2 + L_4^2 + L_7^2}} \qquad \text{or} \qquad L = \pm \frac{1}{\sqrt{L_2^2 + L_5^2 + L_8^2}} \qquad (2.24)$$

The camera rotation parameters can be obtained as,

$$\omega = \sin^{-1}\left(\frac{LL_8}{\cos\phi}\right)$$

$$\phi = -\sin^{-1}(LL_7)$$

$$k = -\sin^{-1}\left(\frac{LL_4}{\cos\phi}\right) \qquad (2.25)$$

The camera location parameters can be obtained as,

$$\begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}^{-1} \begin{bmatrix} LL_3 \\ LL_6 \\ -L \end{bmatrix} \qquad (2.26)$$

The estimated parameters using this method are only approximate. Assuming image co-ordinates are corrected for lens distortions the closeness of the estimations to their correct values depend on the accuracy of the measurement of control point locations. These estimations can be used as starting values in an iterative method to obtain more accurate values. Use of more than four control points provide better approximations.

2.3.2 Resection - iterative solution

The linearized collinearity equations can be used to iteratively estimate the camera exterior parameters. The approximate values of the unknown parameters and a minimum of three known control points are required. Use of more than three control points results in more accurate estimates of the parameters. Linearizing a pair of collinearity equations by partial differentiating with respect to exterior parameters X_L, Y_L, Z_L, ω, φ, and κ and image measurements,

$$f_{ix}(x_0) - b_{ix0} + \left(\frac{\partial F_{ix}}{\partial X_L}\right)_0 dX_L + \left(\frac{\partial F_{ix}}{\partial Y_L}\right)_0 dY_L + \left(\frac{\partial F_{ix}}{\partial Z_L}\right)_0 dZ_L$$
$$+ \left(\frac{\partial F_{ix}}{\partial \omega}\right)_0 d\omega + \left(\frac{\partial F_{ix}}{\partial \phi}\right)_0 d\phi + \left(\frac{\partial F_{ix}}{\partial \kappa}\right)_0 d\kappa + \left(\frac{\partial F_{ix}}{\partial b_{ix}}\right)_0 db_{ix} = 0$$

$$f_{iy}(x_0) - b_{iy0} + \left(\frac{\partial F_{iy}}{\partial X_L}\right)_0 dX_L + \left(\frac{\partial F_{iy}}{\partial Y_L}\right)_0 dY_L + \left(\frac{\partial F_{iy}}{\partial Z_L}\right)_0 dZ_L$$
$$+ \left(\frac{\partial F_{iy}}{\partial \omega}\right)_0 d\omega + \left(\frac{\partial F_{iy}}{\partial \phi}\right)_0 d\phi + \left(\frac{\partial F_{iy}}{\partial \kappa}\right)_0 d\kappa + \left(\frac{\partial F_{iy}}{\partial b_{iy}}\right)_0 db_{iy} = 0$$

$$(2.27)$$

where, $(i = 1, 2, ..., m)$ and $m \geq 3$.

In matrix form, Equation 2.27 can be stated as,

$$A \Delta x = b + v$$

The corrections to the initial estimates of camera exterior parameters can be estimated using least squares as,

$$\Delta x = \left(A' \, WA \right)^{-1} A' Wb \qquad (2.28)$$

where, W is the weight matrix of the *2-D* measurements and $(A'WA)^{-1}$ is the cofactor matrix of the estimated exterior parameters and,

$$
A = \begin{bmatrix}
\left(\dfrac{\partial F_{1x}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{1x}}{\partial Y_L} \right)_0 & \left(\dfrac{\partial F_{1x}}{\partial Z_L} \right)_0 & \left(\dfrac{\partial F_{1x}}{\partial \omega} \right)_0 & \left(\dfrac{\partial F_{1x}}{\partial \phi} \right)_0 & \left(\dfrac{\partial F_{1x}}{\partial \kappa} \right)_0 \\
\left(\dfrac{\partial F_{1y}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{1y}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{1y}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{1y}}{\partial \omega} \right)_0 & \left(\dfrac{\partial F_{1y}}{\partial \phi} \right)_0 & \left(\dfrac{\partial F_{1y}}{\partial \kappa} \right)_0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\left(\dfrac{\partial F_{mx}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{mx}}{\partial Y_L} \right)_0 & \left(\dfrac{\partial F_{mx}}{\partial Z_L} \right)_0 & \left(\dfrac{\partial F_{mx}}{\partial \omega} \right)_0 & \left(\dfrac{\partial F_{mx}}{\partial \phi} \right)_0 & \left(\dfrac{\partial F_{mx}}{\partial \kappa} \right)_0 \\
\left(\dfrac{\partial F_{my}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{my}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{my}}{\partial X_L} \right)_0 & \left(\dfrac{\partial F_{my}}{\partial \omega} \right)_0 & \left(\dfrac{\partial F_{my}}{\partial \phi} \right)_0 & \left(\dfrac{\partial F_{my}}{\partial \kappa} \right)_0
\end{bmatrix}
$$

$$
\Delta x = \begin{bmatrix} dX_L \\ dY_L \\ dZ_L \\ d\omega \\ d\phi \\ d\kappa \end{bmatrix}
\qquad
b = \begin{bmatrix} -f_{1x}(x_0) + b_{1x0} \\ -f_{1y}(x_0) + b_{1y0} \\ \cdot \\ \cdot \\ -f_{mx}(x_0) + b_{mx0} \\ -f_{mx}(x_0) + b_{mx0} \end{bmatrix}
\qquad
v = \begin{bmatrix} -db_{1x} \\ -db_{1y} \\ \cdot \\ \cdot \\ -db_{mx} \\ -db_{my} \end{bmatrix}
$$

After each iteration, updated values of the unknown exterior parameters can be computed as,

$$
x = \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ \omega \\ \phi \\ \kappa \end{bmatrix} = \begin{bmatrix} X_{L0} \\ Y_{L0} \\ Z_{L0} \\ \omega_0 \\ \phi_0 \\ \kappa_0 \end{bmatrix} + \Delta x \qquad (2.29)
$$

The iterations need to be continued until the corrections are smaller than some preset threshold. The partial differentials in Equation 2.27 are given in the Appendix.

2.3.3 Relative orientation

Figure 2.2 illustrates an image pair I_1 and I_2 whose local axes systems are (x_1,y_1,z_1) and (x_2,y_2,z_2). The images of the object space point $A_i(X_i,Y_i,Z_i)$ on images I_1 and I_2 are at points $p_{1i}(x_{1i},y_{1i})$ and $p_{2i}(x_{2i},y_{2i})$. The coplanarity created by points A_i, O_1, O_2, p_{1i}, and p_{2i} can be used to estimate the relative orientation of two images I_1 and I_2 (Atkinson, 1996; Ghosh, 1972).
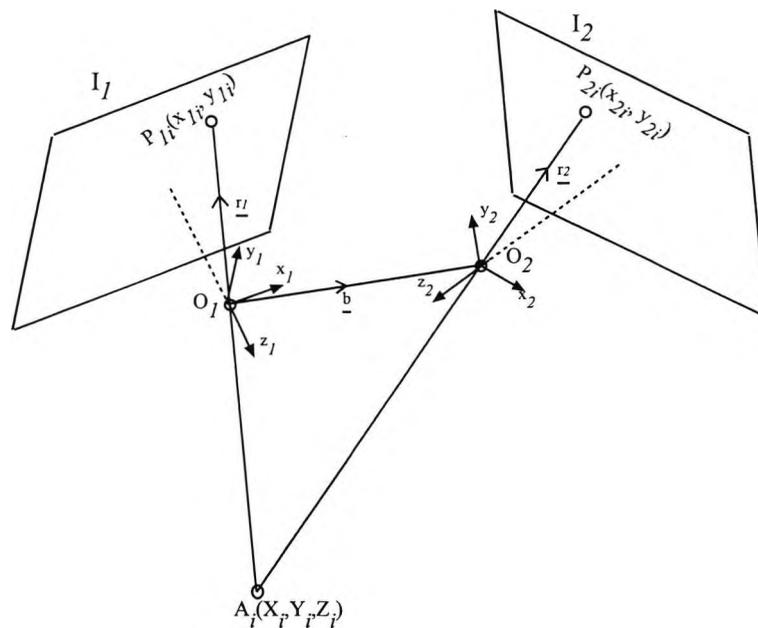
Figure 2.2 Geometry of coplanarity

Vectors $\underline{r_1}$, $\underline{r_2}$, and $\underline{b}$ are coplanar. Considering (x_1,y_1,z_1) as the primary axis system, three vectors can be stated as (assuming the same scale),

$$\underline{r1} = \begin{bmatrix} x_{1i} & y_{1i} & -f_1 \end{bmatrix}^t, \quad \underline{r2} = R' \begin{bmatrix} x_{2i} & y_{2i} & -f_2 \end{bmatrix}^t, \text{ and } \underline{b} = \begin{bmatrix} b_x & b_y & b_z \end{bmatrix}^t \qquad (2.30)$$

where R is the $(3x3)$ rotational matrix.

Due to the coplanarity, the triple scalar product of the three vectors should be zero. That is,

$$\underline{b}.(\underline{r1} \times \underline{r2}) = 0 \qquad (2.31)$$

The condition in matrix form,

$$\begin{bmatrix} b_x & b_y & b_z \\ x_{1i} & y_{1i} & z_{1i} \\ (m_{11}x_{2i}+m_{21}y_{2i}-m_{31}f_2) & (m_{12}x_{2i}+m_{22}y_{2i}-m_{32}f_2) & (m_{13}x_{2i}+m_{23}y_{2i}-m_{33}f_2) \end{bmatrix} = 0$$

Substituting t_1, t_2, and t_3 for $(m_{11}x_{2i}+m_{21}y_{2i}-m_{31}f_2)$, $(m_{12}x_{2i}+m_{22}y_{2i}-m_{31}f_2)$, and $(m_{11}x_{2i}+m_{21}y_{2i}-m_{31}f_2)$,

$$\begin{bmatrix} b_x & b_y & b_z \\ x_{1i} & y_{1i} & -f_1 \\ t_1 & t_2 & t_3 \end{bmatrix} = 0$$

Hence, the coplanarity equation can be written as,

$$b_x\left(y_{1i}t_3 + t_2 f_1\right) + b_y\left(x_{1i}t_3 + t_1 f_1\right) + b_z\left(x_{1i}t_2 - t_1 y_{1i}\right) = 0 \qquad (2.32)$$

Assuming b_x is not zero, the condition can be stated as,

$$\left(y_{1i}t_3 + t_2 f_1\right) + \frac{b_y}{b_x}\left(x_{1i}t_3 + t_1 f_1\right) + \frac{b_z}{b_x}\left(x_{1i}t_2 - t_1 y_{1i}\right) = 0 \qquad (2.33)$$

Equation 2.33 is non-linear and hence should be linearized using the Taylor expansion. The five relative orientation parameters $\left(\frac{b_y}{b_x}\right)$, $\left(\frac{b_z}{b_x}\right)$, ω, ϕ, and κ can be estimated if five pairs of corresponding image points are known. When more than five points are available, a least squares solution can be obtained. Initial estimates of unknown relative orientation parameters are required.

2.4 *3-D* location estimation - Intersection

The process of the estimation of co-ordinates of an unknown *3-D* location using known camera exterior parameters and image measurements is termed intersection. Direct or iterative methods can be used, the choice of the method depends on the speed or statistical rigourousness required (Atkinson, 1986).

2.4.1 Intersection - non-rigorous solution

In non-rigourous direct intersection, camera exterior parameters and image measurements are considered as constants. By rearranging the collinearity equations (Equation 2.14), 3-D co-ordinates of a point can be estimated. A minimum of two image points in each image are required which results in four equations to solve for three unknowns. Hence, least squares can be used to obtain a solution. Estimations using this method are less reliable as the random errors are not taken into consideration. Assuming *m* 3-D points and *n* cameras, there will be *3m* unknowns and *2mn* equations provided each camera views the same set of 3-D points. A set of re-arranged equations can be stated as,

$$
\begin{aligned}
\left(x_{ji} m_{j31} + f_j m_{j11} \right) X_i + \left(x_{ji} m_{j32} + f_j m_{j12} \right) Y_i + \left(x_{ji} m_{j33} + f_j m_{j13} \right) Z_i \\
= x_{ji} \left(m_{j31} X_{jL} + m_{j32} X_{jL} + m_{j33} X_{jL} \right) + f_i \left(m_{j11} X_{jL} + m_{j12} Y_{jL} + m_{j13} Z_{jL} \right)
\end{aligned}
$$

$$
\begin{aligned}
\left(y_{ji} m_{j31} + f_j m_{j21} \right) X_i + \left(y_{ji} m_{j32} + f_j m_{j22} \right) Y_i + \left(y_{ji} m_{j33} + f_j m_{j23} \right) Z_i \\
= y_{ji} \left(m_{j31} X_{jL} + m_{j32} X_{jL} + m_{j33} X_{jL} \right) + f_i \left(m_{j21} X_{jL} + m_{j22} Y_{jL} + m_{j23} Z_{jL} \right)
\end{aligned}
\tag{2.34}
$$

where $\left(i = 1, 2, \ldots\ldots m \right)$ and $\left(i = 1, 2, \ldots\ldots n \right)$.

In the matrix form, set of equations can be represented as,

$$
A \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = B
$$

Using least squares unknowns can be estimated as,

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \left(A' A \right)^{-1} \left(A' B \right)
\tag{2.35}
$$

2.4.2 Intersection - iterative solution

A more refined estimation of the *3-D* object point co-ordinates can be obtained by iterative intersection. In this method, linearized collinearity equations are used. Initial estimates of the *3-D* co-ordinates (X_0, Y_0, Z_0) are required which may be obtained using the non-rigorous intersection method. Considering images of a *3-D* point in *m* cameras, a set linearized collinearity equations can be stated as,

$$f_{ix}(x_0) - b_{ix0} + \left(\frac{\partial F_{ix}}{\partial X}\right)_0 dX + \left(\frac{\partial F_{ix}}{\partial Y}\right)_0 dY + \left(\frac{\partial F_{ix}}{\partial Z}\right)_0 dZ + \left(\frac{\partial F_{ix}}{\partial b_{ix}}\right)_0 db_{ix} = 0$$

$$f_{iy}(x_0) - b_{iy0} + \left(\frac{\partial F_{iy}}{\partial X}\right)_0 dX + \left(\frac{\partial F_{iy}}{\partial Y}\right)_0 dY + \left(\frac{\partial F_{iy}}{\partial Z}\right)_0 dZ + \left(\frac{\partial F_{iy}}{\partial b_{iy}}\right)_0 db_{iy} = 0 \quad (2.36)$$

where, $(i = 1, 2, ..., m)$ and $m \geq 3$.

In matrix form,

$$A \, \Delta x = b + v$$

Using least squares, corrections to unknowns can be obtained as,

$$\Delta x = \left(A'WA\right)^{-1} A'Wb \qquad (2.37)$$

where, *W* is the weight matrix of the *2-D* measurements. *(A'WA)⁻¹* is the cofactor matrix where,

$$A = \begin{bmatrix} \left(\dfrac{\partial F_{1x}}{\partial X}\right)_0 & \left(\dfrac{\partial F_{1x}}{\partial Y}\right)_0 & \left(\dfrac{\partial F_{1x}}{\partial Z}\right)_0 \\[1em] \left(\dfrac{\partial F_{1y}}{\partial X}\right)_0 & \left(\dfrac{\partial F_{1y}}{\partial Y}\right)_0 & \left(\dfrac{\partial F_{1y}}{\partial Z}\right)_0 \\[1em] \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \left(\dfrac{\partial F_{mx}}{\partial X}\right)_0 & \left(\dfrac{\partial F_{mx}}{\partial Y}\right)_0 & \left(\dfrac{\partial F_{mx}}{\partial Z}\right)_0 \\[1em] \left(\dfrac{\partial F_{my}}{\partial X}\right)_0 & \left(\dfrac{\partial F_{my}}{\partial Y}\right)_0 & \left(\dfrac{\partial F_{my}}{\partial Z}\right)_0 \end{bmatrix}$$

$$\Delta x = \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} \qquad b = \begin{bmatrix} -f_{1x}(x_0)+b_{1x0} \\ -f_{1y}(x_0)+b_{1y0} \\ \cdot \\ \cdot \\ -f_{mix}(x_0)+b_{mx0} \\ -f_{mx}(x_0)+b_{mx0} \end{bmatrix} \qquad v = \begin{bmatrix} -db_{1x} \\ -db_{1y} \\ \cdot \\ \cdot \\ -db_{mx} \\ -db_{my} \end{bmatrix}$$

After each iteration, the updated X, Y, and Z values can be computed as,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} + \begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} \qquad (2.38)$$

Iterations should be continued until corrections are smaller than a pre-determined threshold. The resulting *3-D* estimations are more accurate than those produced by non-rigorous solution due to the minimisation of residuals (Atkinson, 1996). The partial differentials in Equation 2.36 are given in Appendix.

2.5 Bundle Adjustment

The separate estimation of camera exterior parameters (resection) and *3-D* object point co-ordinates (intersection) were discussed in the previous sections. These quantities can also be estimated simultaneously. This estimation process is known as the bundle adjustment which uses the functional model based on linearized collinearity equations (Atkinson, 1996). Initial estimates of the unknowns are required and are updated after each iteration until the adjustment to each quantity is negligibly small.

Considering co-ordinates of the ith *3-D* point (X_i, Y_i, Z_i) and exterior parameters $X_L, Y_L, Z_L, \omega, \varphi, \kappa$ of jth camera, linearized collinearity equations can be written as,

$$f_{ijx}(x_0) - b_{ijx0} + \left(\frac{\partial F_{ijx}}{\partial X_i}\right)_0 dX_i + \left(\frac{\partial F_{ijx}}{\partial Y_i}\right)_0 dY_i + \left(\frac{\partial F_{ijx}}{\partial Z_i}\right)_0 dZ_i$$

$$+ \left(\frac{\partial F_{ijx}}{\partial X_{Lj}}\right)_0 dX_{Lj} + \left(\frac{\partial F_{ijx}}{\partial Y_{Lj}}\right)_0 dY_{Lj} + \left(\frac{\partial F_{ijx}}{\partial Z_{Lj}}\right)_0 dZ_{Lj}$$

$$+ \left(\frac{\partial F_{ijx}}{\partial \omega_j}\right)_0 d\omega + \left(\frac{\partial F_{ijx}}{\partial \phi_j}\right)_0 d\phi + \left(\frac{\partial F_{ijx}}{\partial \kappa_j}\right)_0 d\kappa + \left(\frac{\partial F_{ijx}}{\partial b_{ijx}}\right)_0 db_{ijx} = 0$$

$$f_{ijy}(x_0) - b_{ijy0} + \left(\frac{\partial F_{ijy}}{\partial X_i}\right)_0 dX_i + \left(\frac{\partial F_{ijy}}{\partial Y_i}\right)_0 dY_i + \left(\frac{\partial F_{ijy}}{\partial Z_i}\right)_0 dZ_i$$

$$+ \left(\frac{\partial F_{ijy}}{\partial X_{Lj}}\right)_0 dX_{Lj} + \left(\frac{\partial F_{ijy}}{\partial Y_{Lj}}\right)_0 dY_{Lj} + \left(\frac{\partial F_{ijy}}{\partial Z_{Lj}}\right)_0 dZ_{Lj}$$

$$+ \left(\frac{\partial F_{ijy}}{\partial \omega_j}\right)_0 d\omega + \left(\frac{\partial F_{ijy}}{\partial \phi_j}\right)_0 d\phi + \left(\frac{\partial F_{ijy}}{\partial \kappa_j}\right)_0 d\kappa + \left(\frac{\partial F_{ijy}}{\partial b_{ijy}}\right)_0 db_{ijy} = 0$$

(2.39)

In matrix form the system of equations can be written as,

$$A\Delta x = b + v \qquad (2.40)$$

where, A is the design matrix , Δx is the matrix of corrections to unknowns, b is the matrix of image measurements, and v is the matrix of image residuals.

The corrections to the unknown 3-D object points and camera exterior parameters (Δx) can be estimated as,

$$\Delta x = \left(A' WA\right)^{-1} A' Wb \qquad (2.41)$$

$C = \left(A'WA\right)^{-1}$ is the cofactor matrix.

After each iteration, the updated value of unknowns can be calculated. The resulting 3-D point co-ordinates and camera exterior parameters are known as adjusted values which are the most probable estimates. When camera interior parameters are also estimated along with 3-D point co-ordinates and camera exterior parameters, the adjustment process is known as the self-calibration bundle adjustment.

The image residuals may be computed as,

$$v = A\Delta x - b \qquad (2.42)$$

Hence, standard deviation of the residuals or precision of image measurements can be computed as,

$$\sigma_0 = \sqrt{\frac{v^t W v}{r}} \qquad (2.43)$$

where r is the number of degrees of freedom.

The solution to the system of equations is possible only if the matrix $(A^t W A)$ (in Equation 2.41) is non-singular. But this matrix is always singular hence, it is required to include constraints to make it non-singular (Atkinson, 1996).

2.6 Image residuals

When 3-D object points and camera exterior parameters are considered as unknowns (bundle adjustment), the least squares estimation gives the most probable values for the unknowns. Hence, Equation 2.42 gives the most probable values for the image residuals and the standard deviation of residuals can be obtained using Equation 2.43. The results of this statistical analysis of image residuals can be used in image point correspondence algorithms where the epipolar constraint is used (chapter 3). Considering an image point in one image of a stereo pair, the corresponding point in the other image can be found by searching within a band which symmetrically covers either side of the corresponding epipolar line. The statistical knowledge of the image residuals can be used to estimate the optimum width of this epipolar search band (Ariyawansa and Clarke, 1999).

2.7 Summary of the chapter

In high precision close range photogrammetric measurements the random errors are taken into account by the least squares technique. Assuming individual measurements are independent and random errors follow a normal distribution, the least squares technique estimates unknowns by minimising the sum of the squares of the residuals. In addition, the statistics of the estimations are given by the covariance matrix.

The ideal collinear condition between the object point, the camera perspective centre, and the image point is an important relationship used in photogrammetry which results in collinearity equations. However, in reality a ray that passes through the perspective

centre is deviated due to the imperfections of the lens system. Hence, for high precision applications the errors caused by the lens system should be taken into consideration.

The estimation of camera exterior parameters is normally the first step of a measurement process. Often a closed-form solution is used to obtain approximate parameters and then an iterative method is used to obtain more accurate parameters. The *2-D DLT* method is a closed-form solution that requires a minimum of four coplanar control point. The precision of the estimations depend on the accuracy with which the control points are measured. The iterative method which uses linearized collinearity equations produces more refined parameters and requires a minimum of three control points. Relative orientation expresses exterior parameters of one camera with respect that of another.

When camera exterior parameters are known and image measurements are given the estimates of *3-D* locations may be obtained using intersection. Image point locations can be corrected for lens distortions using pre-estimated camera interior parameters. The rearranged non-linear collinearity equations can be used to obtain approximate values for *3-D* locations. Normally these approximate values are used by an iterative method as initial values to obtain more accurate estimations. The iterative method uses least squares and provides statistical quality of estimations.

The bundle adjustment iteratively estimates both the camera exterior parameters and the *3-D* locations. The approximations of the unknowns are required. The estimations provide the most probable values for the unknowns and the statistical quality of the estimations are provided by the covariance matrix. The self calibrating bundle adjustment estimates camera interior parameters in addition to the exterior parameters and the *3-D* locations.

Chapter 3

Multiple view correspondence methods

Automatically finding corresponding features in images has been a research topic for many years. Feature correspondences are required for tracking, *3-D* location estimation, object identification, and camera lens system calibration to name a few. Correspondences may be required from a sequence of images acquired from a single camera or a set of images acquired from multiple cameras located at different view points. The camera(s) and/or object(s) may be moving or stationary (Koch, 1994; Wang and Duncan, 1996; Kang *et al.* 1997).

Extracting *3-D* information of a scene using two or more images is an important passive method in high precision close-range photogrammetry and computer vision. Applications may range from obtaining the structure of a stationary object to guiding a robot to perform an assembly task. The extraction of the image features of interest is the first step. In high precision applications, accurate location of image features is important for the subsequent establishment of correspondences and *3-D* localisation. Images of the natural features of objects cannot be located as accurately due to the radiometric and geometric distortions which are dependant upon the view point. Hence, in high precision applications object points of interest are signalised using artificial circular retro-reflective features which are usually termed targets. Due to the special reflective characteristics these targets can be located accurately in image space.

In computer vision literature, the word 'matching' is widely used instead of the word 'correspondences'. The reason is that most algorithms use intensity based matching of small areas of images (windows) in order to establish correspondences. These algorithms use both radiometric and geometric information. In this thesis, both words are used as and when appropriate. However, the material in this thesis mainly involve the discussion and development of algorithms which use only the geometric techniques. Hence, the word 'correspondence' is used widely.

In this chapter, feature extraction techniques are discussed, matching or correspondence algorithms are categorised. Feature extraction and correspondence

methods that are used in high-precision close-range photogrammetric applications are discussed in detail. The traditional geometric techniques and the new geometric techniques developed by the computer vision researchers are discussed.

3.1 Classification of image matching methods

There has been a vast number of image matching methods developed over the years. In general these methods exploit the geometry between images and intensity information for the establishment of correspondences. Image matching methods can be categorised in a number of ways which may be depending on the imaging geometry such as parallel or non-parallel optical axes, the number of cameras used such as binocular, trinocular and multiocular, and the image intensity matching technique used which may be correlation or least squares. In this chapter image matching is broadly categorised as area based matching *(ABM)* and feature based matching *(FBM)*.

There has been a significant amount of research on *ABM* techniques by the computer vision researchers (Rosenfeld, 1998). The *ABM* techniques are based on the fact that the projections of a *3-D* point in different images are likely to have similar gray values (Fua, 1997). Most of the past *ABM* algorithms used correlation between intensity patterns of the local neighbourhood (window) of points in images (Ruther, 1996; Kanade *et al*, 1999). Initially, a point of interest is chosen in one image then a cross-correlation measure is used to search for a point with a matching neighbourhood in the other images. If the relative or absolute orientation of an image pair is known the search can be constrained using epipolar geometry (Okutomi and Kanade, 1993). A recent development in *ABM* is the use of Least Squares *(LS)* technique for image matching (Atkinson, 1996; Baltsavias, 1991). *LS* matching algorithms are considered to be more robust and accurate compared to correlation based methods. The drawbacks of *ABM* algorithms are that they are sensitive to the angular separation of stereo pairs, contrast, and illumination characteristics (Dhond and Aggarwal, 1989).

FBM algorithms use natural features that are extracted from images rather than using image intensity values directly (Ayache and Hansen, 1988). Features commonly used are corners, edge points, and edge segments. *FBM* algorithms, depending on the features used, are sensitive to noise and do not perform well in highly textured images (Dhond and Aggarwal, 1989). In high-precision applications, *FBM* techniques are

preferred to *ABM* techniques due to the use of signalising points whose images can be distinguished from the background and can be located accurately. In certain applications signalising points (targets) with some unique structures are used.   An example would be the coded targets that are widely used in measurement applications. When such targets are used simple projective geometric constraints can be used to establish correspondences. The use of features alone without using the neighbouring intensity information may increase the likelihood of establishing false correspondences in complex object space situations but these methods are simpler and faster than *ABM* methods.

3.2 Feature extraction techniques

The image features of interest depend on the application. For low accuracy applications, natural features such as corners, edge points, or edge segments can be extracted using an interest operator. There are numerous such operators available. Baltsavias, 1991 pointed out that for feature detection Hadem, Moravec, Forestner, and Hannah operators and Canny edge detector have been used. These operators perform intensity based operations to extract features. In most cases, some quantitative attributes such as high grey level variance, steep correlation, distance to other local features are also calculated for each feature to aid the subsequent correspondence process (Baltsavias, 1991).

In high-accuracy applications, the precision of image feature location is important as it affects the accuracy of *3-D* localisation. Natural features such as those mentioned previously cannot be located accurately. Hence, artificial targets are placed on the object points of interest at which measurements are to be made. Normally, these targets are small (radius of *2-10 mm*) and of circular shape to enable accurate location estimation (Baltsavias, 1991). The use of targets that are made of retro-reflective material and the use of special illumination ensures that these targets are distinguishable from the background (Clarke and Wang, 1996; Shortis *et al,* 1994). In most applications, background clutter can be considerably reduced or completely removed by setting a small aperture since the background intensity information is of little or no importance. In many cases, a pre-determined global threshold is used to distinguish targets from image background. Attributes such as peak intensity and area are determined for each target.

Coded targets, which can be easily distinguishable due to their specific patterns and retro-reflective characteristics, are used in many applications. Currently, there are two types in use. In the first type, a central dot is surrounded by broken concentric rings. The number of broken segments in each concentric ring are used to label targets. The second type comprises a central dot surrounded by a number of dots. The number of secondary dots and their differing sizes are used to label the target (Fraser, 1997).

3.2.1 Target detection techniques

Contour tracing with chain codes is a technique used for the detection of image features (West, 1982; Chen, 1995; Pavlidis, 1982). In the first step, the image is binarized using a suitable threshold. A suitable chain coding operator is then used to trace boundaries of the features. Figure 3.1(a) illustrates part of an image with a binarized feature. The grid points marked with circles represent binarized feature pixels. Figure 3.1(b) illustrates two types of chain-code operators: *4*-way and *8*-way. The tracing algorithm scans each line of the binarized image from top to bottom. When a binarized feature pixel is encountered, chain-code operator begins tracing the feature boundary and the direction vectors are noted. The tracing a feature ends when next pixel to be traced is the one that was considered first. Once completely detected, the feature can be removed from the binarized image to enable the detection of the new features.

(a)   (b)
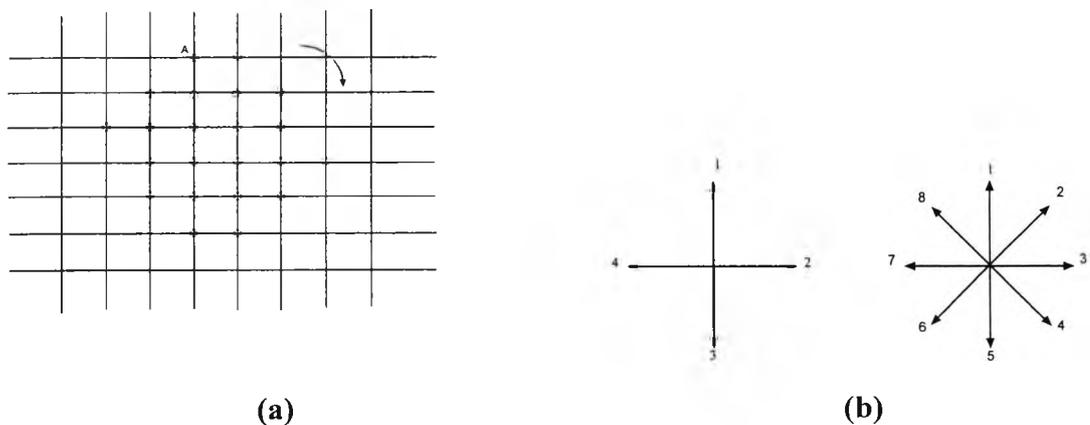
Figure 3.1(a). Binarized feature in a section of an image,   (b). 4 and 8 way operators

Using an 8-way operator in the clockwise direction beginning at point A (Figure 3.1(a)), the chain code for the feature is detected as '*3,4,5,4,6,6,7,8,1,8,2*'. Using this

information, the perimeter length can be obtained directly and the co-ordinates of the periphery and the area can also be obtained.

Another way of implementing this algorithm is to use the true intensity image without binarizing. Once a feature is completely detected and co-ordinates of the periphery are obtained, its internal intensity information may be used for more accurate localisation of the feature.

3.2.2 Target recognition techniques

Once a feature is detected in the image, a number of tests can be performed to identify if the detected feature is a legitimate target. Attributes such as area, perimeter, shape factor, maximum width and length, and peak intensity are commonly used. The perimeter length can be computed using the traced contour co-ordinates. Area can be computed as the number of pixels in the target image. The circular and non-circular features can be distinguished by the shape factor (Chen *et al*, 1992) which can be computed as,

$$Q = A \Big/ \left[\pi \left(L/2 \right)^2 \right] \qquad (3.1)$$

where A is the computed area of the feature and L is the longest distance across the feature. If the shape of the feature is closer to a circle, the value of shape factor Q is closer to 1. Since targets used in high-accuracy applications are circular, the shape factor is a useful measure. The peak intensity of the target image is another recognition factor which is normally used to complement the information given by the shape factor and the area of a feature.

3.2.3 Target location techniques

Once features are recognised as legitimate targets, the location is computed to sub-pixel accuracy. There are number of methods which are outlined and their characteristics are discussed.

Shape centre (average of perimeter): The average of co-ordinates $(\bar{x}, \bar{y})$ of the perimeter of a target image can be computed as,

$$\bar{x} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} x_i \ , \qquad \bar{y} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} y_i \qquad (3.2)$$

where (x_i, y_i) is the ith pixel co-ordinate and n is the total number of pixels in the perimeter.

Binary centroid: In this method, binarized target images are used to compute the binary centroid $(\bar{x}, \bar{y})$ which can be stated as,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot I_i}{\sum_{i=1}^{n} I_i} \ , \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i \cdot I_i}{\sum_{i=1}^{n} I_i} \qquad (3.3)$$

where (x_i, y_i) is the ith pixel co-ordinate at which the binarized intensity I_i which is either one or zero.

Grey scale centroid: The grey scale centroid can be represented by the same equations as (3.3), but I_i represent the actual unthresholded intensity value at pixel location (x_i, y_i).

Squared grey scale centroid: The computation of squared grey scale centroid is the same as gray scale centroid but the squared value of the intensity is used.

Shortis *et al,* 1994 carried out simulation tests to investigate relative merits of each location method for the Gaussian shaped targets. The effects on target location accuracy due to the number of quantization levels of *A/D* conversion, threshold, target size, intensity saturation, *DC* offset of the video signal to the *A/D* converter input was analysed. Intensity methods (gray scale centroid and squared grey scale centroid) have better accuracy performances when number of quantization levels are higher due to the better intensity resolution of pixels around the edge of the target. The threshold methods (shape centre and binary centroid) do not show clear improvement. As the threshold intensity increases, the location accuracy of the intensity methods decreases again due to the removal of pixels on the edge of the target. With the threshold methods, location accuracy increases as target size increases where as it remains stable

for intensity methods when diameter is greater than 6-7 pixels. In general, intensity methods have better location accuracy performance than threshold methods.

3.3 Geometric techniques for establishing correspondences

Using a suitable technique image points of interest are located in each image. Taking a point in one image it is unknown which point in each other image corresponds to the same 3-D world point.

The geometric techniques provide vital information for establishing correspondences. In *ABM* algorithms, geometric techniques are used to constrain the search space. In *FBM* algorithms, geometric techniques are used for establishing feature correspondences between images.

3.3.1 Epipolar geometry

Epipolar geometry is a fundamental projective geometric relationship that exists between a pair of images (Haralick and Shapiro, 1993; Faugeras, 1995; Kanatani, 1993). Considering a stereo pair where the relative or the absolute orientation of images are known the epipolar geometry can be used to establish correspondences. Figure 3.2 illustrates a convergent optical axes image pair I_1 and I_2 whose optical centres are O_1 and O_2 respectively.
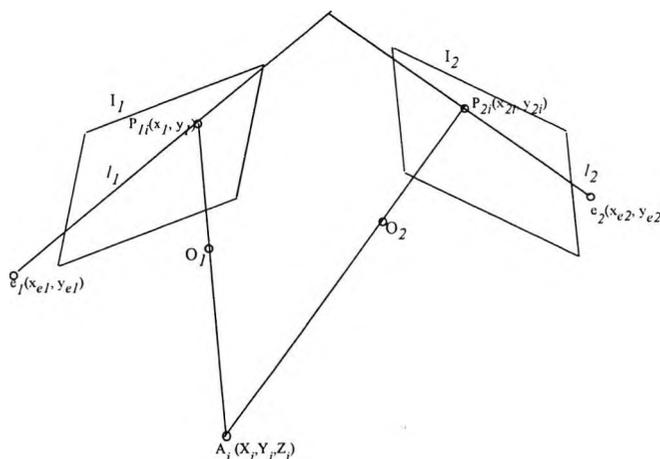
Figure 3.2 Epipolar geometry

Images of a 3-D world point $A_i(X_i, Y_i, Z_i)$ on images I_1 and I_2 are at points $p_{1i}(x_{1i}, y_{1i})$ and $p_{2i}(x_{2i}, y_{2i})$. The line joining the optical centres O_1 and O_2 is termed the baseline. Points $e_1(x_{e1}, y_{e1})$ and $e_2(x_{e2}, y_{e2})$ are termed epipoles and are the images of optical centers O_2 and

O_1 on images I_1 and I_2 respectively. The plane formed by points O_1, O_2, and A is termed the epipolar plane which intersects image planes I_1 and I_2 along the epipolar lines l_{1i} and l_{2i}. Hence, for point $p_{1i}(x_{1i}, y_{1i})$ which lie on epipolar line l_{1i}, the corresponding point $p_{2i}(x_{2i}, y_{2i})$ can be found by searching along the conjugate epipolar line l_{2i}. This condition is termed the epipolar constraint.

3.3.1.1 Determination of conjugate epipolar line

In order to find a corresponding point in the image I_2 for a given point in image I_1, the first step is to determine the equation of epipolar line l_{2i} (Figure 3.3). Given the exterior orientation of cameras and co-ordinates of point p_{1i}, equation of l_{2i} can be determined.
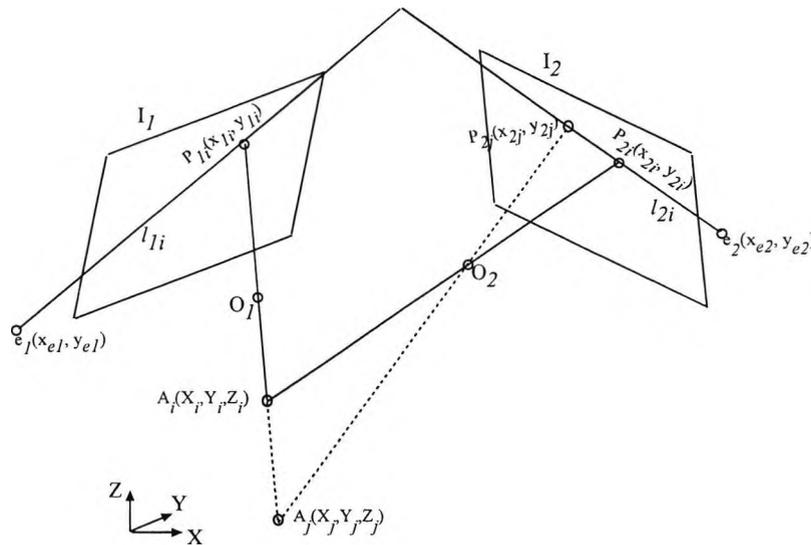
Figure 3.3 Conjugate epipolar line

By projecting a ray from p_{1i} through O_1 to XY-plane of the world co-ordinate system, a 3-D point $A_j(X_j, Y_j, Z_j = 0)$ can be determined. Reprojecting a ray from this point to image I_2 through O_2 a point $p_{2j}(x_{2j}, y_{2j})$ in image I_2 can be determined. Point p_{2j} will also fall on the same epipolar line l_{2i}. If co-ordinates of epipole $e_2(x_{e2}, y_{e2})$ are known, the epipolar line l_2 can be determined as follows,

$$\frac{(y - y_{2j})}{(x - x_{2j})} = \frac{(y_{2j} - y_{e2})}{(x_{2j} - x_{e2})}$$

By rearranging in $y = mx + c$ form,

$$y = \left(\frac{y_{2j} - y_{e2}}{x_{2j} - x_{e2}} \right) x + \left(y_{2j} - \left(\frac{y_{2j} - y_{e2}}{x_{2j} - x_{e2}} \right) x_{2j} \right) \qquad (3.4)$$

3.3.1.2 Determination of search band

With real imaging systems, the epipolar constraint discussed in the previous section will not be satisfied exactly and corresponding points will lie a small distance away from the epipolar line. This is true even after correcting for lens distortion. The reason for this deviation is because of the unknown random errors of the imaging system. Hence, a band that symmetrically covers either side of the ideal epipolar line is considered when searching for corresponding points. This search window is termed the search band b_{2i} and its border lines l_{1bi} and l_{2bi} are termed search band border lines (Figure 3.4). Gradients of border lines l_{1bi} and l_{2bi} are the same as that of the epipolar line l_{2i}. The perpendicular distance between the parallel border lines is termed the search band width (d). A detailed discussion on determining search band width is given in chapter 5.
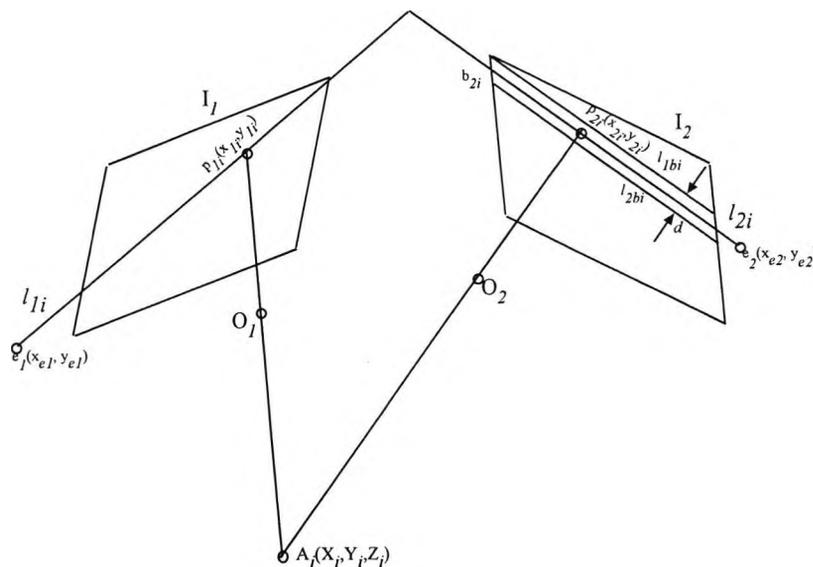
Figure 3.4 Illustration of search band

As illustrated in Figure 3.3, once a point on epipolar line l_2 (p_{2j}) is determined, the next step is to determine a point that lie on each border line. Figure 3.5 illustrates epipolar lines with positive and negative gradients and the determination of points that fall on each border line.
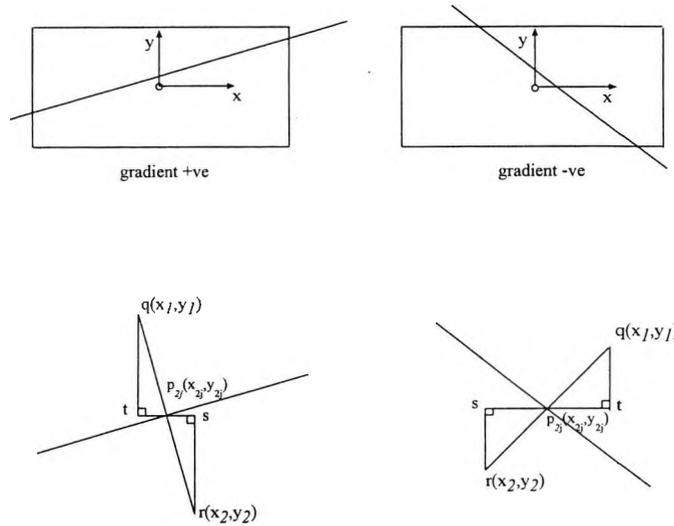
Figure 3.5 Possible epipolar line gradients and points on each border line

Given the co-ordinates of point $p_{2j}(x_{2j}, y_{2j})$ and width of the search band *(qr $\equiv$ d)*, the points $q(x_1, y_1)$ and $r(x_2, y_2)$ that lie on border lines can be determined.

When the gradient is positive,

$$q \equiv \left(\left(x_{2j} - \left(\frac{d}{2} \right) \sin\theta \right), \left(y_{2j} + \left(\frac{d}{2} \right) \cos\theta \right) \right)$$

$$r \equiv \left(\left(x_{2j} + \left(\frac{d}{2} \right) \sin\theta \right), \left(y_{2j} - \left(\frac{d}{2} \right) \cos\theta \right) \right) \tag{3.5}$$

When the gradient is negative,

$$q \equiv \left(\left(x_{2j} + \left(\frac{d}{2} \right) \sin\theta \right), \left(y_{2j} + \left(\frac{d}{2} \right) \cos\theta \right) \right)$$

$$r \equiv \left(\left(x_{2j} - \left(\frac{d}{2} \right) \sin\theta \right), \left(y_{2j} - \left(\frac{d}{2} \right) \cos\theta \right) \right) \tag{3.6}$$

where $\theta = \tan^{-1}(|m|$, and *m* is the gradient of the epipolar line. Special situations occur when $m = 0$ and $m \rightarrow \infty$.

If the intercepts of border lines are c_1 and c_2, the equations can be written as

$$y = mx + c_1, \quad y = mx + c_2 \tag{3.7}$$

If an image point (x_i, y_i) falls within the band then

$$(y_i - mx_i - c_1) > 0 \quad \text{and} \quad (y_i - mx_i - c_2) < 0 \tag{3.8}$$

or

$$(y_i - mx_i - c_1) < 0 \quad \text{and} \quad (y_i - mx_i - c_2) > 0 \tag{3.9}$$

Hence, potentially corresponding points can be identified.

3.3.1.3 Epipolar geometry in parallel optical axis images

A special situation of the epipolar geometry occurs when the local axes systems are parallel and both images lie on a plane which is parallel to the baseline O_1O_2 (Figure 3.6). The x-axes of the local co-ordinate systems are parallel to the baseline. In each image the epipoles are at infinity. The epipolar plane created by points A_i, O_1, and O_2 intersects image planes I_{r1} and I_{r2} along conjugate epipolar lines l_{r1i} and l_{r2i} that are collinear. In ideal geometric situation, the y-co-ordinates of corresponding image points are equal (*i.e.* $y_{r1i} = y_{r2i}$). If image point co-ordinates are sorted in ascending order of y-co-ordinates, potential correspondences can be established without an extensive searching through the image point co-ordinates.
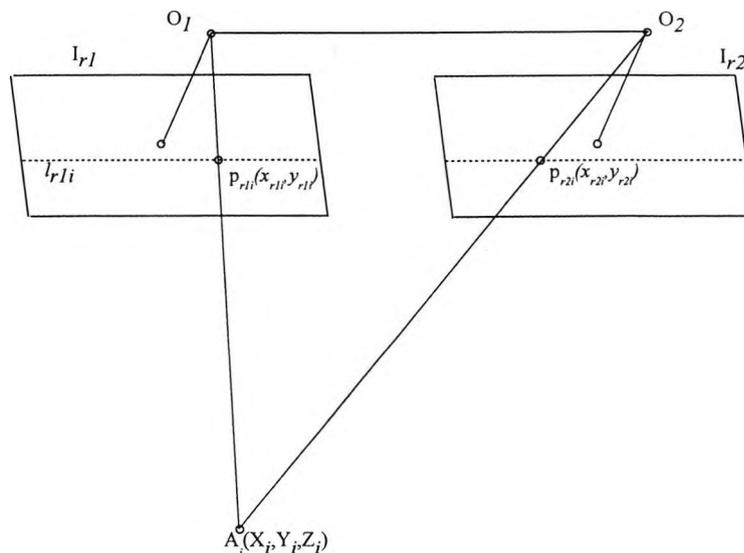
Figure 3.6 Parallel optical axis images

3.3.1.4 Consideration of ambiguities

Ambiguities occur when multiple image points are present on the epipolar lines (Figure 3.7). These occur when multiple *3-D* object points A_1 to A_5 fall on a single epipolar plane. In practice, the errors in the imaging system cause these points to fall near the ideal epipolar lines l_1 and l_2. Due to the consideration of a search band it is likely that these points fall within the band and cause ambiguities.
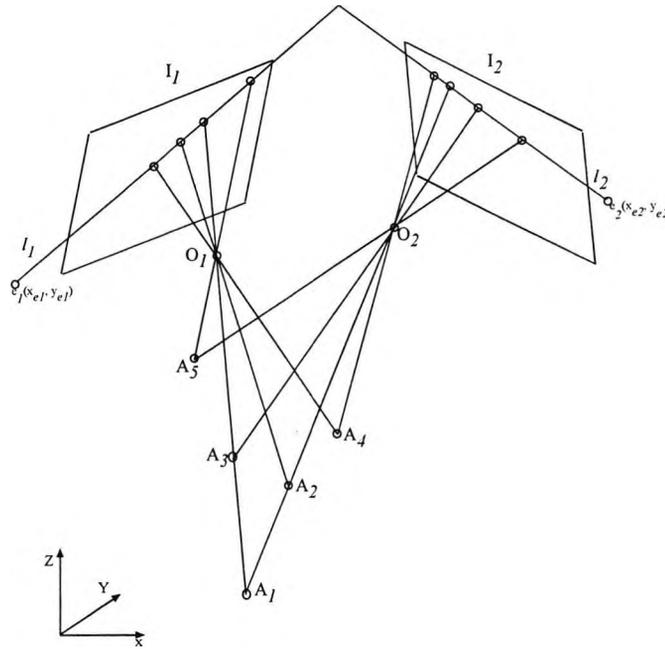
Figure 3.7 Ambiguities due to multiple co-planar *3-D* points

If the object space information is available some ambiguities may be solved intuitively using analytical constraints such as a left-to-right rule. This rule uses the fact that the left most point in one image should be the left most point in the other image. The epipole in each image can be taken as a local reference. The distance from the epipole to each point along the epipolar line can be used as a measure to identify the left to right order of image point location. For this technique to be applied successfully, all *3-D* points should appear on both images.

3.3.1.5 Epipolar geometry between more than two views

The epipolar geometry that exists between a pair of images can be extended to three convergent axis images. Three images can be considered as a combination of three stereo pairs I_1/I_2, I_2/I_3, and I_3/I_1 (Figure 3.8). It is assumed that the absolute or relative

orientation of each stereo pair is known. In this configuration, there are six epipoles e_{12}, e_{21}, e_{23}, e_{32}, e_{31}, and e_{13}. The plane formed by three optical centres O_1, O_2, and O_3 is termed the trifocal plane (Faugeras and Robert, 1996) which is an epipolar plane simultaneously for all three cameras.
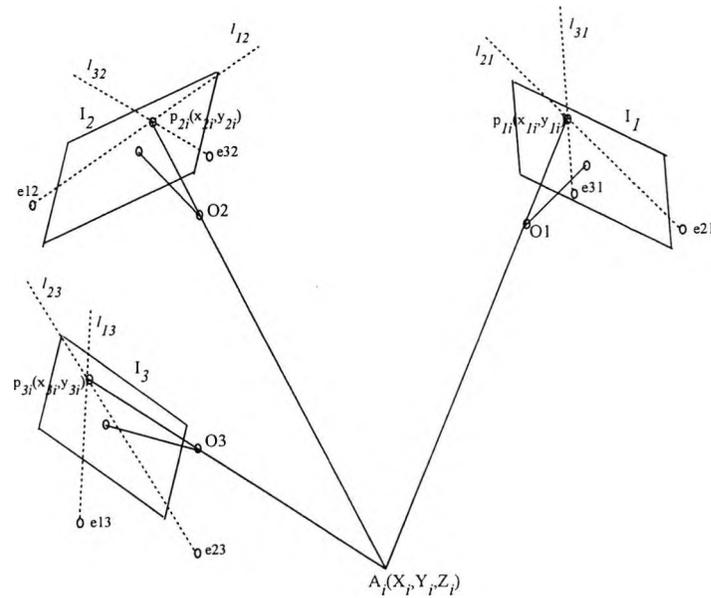
Figure 3.8 Application of epipolar constraint to three convergent axis images

Images of 3-D point $A_i(X_i, Y_i, Z_i)$ in images I_1, I_2, and I_3 are $p_{1i}(x_{1i}, y_{1i})$, $p_{2i}(x_{2i}, y_{2i})$, and $p_{3i}(x_{3i}, y_{3i})$. Considering the 3-D point $A_i(X_i, Y_i, Z_i)$ and image points p_{1i}, p_{2i}, and p_{3i}, the epipolar geometry between image pairs I_1/I_2, I_2/I_3, and I_3/I_1 forms three epipolar planes whose intersections with image planes create three pairs of conjugate epipolar lines l_{12}/l_{21}, l_{23}/l_{32}, l_{31}/l_{13} respectively. Under ideal geometric conditions, pairs of epipolar lines (l_{21}, l_{31} on image I_1; l_{12}, l_{32} on image I_2; and l_{13}, l_{23} on image I_3) should intersect on image points p_{1i}, p_{2i}, and p_{3i} hence forming a closed loop of points. A further geometric relationship is that each image point is a common point for the image plane and the two epipolar planes.

The epipolar geometry can be extended to four convergent axes images. Figure 3.9 illustrates epipolar geometry between four images. As in the case of three images, four images can be considered as a combination of six stereo pairs. Considering a 3-D point $A_i(X_i, Y_i, Z_i)$ and image points p_{1i}, p_{2i}, p_{3i}, and p_{4i}, the epipolar geometry between image pairs I_1/I_2, I_2/I_3, I_3/I_4, and I_4/I_1 forms four epipolar planes whose intersections with image planes creates four pairs of conjugate epipolar lines l_{12}/l_{21}, l_{23}/l_{32}, l_{34}/l_{43}, and l_{41}/l_{14}

respectively. These epipolar lines intersect on points p_{1i}, p_{2i}, p_{3i}, and p_{4i} forming a closed loop.
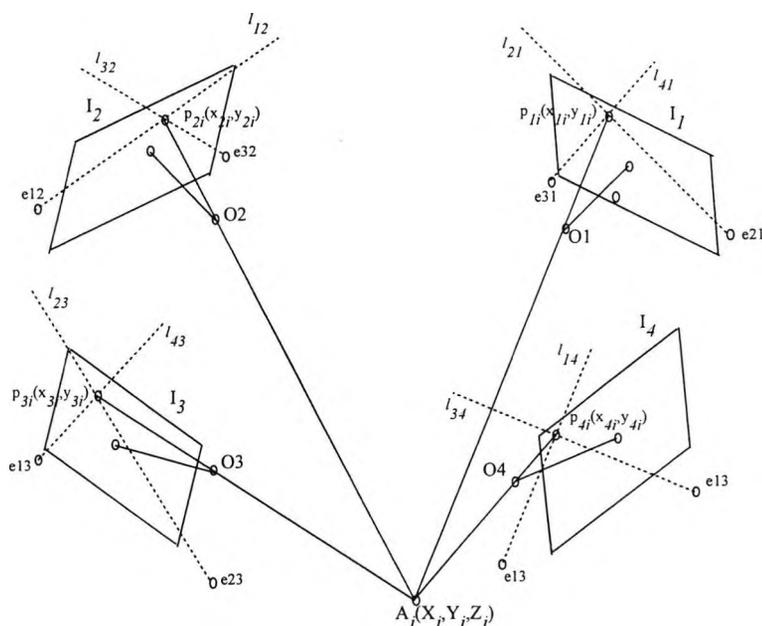
Figure 3.9 The application of epipolar constraint to four convergent axis images

The epipolar geometry can be extended to an arbitrary number of views. The epipolar constraint complemented with other geometric techniques have been used to establish correspondences. Number of algorithms which use these techniques were developed over the years due to both computer vision and photogrammetric research. Some of these algorithms are discussed in sections 3.4 and 3.5.

3.4 Correspondence algorithms in close-range photogrammetry

In all high precision close-range photogrammetric applications (because of the use of special illumination techniques, aperture settings, and the use of targets with retro-reflective characteristics) the images acquired often only consist of the intensity information of the targets. The subpixel locations of these targets are then computed. In order to compute *3-D* locations, the corresponding targets, now their subpixel locations (image points), should be identified. The camera interior parameters are estimated as an off-line process and the exterior parameters are determined to the highest possible accuracy. This section describes number of techniques that have been used in close-range applications.

3.4.1 Epipolar line algorithm

The epipolar line intersection method is commonly used technique used to establish multiple view image point correspondences. Mass, 1992 developed a three and four view correspondence method for use in *3-D* particle tracking velocimetry. Probabilistic models for image space ambiguities, with the knowledge of the density of object space particles, were developed for various three view configurations. The principle of epipolar line intersection was used for the establishment of correspondences. Figure 3.10 illustrates a three view configuration. Consider point p_{1i} *(i=1....n)* and epipolar line l_{21i} in image I_1. Its conjugate epipolar line in image I_2 *(l_{12i})* has potentially corresponding points p_{21}, p_{22}, and p_{23}. Similarly, consider point p_{1i} and epipolar line l_{31i} in image I_1. The conjugate epipolar line in image I_3 *(l_{13i})* has potentially corresponding points p_{31}, p_{32}, p_{33}. The conjugate epipolar lines belonging to points p_{21}, p_{22}, p_{23} intersect the epipolar line l_{13} on image I_3 near points p_{31}, p_{32}, and p_{33}. Mass, 1992 suggests that there is a high probability that only one of the intersecting points is closer to the corresponding image point *(p_{31})*.
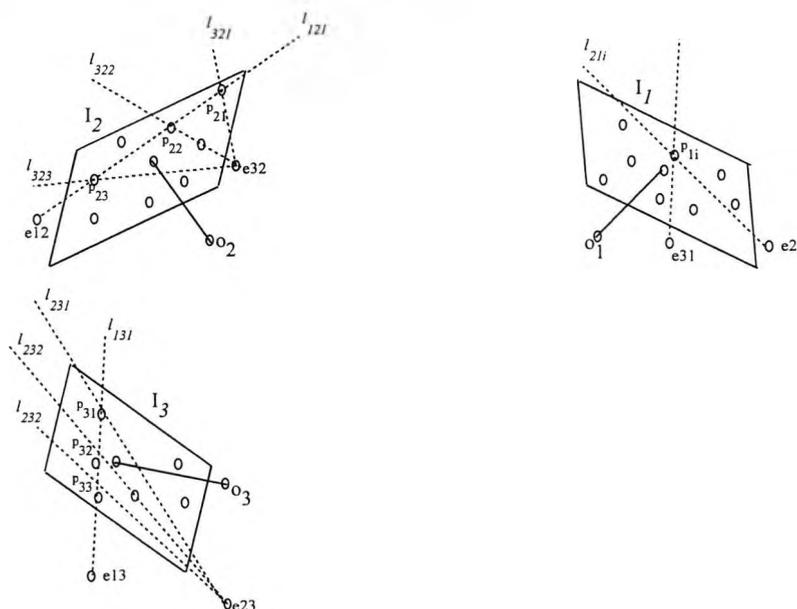
Figure 3.10 Intersection of epipolar lines on third view

A version of the epipolar line method was implemented for the purpose of this thesis. In this implementation multiple view configuration was considered as a combination of stereo pairs. The stereo pair correspondences are established as discussed in section 3.3.1. Image point identity based search is then carried out to distribute

correspondences. In this thesis this technique is referred as epipolar line (*EL*) algorithm from this point onwards.

3.4.2 *3-D* space intersection algorithm

This algorithm uses the epipolar constraint in the *3-D* object space rather than in the *2-D* image space (Chen, 1995). It uses the fact that rays projected from corresponding *2-D* image points should intersect at a single point which is the corresponding *3-D* world point. However, due to the errors of the imaging system these rays do not normally intersect at a point. Instead they intersect with a small perpendicular distance between them (Figure 3.11). The distances between intersecting rays that originate from corresponding points are smaller compared to those between non-corresponding points provided camera exteriors are accurate and image point co-ordinates are corrected for lens distortions. Hence, corresponding points can be distinguished by comparing computed distance with a pre-determined threshold.
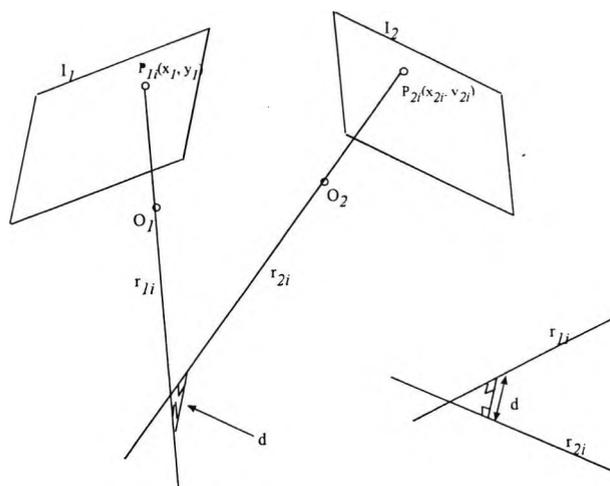
Figure 3.11 Minimum distance between a pair of rays

Points $p_{1i}(x_{1i}, y_{1i})$ and $p_{2i}(x_{2i}, y_{2i})$ in images I_1 and I_2 are two arbitrary points. $O_1(X_{01}, Y_{01})$ and $O_2(X_{01}, Y_{01})$ are optical centres of images I_1 and I_2. Rays r_{1i} and r_{2i} projected into *3-D* space from each point intersect with a perpendicular distance d between them which can be calculated. By rearranging collinearity equations (Equation 2.14), the equation of the *3-D* lines representing rays r_{1i} and r_{2i} can be written as,

$$\frac{(X - XO1)}{p1} = \frac{(Y - YO1)}{q1} = \frac{(Z - ZO1)}{r1},$$

$$\frac{(X - XO2)}{p2} = \frac{(Y - YO2)}{q2} = \frac{(Z - ZO2)}{r2}$$ (3.10)

where,

$$p1 = m1_{11}x + m1_{21}y - m1_{31}f_1, \quad q1 = m1_{12}x + m1_{22}y - m1_{32}f_1, \quad r1 = m1_{13}x + m1_{23}y - m1_{33}f_1,$$

$$p2 = m2_{11}x + m2_{21}y - m2_{31}f_2, \quad q2 = m2_{12}x + m2_{22}y - m2_{32}f_2, \text{ and } r2 = m2_{13}x + m2_{23}y - m2_{33}f_2$$

Hence, the minimum distance between rays r_{1i} and r_{2i} is given by,

$$d = \begin{vmatrix} (X02 - X01) & (Y02 - Y01) & (Z02 - Z01) \\ p1 & q1 & r1 \\ p2 & q2 & r2 \end{vmatrix} \Big/ \sqrt{\begin{vmatrix} p1 & q1 \\ p2 & q2 \end{vmatrix}^2 + \begin{vmatrix} q1 & r1 \\ q2 & r2 \end{vmatrix}^2 + \begin{vmatrix} r1 & p1 \\ r2 & p2 \end{vmatrix}^2}$$ (3.11)

A threshold for the minimum distance value that distinguishes corresponding and non-corresponding points can be established using prior knowledge of *3-D* location accuracy.

Chen, 1995 considers all images simultaneously. An image that views all object space points is selected as the master image. A ray is projected into *3-D* space from a point in this image. The perpendicular distances to all the other rays projected from each image are computed and are compared against a threshold value. Rays that generate distances which are smaller than the threshold are selected as those belonging to corresponding image points.

Implementation of this algorithm by Chen, 1995 considered the use of approximate camera exterior parameters. Using a few control points, approximate exterior parameters are obtained by resection. This enables the establishment of correspondences between fewer image points. The least squares adjustment was then carried out which resulted in more accurate camera exterior parameters. Correspondences were established for the previously uncorresponded image points resulting more corresponded points. Hence, next adjustment produces even more accurate exteriors. This iterative process was carried out until all possible image points were corresponded.

A version of this technique was implemented. Multiple view configurations were considered as discussed in section 3.4.1 for the *EL* algorithm. In this thesis this technique is referred as *3-D* space intersection *(3DSI)* algorithm from this point onwards.

3.4.3 Back projection or drive-back algorithm

This is a commonly used technique in multiple view close-range applications and has two implementations based on whether *3-D* object space points are unknown or known (Fraser, 1997).

When *3-D* object space points are unknown, two images that view all *3-D* points of interest are taken as master images and the point correspondences are established. A *3-D* reconstruction is made using these correspondences. Rays are projected into other views from each computed *3-D* point. Image points that fall within a pre-determined tolerance from the projected points are taken as correspondences. The efficiency of this method depends on the correct establishment of the correspondences between master image pair. It has been implemented successfully by at least one manufacturer of photogrammetric systems (Fraser, 1997).

When *3-D* object space point locations are known, rays can be reprojected into each view from each known *3-D* point and correspondences can be established in the same way.

3.4.4 Epipolar line slope algorithm

The Epipolar line slope algorithm introduced by Sabel *et al*, 1993 uses the line slope to establish image point correspondences in real-time. In the initialisation step of the algorithm, a relationship between the slopes of conjugate epipolar lines is established. Hence, for a selected point in one image, the slope of the conjugate epipolar line in the other image can be computed with lower computational cost compared to the usual epipolar line determination method (Section 3.3.1.1).

During the initialisation step, a set of epipolar lines is generated in one image with regular angular spacing between them and their slopes with respect to the local co-ordinate system are computed. The slopes of the conjugate epipolar lines in the other image can be obtained as discussed in section 3.3.1.1. Sabel, 1993 pointed out that the

relationship between the slopes can be represented by a 2nd order polynomial. The next step is to use the least squares polynomial fitting to obtain an optimum polynomial that represent a relationship between conjugate slopes. The method of using a 2nd order polynomial is discussed here.

Suppose that the following is a set of pairs of conjugate epipolar line slopes,

$$(s_1, s'_1), \ (s_2, s'_2), \ (s_3, s'_3), \ .... \ , \ (s_i, s'_i), \ ...., \ (s_m, s'_n)$$

The mth order polynomial relationship that expresses s' in terms of s can be written as,

$$p(s') = b_0 + b_1.s + b_2.s^2 + ...+ b_m.s^m \tag{3.12}$$

where $b_0, b_1, ...b_m$ are unknowns which can be evaluated by minimising the residuals between actual and calculated (using polynomial) slopes in a second image.

The sum of the squares of the residuals (R) can be stated as,

$$R = \sum_{i=1}^{n} \left(s'_i - p(s_i') \right)^2 \tag{3.13}$$

The least squares principle states that sum of the squares of residuals should be minimised in order to find optimum coefficients. Hence, R should be minimised. If the polynomial required is of 2nd order as considered by Sabel, 1993, the coefficients that minimises the residuals occur when,

$$\frac{\partial R}{\partial b_0} = 0, \quad \frac{\partial R}{\partial b_1} = 0, \quad \frac{\partial R}{\partial b_2} = 0$$

Using above conditions, following equations can be obtained,

$$nb_0 + b_1 \sum_{i=1}^{n} s_i + b_2 \sum_{i=1}^{n} s_i^2 = \sum_{i=1}^{n} s'_i$$

$$b_0 \sum_{i=1}^{n} s_i + b_1 \sum_{i=1}^{n} s_i^2 + b_2 \sum_{i=1}^{n} s_i^3 = \sum_{i=1}^{n} s_i s'_i$$

$$b_0 \sum_{i=1}^{n} s_i^2 + b_1 \sum_{i=1}^{n} s_i^3 + b_2 \sum_{i=1}^{n} s_i^4 = \sum_{i=1}^{n} s_i^2 s'_i \tag{3.14}$$

In matrix form,

$$Ax = B$$

where,

$$A = \begin{bmatrix} n & \sum_{i=1}^{n} s_i & \sum_{i=1}^{n} s_i^{2} \\ \sum_{i=1}^{n} s_i & \sum_{i=1}^{n} s_i^{2} & \sum_{i=1}^{n} s_i^{3} \\ \sum_{i=1}^{n} s_i^{2} & \sum_{i=1}^{n} s_i^{3} & \sum_{i=1}^{n} s_i^{4} \end{bmatrix} \qquad x = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad b = \begin{bmatrix} \sum_{i=1}^{n} s'_i \\ \sum_{i=1}^{n} s_i s'_i \\ \sum_{i=1}^{n} s_i^{2} s_i \end{bmatrix}$$

Hence, the unknown coefficients (x) of a 2nd order polynomial can be obtained as,

$$x = A^{-1} B \qquad\qquad (3.15)$$

Following example illustrates the establishment of a 2nd order polynomial relationship between the conjugate epipolar line slopes. Table 3.1 gives exterior parameters of a stereo pair. Figure 3.12 illustrates the relationship between conjugate epipolar line slopes in graphical form.

camera	X0 (mm)	Y0 (mm)	Z0 (mm)	omega (degrees)	phi (degrees)	kappa (degrees)	focal length (mm)
1000	- 1121.3477	-1128.1358	1974.5010	30.5977	-25.4267	104.7095	16
1001	975.8044	-1091.4825	1939.4547	27.9820	23.7131	35.7754	16

**Table 3.1 Exterior orientation parameters and focal length**
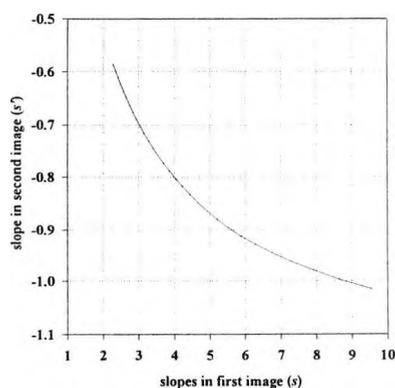
Figure 3.12 Illustration of the relationship between conjugate slopes

Using the least squares technique, the fitting 2nd order polynomial was obtained as,

$$p(s') = (-0.2839 - 0.1639s + 0.0094s^2) \hspace{2cm} (3.16)$$

Hence, for any slope (s) in first image the conjugate slope (s') in the second image can be obtained.

Further work was carried out to investigate this algorithm. For certain instances of the relative orientations between two images, it is not possible to obtain a polynomial relationship. One of these instances is when the value of s' reaches infinity for a particular value of s. It is possible that a mathematical relationship between s and s' can be obtained for these kinds of curves but it is a more complicated process. Hence, the pre-calibration process could become more complicated. However, by careful placement of cameras and determination of the polynomial, the computational advantages of this algorithm can still be exploited. The slope relationship between stereo pairs was also implemented for three view point correspondence establishment.

3.5 Correspondence algorithms in computer vision

In high precision close-range applications, it is customary to estimate camera interior parameters (principal point offset, principal distance, and lens distortion parameters) off-line. Regardless of the number of cameras, accurate exterior parameters (location and rotation) are also estimated. But in most low-precision computer vision applications, it is not always possible to have the camera exterior parameters estimated accurately using a calibration object. Hence, it may be necessary to estimate these parameters on-line using the information available in the scene. Some of the recently developed algorithms are discussed in this section.

3.5.1 Essential matrix

Relative orientation has been a technique widely used in photogrammetry which estimates exterior orientation of one camera with respect to the local co-ordinate system of another (section 2.3.3). Interior parameters are assumed to be estimated off-line. Using a minimum of five corresponding sets of points in a stereo pair, relative orientation can be estimated. Once relative orientation is known the epipolar constraint can be used to establish image point correspondences. Along the lines of this technique, Longuet-Higgins, 1981 developed the essential matrix which can be used to map camera co-ordinates of the image points in one image to corresponding epipolar

lines in another image. The camera interior parameters are required to convert pixel co-ordinates to camera co-ordinates. Using a minimum of eight pairs of corresponding points in a stereo image pair, a *(3x3)* essential matrix can be estimated and can be used to recover complete epipolar geometry between two views.

Figure 3.13 illustrates the image pair I_1 and I_2 whose optical centres are O_1 and O_2. Images of *3-D* point P_t on images I_1 and I_2 are at points p_{1i} and p_{2i} which can be represented by vectors r_1 and r_2 with respect to each local co-ordinate system. Similarly, the co-ordinates of point P_i with respect to local co-ordinate systems are (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) which can be represented by vectors R_1 and R_2. The base vector is b.
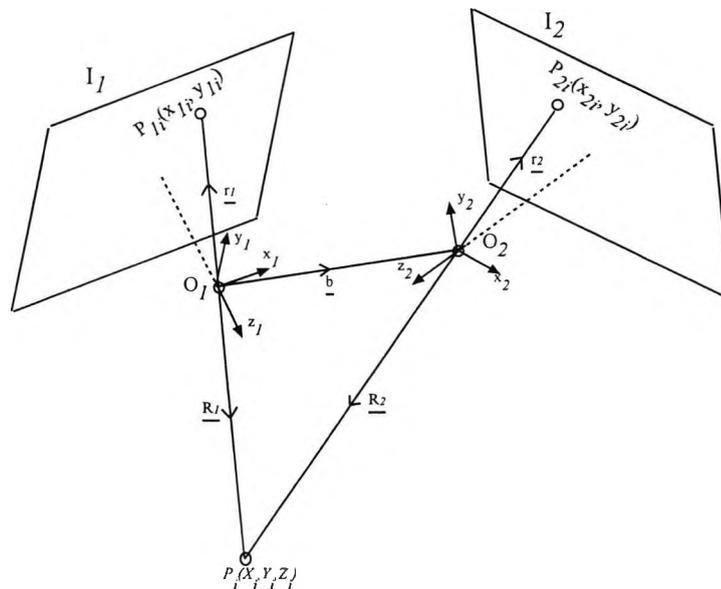
Figure 3.13 Illustration of the coplanarity condition

The vectors $R1$, $R2$, and b are coplanar. Hence, the triple scalar product is equal to zero. That is,

$$\left(R_1 - b \right) \cdot b \times R_1 = 0 \qquad (3.17)$$

It can be stated that

$$R_2 = R\left(R_1 - b \right) \qquad (3.18)$$

where, R is the orthogonal rotation matrix and $R^{-1} = R^t$.

Hence,

$$R^t \underline{R_2} = (\underline{R_1} - \underline{b}) \qquad (3.19)$$

The vector product $\underline{b} \times \underline{R_1}$ can be written as a product of skew symmetric matrix S and R_1. Hence,

$$\underline{b} \times \underline{R_1} = SR_1{}^t$$

where,

$$S = \begin{bmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{bmatrix}$$

Rearranging Equation (3.17) in matrix form and substituting for $\underline{b} \times \underline{R_1}$,

$$(\underline{R_1} - \underline{b})^t \, S\underline{R_1} = 0 \qquad (3.20)$$

Substituting (3.19) into (3.20),

$$(R' \underline{R_2})^t \, S\underline{R_1} = 0$$

Hence,

$$\underline{R_2}' \, E \, \underline{R_1} = 0 \qquad (3.21)$$

where, $E = RS$.

Dividing both sides of Equation 3.21 by $Z_1 Z_2$,

$$\underline{r_2}^t \, E \, \underline{r_1} = 0 \qquad (3.22)$$

The matrix E is termed the essential matrix. The nine unknowns of the essential matrix can be evaluated given eight pairs of corresponding points which is usually called 8-point algorithm. Once the essential matrix has been estimated, for a given point in one

image the corresponding epipolar line in the other image can be computed. The rotation matrix and the translational vector can be derived using this matrix.

3.5.2 Fundamental matrix

A further development by Luong and Faugeras,1996 is the fundamental matrix which can be used to map pixel co-ordinates in one image to the corresponding epipolar lines of another image. Both camera exterior and interior parameters are included in the fundamental matrix. The complete epipolar geometry can be recovered using this matrix.

If $\overline{r1}$ and $\overline{r2}$ are the points in pixel co-ordinates corresponding to $\underline{r_1}$ and $\underline{r_2}$ camera co-ordinates, their relationship can be stated as,

$$\underline{r1} = M_1^{-1}\overline{r1} \qquad\qquad \underline{r2} = M_2^{-1}\overline{r2} \qquad\qquad (3.23)$$

where M_1 and M_2 are intrinsic parameter matrices of cameras for mapping pixel co-ordinates to the camera co-ordinates.

Substituting Equations 3.23 into Equation 3.22,

$$\overline{r2}'\, F\, \overline{r1} = 0 \qquad\qquad (3.24)$$

where $F = M_2^{-1\,t} E\, M_1^{-1}$ is termed the fundamental matrix.

The fundamental matrix can be estimated in the same way as the essential matrix by using pixel co-ordinates of eight pairs of corresponding image points. When the exterior or interior parameters are unknown, the fundamental matrix is the key concept for recovering epipolar geometry.

The fundamental matrix theory has been extensively studied by computer vision researchers (Luong and Faugeras, 1996; Faugeras and Robert, 1996). Various methods of determining the matrix have been studied and stability analysis carried out. The theory of the fundamental matrix is complete for two views (Faugeras and Robert, 1996) and is being extended to more views. The fundamental matrix based techniques have been developed to predict points, lines, and curves in a third view given

correspondences in another two views (Faugeras and Robert, 1996). For the three view case, fundamental matrices can be estimated for each stereo pair separately but any relationship between these matrices is not known.

3.5.3 Multiple baseline stereo

The multiple baseline stereo algorithm (Okutomi and Kanade, 1993) uses multiple stereo pairs with various base lines to obtain precise distance measurements and to avoid ambiguity problems. When the baseline length (B) is smaller, stereo depth estimation will be less accurate due to the narrower triangulation angle. Correspondences can be established more accurately if radiometric or geometric distortion is minimal. For more accurate distance estimates, a longer baseline length is necessary. However, the chances of making incorrect correspondences are higher due to the less overlapping images of the object space and due to the higher radiometric and geometric distortions. The multiple baseline algorithm provides an optimum solution with respect to accuracy of depth estimates and correctness of correspondences.

Figure 3.14 illustrates an stereo pair whose optical centres are O_1 and O_2 and the baseline length is B. The local co-ordinate systems of each image are parallel and the focal length of each image is f.
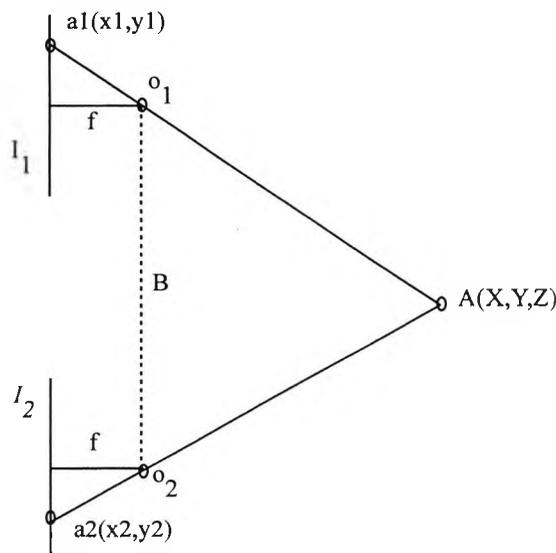

Figure 3.14 Parallel axes stereo pair

The images of *3-D* point *A(X,Y,Z)* on images I_1 and I_2 are at points $a_1(x_1,y_1)$ and $a_2(x_2,y_2)$. The disparity $d = (x_1 - x_2)$, the focal length (*f*), baseline length (*B*), and the depth (*Z*) are related by,

$$\frac{Z}{B} = \frac{f}{(x_1 - x_2)}$$

$$\frac{d}{B} = \frac{f}{Z} \qquad (3.25)$$

The Equation 3.25 shows that the ratio between the disparity and baseline length is a constant for a particular *3-D* point since depth is a constant. This idea was exploited in the multiple baseline configuration. A number of cameras are placed along a line to create several stereo pairs (Figure 3.15). The placement of cameras may be along the *x* or *y* directions of the base camera $C_{(0,0)}$ whose axis system is parallel to that of each secondary camera. The base camera is paired with each camera (in *x* or *y* direction) to make multiple baseline stereo pairs. For each stereo pair the depth is the same hence, the ratio between disparity and baseline length is the same.
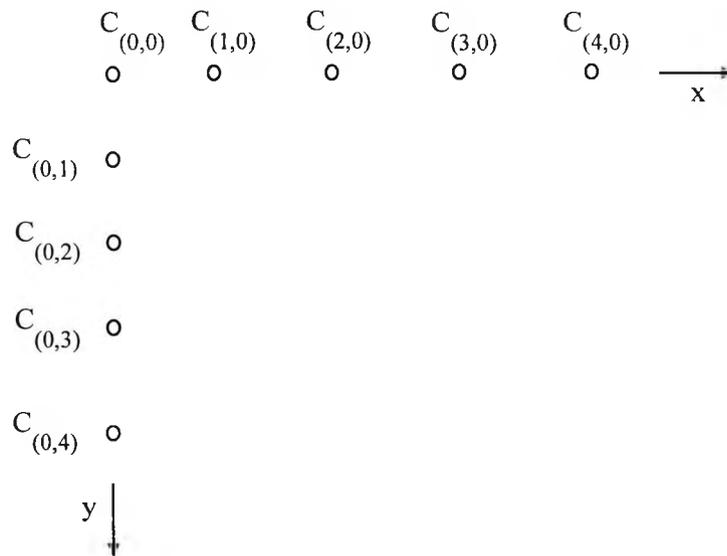

Figure 3.15 Multiple baseline stereo pairs

The intensity information is used in this algorithm. A point is selected in the base image and a small window is created around it. This window is slid along the other parallel and collinear epipolar lines in the other image and the sum of squared differences *(SSD)* of intensities are computed for each disparity value. *SSDs* are

computed for each image *(SSD₁, SSD₂, SSD₃, SSD₄)*. When each *SSD* is plotted against inverse of the depth there is a minimum but it is not very distinct. When the baseline is longer, more minima exist due to the ambiguities. Hence, *SSDs* of each image are added to produce a sum of *SSDs (SSSD)*. Okutomi and Kanade, 1993 showed that when the *SSSD* is plotted against inverse depth, it clearly illustrates a minimum at the correct disparity. Hence, corresponding points in secondary images can be identified and the accurate depth estimate can be obtained using the redundant observations.

3.6 Summary of the chapter

In this chapter a survey of correspondence techniques developed by computer vision and close range photogrammetric research has been presented and discussed. The feature extraction techniques used for high precision applications were discussed in detail.

The images of natural features on objects cannot be located accurately. Hence, artificial retro-reflective targets are used to signalise object points of interest. With the use of special illumination and appropriate camera apeture set up these target images are distinguishable from the background clutter. Target images can be located accurately.

The epipolar geometry is the widely used projective geometric relationship. Provided relative or absolute orientation of a stereo pair are given epipolar constraint can be used to establish correspondences. This relationship can also be extended to more than two views by considering multiple views as a combination of stereo pairs. A number of techniques that use the epipolar geometry as the basis were discussed. The new developments in computer vision research, the essential and the fundamental matrices were also discussed. In close-range measurement applications approximate or accurate camera exterior parameters are always known prior to establishing correspondences. Hence, fundamental and essential matrices may be estimated using the known epipolar geometry. These matrices may then be used for determining the conjugate epipolar lines instead of the method described in section 3.3.1.1.

Chapter 4

Multiple view point correspondences using rectification - *PEL* algorithm

In this chapter, the development of an algorithm based on image rectification principles for fast multiple view point correspondence establishment is discussed. The algorithm uses parallel epipolar lines for the establishment of correspondences, hence it is termed the *PEL* algorithm. The algorithm considers images acquired from multiple views as a set of stereo pair combinations. Potential correspondences are established in each stereo pair combination following which an efficient search based on image point identities is performed to establish sets of corresponding points. The use of collinear epipolar lines and a technique for determining search band in rectified image space results in reduced searching and computations hence, faster establishment of correspondences.

Convergent Vs parallel optical axis stereo point correspondences

Figure 4.1 illustrates a pair of convergent axis images I_1 and I_2 where O_1 and O_2 are optical centres. The epipoles of images I_1 and I_2 are points $e_1(x_1, y_1)$ and $e_2(x_2, y_2)$. The points $p_{1i}(x_{1i}, y_{1i})$ and $p_{2i}(x_{2i}, y_{2i})$ are images of the *3-D* world point $A_i(X_i, Y_i, Z_i)$ on images I_1 and I_2. Considering point $p_{1i}(x_{1i}, y_{1i})$ that lies on epipolar line l_{1i} in image I_1, according to the epipolar constraint the corresponding point $p_{2i}(x_{2i}, y_{2i})$ in image I_2 can be found within the band b_{2i}.
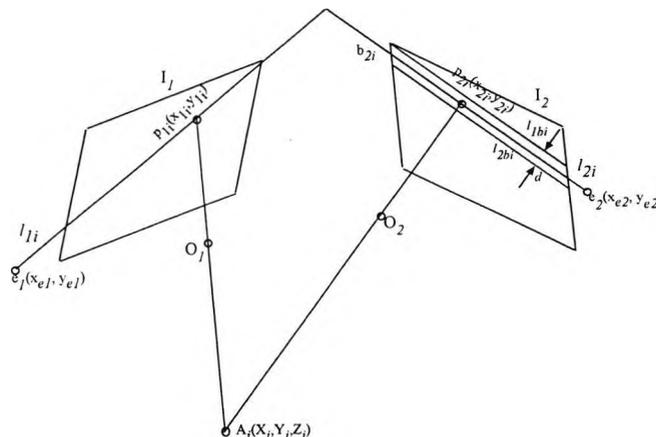

Figure 4.1 Application of epipolar constraint in image space

In order to find a corresponding point in image I_2 for a selected point in image I_1, the first step is the determination of an appropriate search band. Secondly, each point in image I_2 should be checked to find if it falls within this search band. Suppose there are n points in each image. For n points in image I_1, n search band determinations and n^2 checks should be performed which amounts to a significant number of computations to be carried out in real-time. Also, for an increasing number of points, the time for the correspondence establishment increases non-linearly.

An alternative method of applying the epipolar constraint was discussed in section (3.4.2) where the epipolar constraint is applied in *3-D* object space rather than *2-D* image space. In this method, the fact that a pair of rays projected into *3-D* space from a pair of corresponding points should intersect with a small perpendicular distance between them is used (Figure 4.2). Hence, corresponding points can be distinguished as their perpendicular distance of intersection is smaller than non-corresponding candidates.
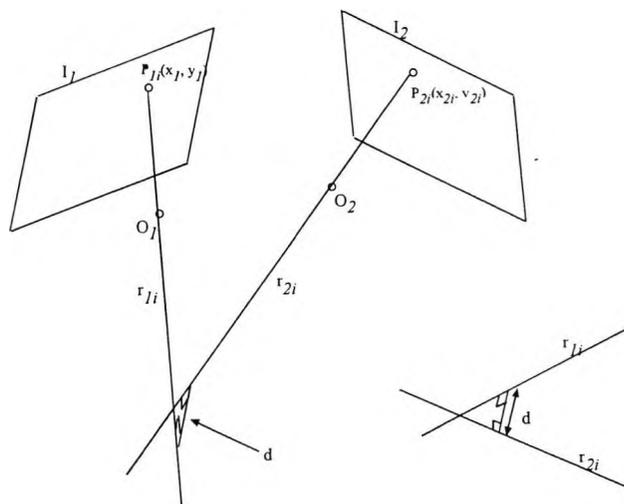
Figure 4.2 Application of epipolar constraint in *3-D* object space

If n points are present in each image, n^2 minimum distance computations should be performed which requires a larger number of computations compared to the case of the application of the constraint in *2-D* image space. Hence, this method exhibits a much steeper non-linear rise in time required to establish correspondences for an increasing number of image points.

The advantage of the *PEL* algorithm is that the searching required to find correspondences in stereo pairs is reduced. Assuming a stereo pair with n points in each image, the *PEL* algorithm require searching which is approximately proportional to n. Hence, the time requirement increases approximately linearly for increasing number of points. Figure 4.3 illustrates a pair of convergent images I_1/I_2 and the corresponding rectified images I_{r1}/I_{r2}. The optical centres O_1 and O_2 are common to convergent and rectified images. Points p_{1i} and p_{2i} are the images of *3-D* world point A_i on images I_1 and I_2. Due to their position and orientation, the epipolar plane formed by points A_i, O_1, and O_2, intersect rectified image planes along collinear epipolar lines l_{r1i} and l_{r2i}. Images of the same *3-D* point A_i on images I_{r1} and I_{r2} (*i.e.* points p_{r1i} and p_{r2i}) should ideally lie on these epipolar lines. Hence, for a selected point in image I_{r1}, corresponding point can be found by searching along the conjugate collinear epipolar line.
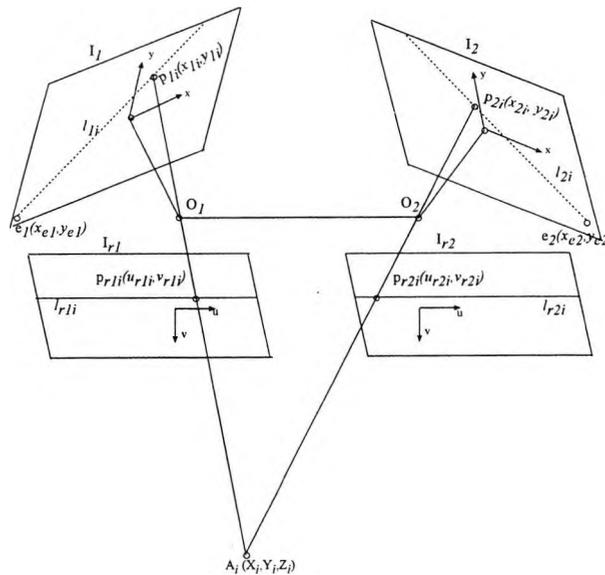
Figure 4.3 Convergent and corresponding rectified image pairs

The *PEL* correspondence process begins with the transformation of points from a pair of convergent optical axes images to a pair of parallel axes rectified images. Image point location errors that exist in convergent images are automatically transferred into rectified images. Hence, a search is required to be performed along a band that is on either side of the epipolar line. The rest of this chapter describes each stage of the development of algorithm. First, the theoretical foundation is developed at stereo pair level. Secondly, the implementation in multiple view configuration is described.

4.1 Determination of the parameters of algorithm

During the initialisation process, the unknown parameters of the algorithm are determined. Considering a stereo pair, the rectified image plane, the rotational parameters of rectified images, and the convergent to rectified image point transformation parameters (*i.e.* rectification parameters) are determined. Once the rectification parameters are known, image points in convergent images can be transformed into rectified images. The next step is the establishment of correspondences. As in convergent image space, a search band is required to constrain the search for corresponding points. Using the geometric relationship that exist between convergent and corresponding rectified image planes, search band parameters are also determined during the initialisation.

4.1.1 Rectified image plane, focal length, and rotational parameters

Rectified image plane:

This algorithm was developed with the rectified image plane positioned perpendicular to the reference plane formed by points O_1, O_2, and reference point R_r (Figure 4.4). The rectified image plane is parallel to the baseline O_1O_2. The perpendicular distance to this plane from baseline O_1O_2 is the focal length of the rectified images. The orientation of the reference plane is used to determine rotational parameters of the rectified images of each stereo pair combination of the camera network.
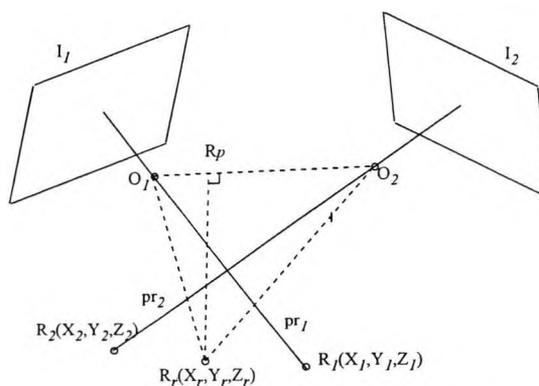

Figure 4.4 Reference plane $O_1O_2R_r$

The principal rays pr_1 and pr_2 of images I_1 and I_2 intersect the *XY*-plane of the world co-ordinate system at points $R_1(X_1,Y_1,Z_1)$ and $R_2(X_2,Y_2,Z_2)$. Reference point $R_r(X_r,Y_r,Z_r)$ is the mid-point of the line joining points R_1 and R_2.

Point R_p is on the base line O_1O_2. The co-ordinates of any point on line O_1O_2 can be written in parametric form as,

$$X = X0_1 + t(X0_1 - X0_2)$$

$$Y = Y0_1 + t(Y0_1 - Y0_2)$$

$$Z = Z0_1 + t(Z0_1 - X0_2)$$ (4.1)

where t is a variable.

Vectors $\overline{O_1O_2}$ and $\overline{R_pR_r}$ are perpendicular hence, the scalar product between them is equal to zero. That is, $\overline{R_pR_r}.\overline{O_1O_2} = 0$.

$\therefore$ t can be evaluated as, $$t = \left(\frac{((x_2 - x0_1)(x0_2 - x0_1) + (y_2 - y0_1)(y0_2 - y0_1) + (z_2 - z0_1)(z0_2 - z0_1))}{((x0_2 - x0_1)^2 + (y0_2 - y0_1)^2 + (z0_2 - z0_1)^2)} \right)$$

(4.2)

Knowing t, the co-ordinates of point $P(X_p, Y_p, Z_p)$ can be calculated.

A unit vector $\bar{u}$ which is perpendicular to plane $O_1R_rO_2$ can be determined by taking vector product between vectors $\overline{O_1R_r}$ and $\overline{O_2R_r}$ as,

$$\bar{u} = \left(\frac{\overline{O_1R_r} \times \overline{O_2R_r}}{\left| \overline{O_1R_r} \right| \left| \overline{O_2R_r} \right| \sin(O_1R_rO_2)} \right)$$ (4.3)

where angle $O_1R_rO_2$ is obtained by taking scalar product between vectors $\overline{O_1R_r}$ and $\overline{O_2R_r}$ as,

$$angle(O_1R_rO_2) = \cos^{-1} \left(\frac{\overline{O_1R_r}.\overline{O_2R_r}}{\left| \overline{O_1R_r} \right| \left| \overline{O_2R_r} \right|} \right)$$ (4.4)

Focal length of rectified images:

Figure 4.5 illustrates a pair of convergent images I_1 and I_2 whose focal lengths are f_1 and f_2. The base line is O_1O_2 and rectified image planes are I_{r1} and I_{r2}. The focal length

of the rectified images f_r, is the perpendicular distance between the base line and the rectified image plane.
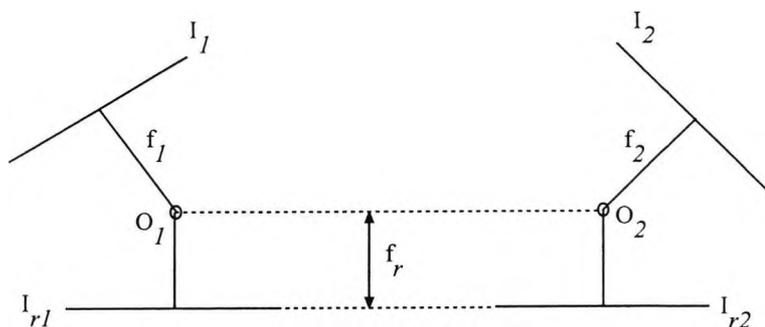
Figure 4.5 Focal length of rectified images

Rotational parameters:

With respect to plane $O_lR_rO_2$, the local co-ordinate systems of rectified images have orientations as illustrated in Figure 4.6. In general, it can be defined that the: **x-axis** is along the line created by intersection of rectified image plane and reference plane $O_lR_rO_2$; **y-axis** is perpendicular to plane $O_lR_rO_2$ and has the direction of unit vector $\bar{u}$ (Equation 4.3); **z-axis:** lies on the plane $O_lR_rO_2$ and has the direction of vector $\overline{R_pR_r}$.
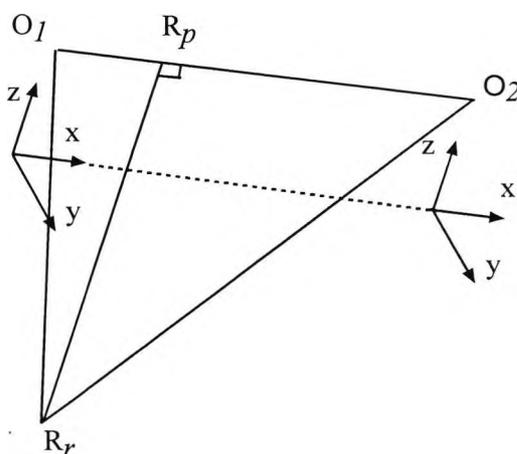
Figure 4.6 Co-ordinate systems on rectified images with respect to reference plane

Once the orientation of the reference plane $O_lR_rO_2$ is known, the world co-ordinate system *XYZ* can be rotationally transformed to obtain the same orientation as the

rectified image co-ordinate system xyz (Figure 4.7). Three rotational parameters ω, ϕ, and κ are determined by sequential rotations around X, Y, and Z axis.

From $\triangle ABR_r$, the magnitudes of AB and BR_r can be obtained as $|AB| = \left| Y_p - Y_r \right|$ and $|BR_r| = \left| Z_p - Z_r \right|$. Hence, the magnitude of AR_r is $|AR_r| = \left| \sqrt{\left(Y_p - Y_r \right)^2 + \left(Z_p - Z_r \right)^2} \right|$.

From $\triangle AR_pR_r$, magnitude of AR_p can be obtained as $\left| AR_p \right| = \left| X_p - X_r \right|$. Hence, ω and ϕ can be obtained as,

$$\omega = angle AR_r B = \tan^{-1}\left(\frac{AB}{BR_r} \right) \tag{4.5}$$

$$\phi = angle AR_r Rp = \tan^{-1}\left(\frac{AR_p}{AR_r} \right) \tag{4.6}$$

Let a unit vector perpendicular to plane AR_pR_r (Figure 4.7) be $\bar{v}$. Taking the vector product between vectors $\overline{R_p A}$ and $\overline{R_r A}$, $\bar{v} = \dfrac{\overline{R_p A} \times \overline{R_r A}}{\left| R_p A \right| \left| R_r A \right| \sin\left(angle R_p AR_r \right)}$ where angle $R_p AR_r$ $= 90^\circ$. Hence,

$$\bar{v} = \frac{\overline{R_p A} \times \overline{R_r A}}{\left| R_p A \right| \left| R_r A \right|} \tag{4.7}$$

κ is the angle between unit vectors $\bar{v}$ and $\bar{u}$. Each unit vector should be in a direction so that each makes an angle greater than 90° with vector $\overline{BR_r}$. If this angle is smaller than 90°, the direction of the particular unit vector needs to be reversed by multiplying it with -1. Taking the scalar product between unit vectors $\bar{u}$ and $\bar{v}$,

$$\kappa = \cos^{-1}\left(\frac{\bar{u}.\bar{v}}{\left| \bar{u} \right| \left| \bar{v} \right|} \right) \tag{4.8}$$

By considering the direction cosines of the unit vector $\bar{u}$ (*i.e.* orientation of plane $O_1 R_r O_2$), the signs of ω, ϕ, and κ can be determined for any pair of rectified images.
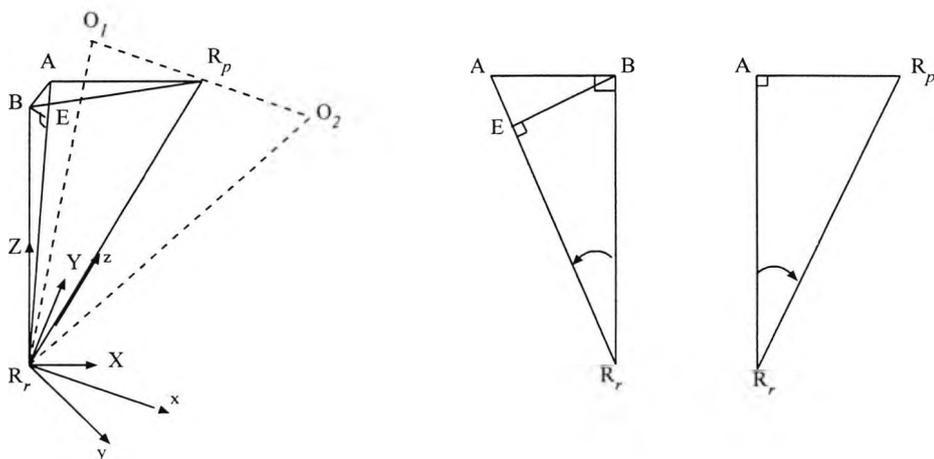
Figure 4.7 Transformation of *XYZ* to *xyz*

4.1.2 Rectification parameters

A point *(x,y)* in one image plane can be projectively transformed into a point *(u,v)* in another non-parallel image plane using the following projective transformation equations (Wolf, 1983),

$$u = \frac{ax + by + c}{gx + hy + 1}$$

$$v = \frac{dx + ey + f}{gx + hy + 1} \tag{4.9}$$

where a,b,c,d,e,f,g, and h are projective transformation or rectification parameters.

For a given stereo pair, a set of rectification parameters is required for each image. To determine each set of parameters, a minimum of four points is required, these being well distributed in each rectified image space whose corresponding points in the convergent images are known. In this algorithm, four corner points of each convergent image are projected onto the *XY*-plane of the world co-ordinate system and are then projected back onto each corresponding rectified image plane to obtain four corresponding points. If corner points are (x_1,y_1), (x_2,y_2), (x_3,y_3), and (x_4,y_4) and the transformed points are (u_1,v_1), (u_2,v_2), (u_3,v_3), and (u_4,v_4) a set of equations can be written as,

$$ax_1 + by_1 + c - gx_1u_1 - hy_1u_1 = u_1$$

$$ax_2 + by_2 + c - gx_2u_2 - hy_2u_2 = u_2$$

$$ax_3 + by_3 + c - gx_3u_3 - hy_3u_3 = u_3$$

$$ax_4 + by_4 + c - gx_4u_4 - hy_4u_4 = u_4$$

$$dx_1 + ey_1 + f - gx_1u_1 - hy_1u_1 = v_1$$

$$dx_2 + ey_2 + f - gx_2u_2 - hy_2u_2 = v_2$$

$$dx_3 + ey_3 + f - gx_3u_3 - hy_3u_3 = v_3$$

$$dx_4 + ey_4 + c - gx_4u_4 - hy_4u_4 = v_4 \qquad (4.10)$$

In matrix form,

$$A\,X = b \qquad\qquad (4.11)$$

where,

$$
A = \begin{vmatrix}
x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 \\
x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2u_2 & -y_2u_2 \\
x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3u_3 & -y_3u_3 \\
 & & & \cdot & & & & \\
x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4u_4 & -y_4u_4 \\
0 & 0 & 0 & x_1 & y_1 & 1 & -x_1v_1 & -y_1v_1 \\
0 & 0 & 0 & x_2 & y_2 & 1 & -x_2v_2 & -y_2v_2 \\
0 & 0 & 0 & x_3 & y_3 & 1 & -x_3v_3 & -y_3v_3 \\
 & & & \cdot & & & & \\
0 & 0 & 0 & x_4 & y_4 & 1 & -x_4v_4 & -y_4v_4
\end{vmatrix}
\qquad
X = \begin{vmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{vmatrix}
\qquad
b = \begin{vmatrix} u_1 \\ u_2 \\ u_3 \\ \cdot \\ u_4 \\ v_1 \\ v_2 \\ v_3 \\ \cdot \\ v_4 \end{vmatrix}
$$

Unknowns can be determined as follows,

$$X = A^{-1}b. \qquad\qquad (4.12)$$

Example of rectification using this method:

For the illustration of the rectification method described above, a stereo pair was selected. The exterior parameters of convergent and rectified images are given in Table 4.1.

convergent/ rectified	X0 (mm)	Y0 (mm)	Z0 (mm)	omega (deg)	phi (deg)	kappa (deg)	focal length (mm)
convergent-1	2530.7640	-4.3070	3063.0021	9.2558	48.4996	11.2267	16
rectified-1	2530.7640	-4.3070	3063.0021	-22.7178	20.9302	-40.7602	32
convergent-2	18.9203	2504.8536	3054.0878	-32.9053	12.5809	-4.4002	16
rectified-2	18.9203	2504.8536	3054.0878	-22.7178	20.9302	-40.7602	32

Table 4.1 Exterior parameters of convergent and rectified images

The focal length of the rectified images was chosen to be *32mm*. Two sets of rectification parameters *(a,b,c,d,e,f,g,h)* were determined using the previously described method for convergent-*1*/rectified-*1* and convergent-*2*/rectified-*2* as follows,

$$(1.2170, -1.5871, -16.3436, 1.7819, 1.3668, -0.0132, 0.0195, -0.0253),$$
$$(1.5849, -1.2199, 16.1477, 1.3660, 1.7755, 0.0132, -0.0250, 0.0192)$$

Nine points which are randomly distributed in *3-D* space were imaged onto a pair of convergent images. These points were then projectively transformed into rectified images. Table 4.2 illustrates image point co-ordinates on convergent and rectified images. It can be seen that the y-co-ordinates of the corresponding points in the rectified images are equal.

convergent-1		rectified-1		convergent-2		rectified-2	
x(*mm*)	y(*mm*)	x(*mm*)	y(*mm*)	x(*mm*)	y(*mm*)	x(*mm*)	y(*mm*)
1.0083	1.2916	-17.3942	3.5960	1.2005	1.0753	16.8969	3.5960
1.0334	-1.2525	-12.4531	0.1105	1.0868	-0.7840	19.6570	0.1105
-0.8735	1.1733	-20.2125	0.0355	-1.3635	1.0627	12.0342	0.0355
-0.8238	-1.1361	-15.3475	-2.9957	-1.2322	-0.7736	14.9006	-2.9957
1.0234	0.0183	-14.8388	1.8003	1.1406	0.1020	18.3177	1.8003
2.0568	2.7114	-18.6788	7.5748	2.6244	2.1346	18.1502	7.5748
-1.6391	-2.9233	-13.1500	-6.6496	-2.2205	-2.1010	14.9643	-6.6496
3.3193	4.1481	-19.6841	12.0590	4.1656	3.3029	19.5154	12.0590
1.7286	-2.1761	-9.9074	0.0851	1.8063	-1.3526	22.2433	0.0851

Table 4.2 Co-ordinates of convergent and rectified image points

A graphical illustration of the convergent and rectified image co-ordinates are given in Figure 4.8. It can be seen that y-co-ordinates of the corresponding points in rectified images are equal as expected with this ideal data set.
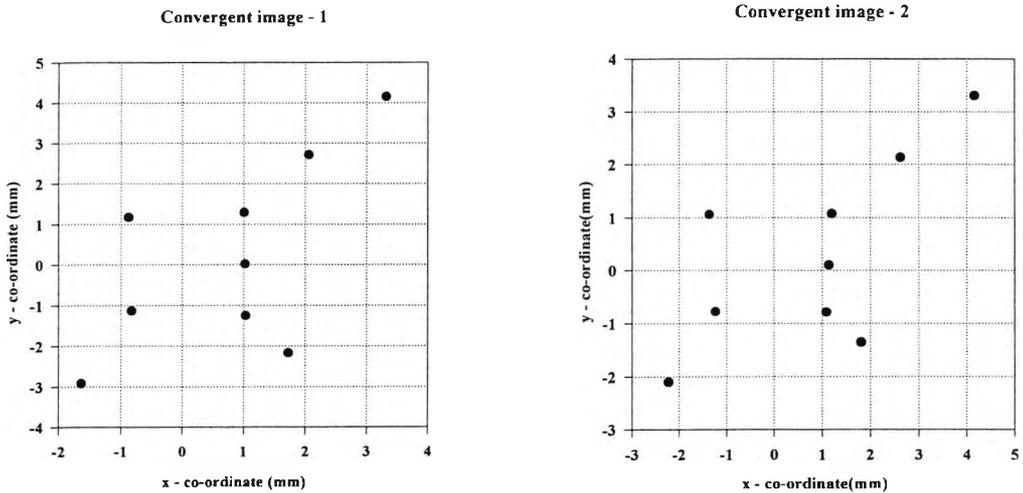
(a) Convergent images
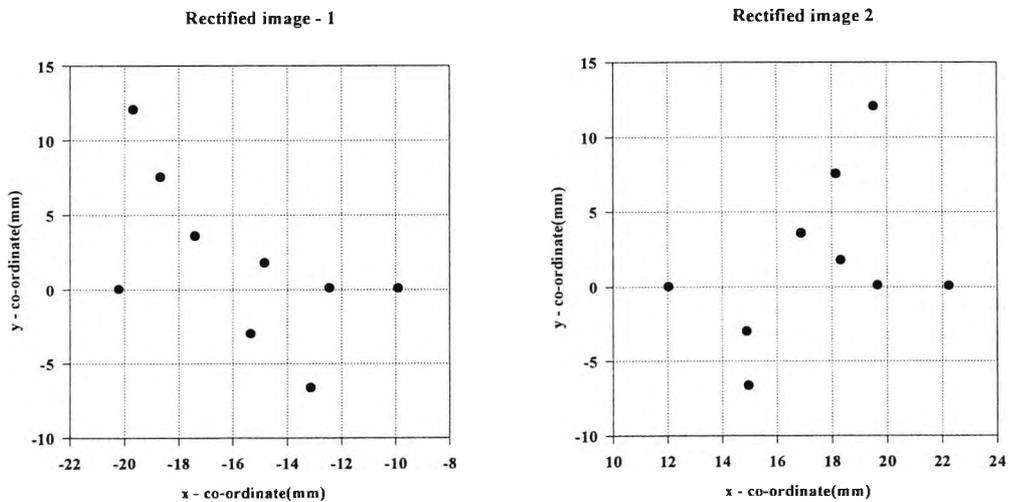
(b) Rectified images

Figure 4.8 Illustration of convergent and rectified image points

4.2 Correspondences in rectified image space

Once the image point co-ordinates are transformed into the rectified image space, the next task is the establishment of correspondences. As in the case of convergent images, a search band is required to constrain the search for corresponding points. This section discusses how this search band is determined using the knowledge of the search bands in the convergent images. In the first step, the geometric relationship

between the search bands in the convergent and the rectified images is established. This relationship allows determination of the search band for any epipolar line in the rectified image. Preliminary correspondence speed performance tests were carried out and comparisons were made with the *EL* and the *3DSI* methods. In the second step, the relationship established in the first step and the boundaries of the rectified images are used along with efficient sorting and searching methods to further speed up correspondences.

4.2.1 Search band in rectified images

The area covered by a search band is defined by its two border lines. If the parameters of the border lines (*i.e.* slopes and intercepts) with respect to the image co-ordinate systems are known, image points that lie within search band can be found. In order to establish correspondences reliably, the search band must be determined correctly. For a given stereo pair, the search bands in convergent images can be projectively transformed into rectified images for the purpose of this analysis. Figure 4.9 shows a convergent image pair, a corresponding rectified image pair, and the transformed search bands of I_2 on I_{r2}. In image I_2, a search band covering the epipolar line l_{2i} has border lines l_{1bi} and l_{2bi}. The corresponding epipolar line, and the transformed border lines in rectified image I_{r2} are l_{r2i}, l_{r2bi}, and l_{r1bi}.
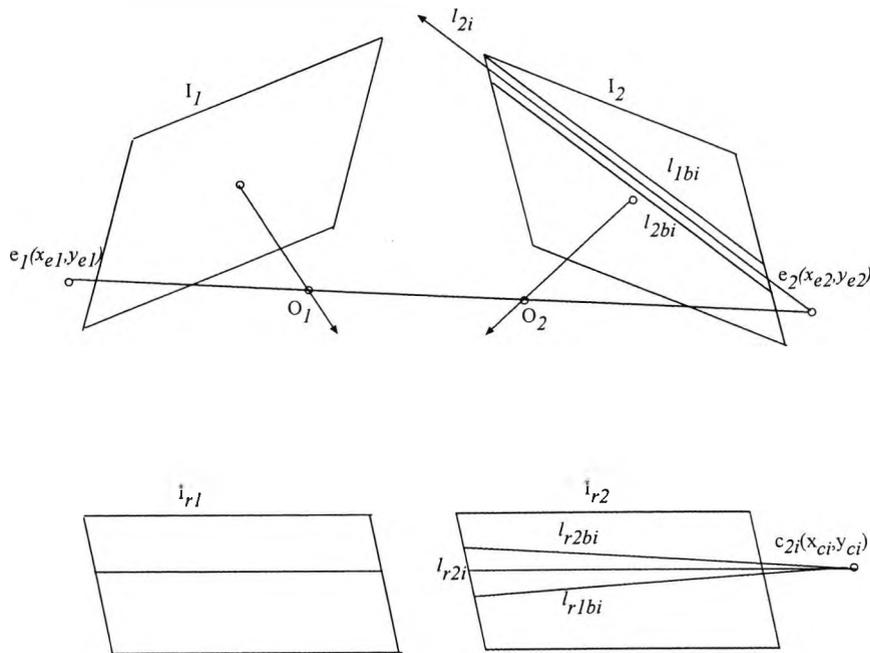

Figure 4.9 Transformation of search band from convergent to rectified image

The characteristics of the transformed search band border lines on rectified images are dependent on the relative orientation between the local axis systems of the convergent and rectified image planes. When a stereo pair has a narrow angular separation, the angle between I_2 and I_{r2} is smaller and vice versa. Simulation tests were carried out to analyse the behaviour of the transformed search bands. A number of stereo pairs with different angular separations and base line orientations were selected for the tests. Table 4.3 shows the absolute locations, orientations, angular separations (approximate) of convergent axis stereo pairs, and the angles between the convergent and the rectified image planes (approximate).

image pair	X0 (mm)	Y0 (mm)	Z0 (mm)	omega (deg)	phi (deg)	kappa (deg)	angular sep: (deg)	angle between planes (deg)
1	1474.8944	-4.3070	3003.7424	9.2558	24.5394	11.2267	50°	27°
	-1472.9510	22.5409	2982.7308	5.2247	-24.8201	11.2995		25°
2	2474.8944	-4.3070	3003.7424	9.2558	37.7799	11.2267	80°	43°
	-2507.3751	0.3366	3024.5680	6.7528	-33.0921	-12.8005		48°
3	5030.7640	-4.3070	3063.0021	9.2558	67.7303	11.2267	113°	60°
	-4962.3294	1.2190	3025.2061	5.6585	267.2032	-44.3246		56°
4	4974.8944	-4.3070	2503.7424	9.2558	60.0106	11.2267	43°	46°
	3519.1430	3540.3876	3018.5142	-42.7842	22.2320	-12.0521		60°
5	-3526.9405	-3536.7584	2500.9539	55.2386	-36.6681	-4.4233	85°	41°
	3497.8085	-3511.7127	2988.0790	41.2868	36.8887	0.0353		55°

Table 4.3 Stereo pairs for simulation tests

Referring to Figure 4.9, for each stereo pair a set of epipolar lines l_{2i} *(i = 1,....,n)* was created on image I_2 with regular angular spacing between them. Corresponding epipolar lines in the rectified image I_{r2} are l_{r2i} *(i = 1,....,n)*. The search band border lines in image I_2: l_{1bi}, l_{2bi} *(i = 1,....,n)* were transformed into the rectified image I_{r2} resulting in corresponding border lines l_{r2bi}, l_{r1bi} *(i = 1,....,n)*. A characteristic of any pair of border lines is that they intersect at a point $c_{2i}(x_{ci},y_{ci})$ which lies on the epipolar line l_{r2i} *(i = 1,....,n)*. Each y-co-ordinate in rectified image space represent an epipolar line. The slopes and the crossing points of transformed border lines were noted for each stereo pair. Figure 4.10 illustrates slopes and x-co-ordinates of crossing points for varying y-co-ordinates for each stereo pair. The search band width in image I_2 was *0.016mm* and focal lengths of rectified images were *32mm* for this test.
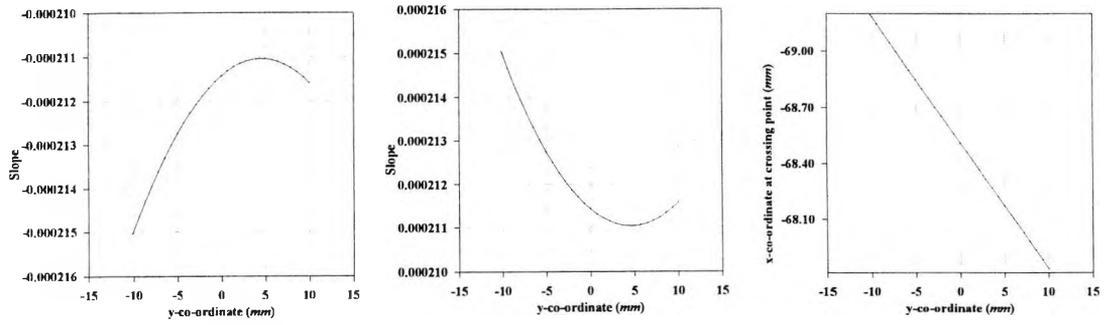
border line l_{r1bi} border line l_{r2bi} x-co-ordinate

(a). Image pair - 1
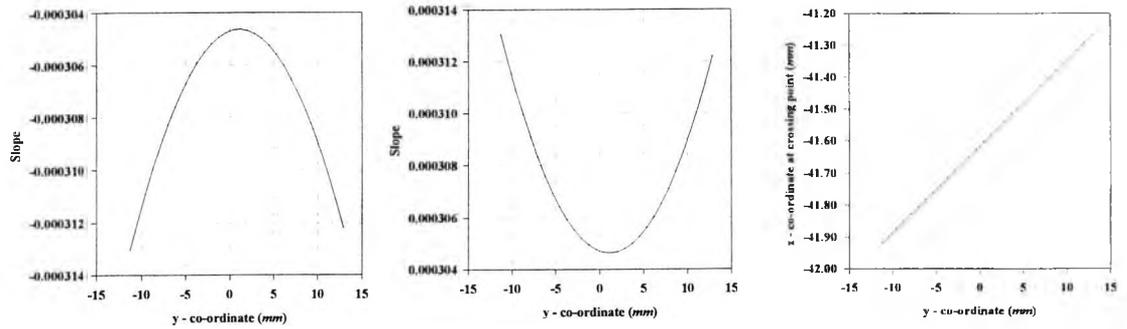
border line l_{r1bi} border line l_{r2bi} x-co-ordinate

(b). Image pair - 2
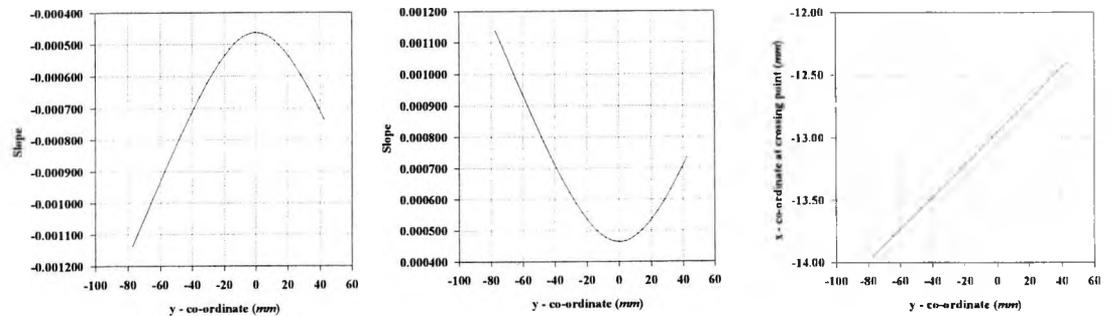
border line l_{r1bi} border line l_{r2bi} x-co-ordinate

(c). Image pair - 3
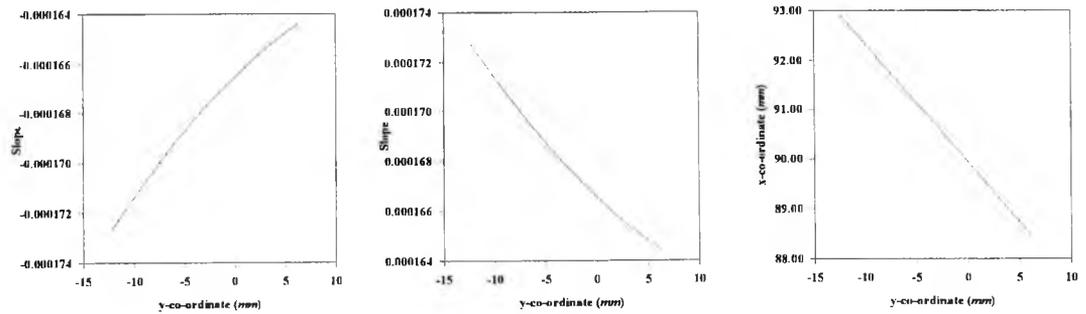
border line l_{r1bi} border line l_{r2bi} x-co-ordinate

(d). Image pair - 4

| border line l_{r1bi} | border line l_{r2bi} | x-co-ordinate |

(e). Image pair - 5
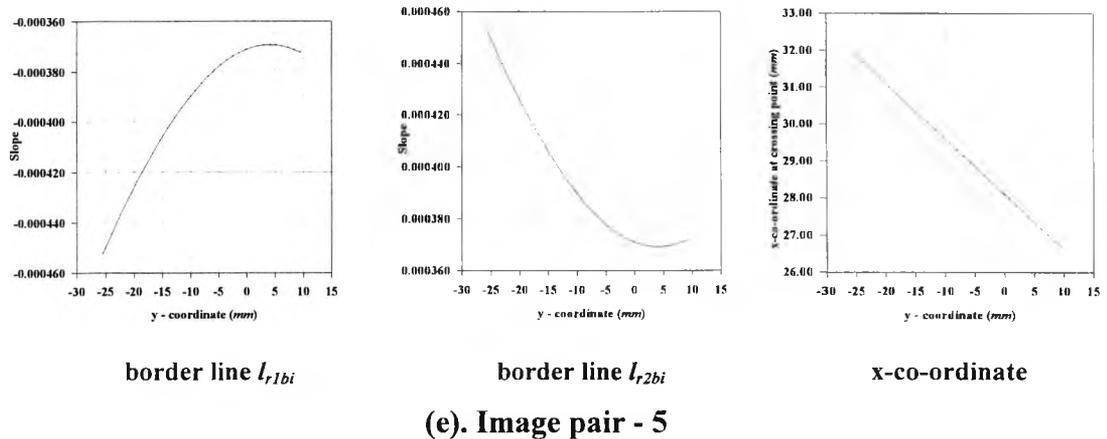
Figure 4.10 Variation of border line slopes and x-co-ordinate at crossing points in rectified image

It can be seen (Figure 4.10) that the variation of the border line slopes over the entire rectified image depends on the stereo pair. As the angular separation increases the value of slopes increases. If the base line slope is smaller, the variation of the slopes follows a symmetric curve ((a), (b), and (c)). When the base line slope is larger and angular separation is smaller, the variation of slopes is relatively smaller ((d), (e)). These slope variations were analysed for a range of band widths covering *0.0016 - 0.16mm*. The Pulnix cameras that were used for experiments have pixels of dimensions *8.2x8.4μm* hence, the range *0.0016-0.16mm* was selected to cover possible band widths. It was noted that as the width of the convergent image search band increases the slopes of the transformed border lines on the rectified images increases. There is a linear relationship between the search band width in convergent image and the slopes of the transformed search band border lines. A ten fold increase in band width causes a tenfold increase in slopes.

For any stereo pair, the x-co-ordinate at the border line intersecting point varies linearly as the y-co-ordinate is varied. For a given stereo pair, when the rectified image focal length is increased the crossing point moves away from centre of image. The centre of the rectified image is referred as the point which is the base of the perpendicular from optical centre of image. For a given stereo pair the crossing point is dependent only on the focal length of the rectified image.

Due to the predictable nature of the slope variation, it can be mathematically represented by a *2nd* order polynomial. Similarly, the linear variation of the x-co-ordinate at the crossing point can be represented by a straight line. Each of these relationships is a function of the y-co-ordinate in the rectified image space. For a given stereo pair, these relationships can be determined during the initialisation process. Hence, for a given point in a rectified image, the search band slopes and the x-co-ordinate at crossing point can be computed. This technique leads to minimised computations to find the search band and to the author's knowledge has not been reported in any of the literature.

Polynomial representation of the variation of the search band border line slope:

A second order polynomial representing the slope s of a border line as a function of y-co-ordinate (y_i) can be stated as,

$$s(y_i) = b_0 \cdot y_i^2 + b_1 \cdot y_i + b_2 \qquad (4.13)$$

where b_0, b_1, and b_2 are the coefficients of the polynomial which can be estimated by minimising the residuals using the least squares technique. The residual is defined as the difference between the estimated and true values which can be represented as,

$$R = s_i - s(y_i) \qquad (4.14)$$

The least squares criteria states that sum of the squares of the residuals (*SSR*) should be minimised to find the estimates for the unknowns. Hence,

$$SSR = \sum_{i=1}^{n} \left(s_i - s(y_i) \right)^2 \qquad (4.15)$$

should be minimised. Substituting for $s(y_i)$,

$$SSR = \sum_{i=1}^{n} \left(s_i - \left(b_0 y_i^2 + b_1 y_i + b_2 \right) \right)^2$$

SSR is a function of b_0, b_1, b_2. Hence, suitable values for these variables which minimise *SSR* can be obtained by partially differentiating with respect to each variable. That is,

$$\frac{\partial(SSR)}{\partial b_0} = 0$$

$$\frac{\partial(SSR)}{\partial b_1} = 0$$

$$\frac{\partial(SSR)}{\partial b_2} = 0 \tag{4.16}$$

which results in three linear equations,

$$nb_0 + b_1 \sum_{i=1}^{n} y_i + b_2 \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} s_i$$

$$b_0 \sum_{i=1}^{n} y_i + b_1 \sum_{i=1}^{n} y_i^2 + b_2 \sum_{i=1}^{n} y_i^3 = \sum_{i=1}^{n} s_i y_i$$

$$b_0 \sum_{i=1}^{n} y_i^2 + b_1 \sum_{i=1}^{n} y_i^3 + b_2 \sum_{i=1}^{n} y_i^4 = \sum_{i=1}^{n} s_i y_i^2 \tag{4.17}$$

In matrix form,

$$AX = B$$

Hence, unknowns b_0, b_1, and b_2 can be estimated as,

$$X = A^{-1}B \tag{4.18}$$

Straight line representation of the variation of x-co-ordinate at crossing point of border lines:

The x-co-ordinate variation can be represented by a straight line as a function of y_i as,

$$x(y_i) = b_0 \cdot y_i + b_1 \tag{4.19}$$

where b_0 and b_1 are the gradient and intercept. As in the previous case, the sum of the squares of residuals is,

$$SSR = \sum_{i=1}^{n}\left(x_i - \left(b_0.y_i + b_1\right)\right)^2 \qquad (4.20)$$

To obtain optimum values for b_0 and b_1,

$$\frac{\partial(SSR)}{\partial b_0} = 0 \text{ and } \frac{\partial(SSR)}{\partial b_1} = 0 \qquad (4.21)$$

which results in two linear equations,

$$b_0 \sum_{i=1}^{n} y_i^2 + b_1 \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i y_i$$

$$b_0 \sum_{i=1}^{n} y_i + nb_1 = \sum_{i=1}^{n} x_i \qquad (4.22)$$

In matrix form,

$$AX = B$$

Hence, unknowns b_0 and b_1 can be estimated as,

$$X = A^{-1}B \qquad (4.23)$$

However, it can be seen that in each case the x-co-ordinate variation follows a straight line. Hence, the equation of the line can be determined by simply noting two points that lie on the line.

Approximated polynomials and straight lines for slope and x-co-ordinate variations given in Figure 4.10 are as follows,

pair-1:

l_{r1bi} : $s(y) = -0.0000000184 y_i^2 + 0.0000001689 y_i - 0.00021143$

l_{r2bi} : $s(y) = 0.0000000184 y_i^2 - 0.0000001689 y_i + 0.000211424$

$\qquad x(y) = 0.067 y - 68.5106$

pair-2:

l_{r1bi} : $s(y) = -0.0000000547 y_i^2 + 0.0000001224 y_i - 0.0003047109$

l_{r2bi} : $s(y) = 0.0000000547 y_i^2 - 0.0000001224 y_i + 0.0003047059$

$\qquad x(y) = 0.0269 y - 41.6158$

pair-3:

l_{r1bi} : $s(y) = -0.0000001226 y_i^2 - 0.0000001387 y_i - 0.0004836085$

l_{r2bi} : $s(y) = 0.0000001226 y_i^2 + 0.0000001387 y_i + 0.0004836029$

$\qquad x(y) = 0.0129 y - 12.9546$

pair-4:

l_{r1bi} : $s(y) = 0.0000000090 y_i^2 - 0.0000003952 y_i + 0.0001665325$

l_{r2bi} : $s(y) = -0.0000000090 y_i^2 + 0.0000003951 y_i + 0.0001665192$

$\qquad x(y) = -0.2408 y + 89.9376$

pair-5:

l_{r1bi} : $s(y) = 0.0000000915 y_i^2 - 0.0000008759 y_i + 0.0003715174$

l_{r2bi} : $s(y) = -0.00000000915 y_i^2 + 0.0000008758 y_i - 0.0003714762$

$\qquad x(y) = -0.1497 y + 28.0985$

It can be concluded that for any stereo pair, a polynomial representation for the variation of search band border line slopes and straight line representation for variation of the x-co-ordinate at the intersecting points can be obtained. Further tests indicated that a 2nd order approximation is sufficient for slope variations of any stereo

pair. These tests include the variation of the sensor size which causes the epipolar lines to be spread in a wide area of the convergent image space.

4.2.2 Preliminary speed performance tests

Tests were carried out with the *PEL* algorithm at this stage of development to assess its speed performance. A Pentium processor based *PC* operating at *200MHz* was used and the correspondence speed performances of the *PEL* algorithm were compared with the *EL* and *3DSI* algorithms. For a varying number of densely distributed image points n ($n = 10 - 400$) in each image, the time taken by each algorithm to establish correspondences in a stereo pair was noted (Figure 4.11 (a)). No occlusions were present as the images were of a simulated planar target field. It can be seen that the *3DSI* algorithm has poor speed performance compared to others. This is due to the costly minimum distance computations that are required to be performed n^2 times assuming n points in each image. The *EL* algorithm has better speed performance than the *3DSI* algorithm due to fewer computations being required to be performed n^2 times. The *PEL* algorithm at this stage has better speed performance than the *EL* and the *3DSI* algorithms when the number of image points n is greater than about *130* and *30* respectively (Figure 4.11 (b) and (c)). The *PEL* algorithm requires an n^2 search, but has reduced computations due to the availability of pre-computed search band border line parameters. The application of epipolar constraint involves fewer computations hence, it has better speed performance. For fewer than *130* and *30* points in each case, the *PEL* algorithm has lower performance due to the overhead of parameter computation in the initialisation stage which took *1ms* for a stereo pair.


(a)   (b)   (c)
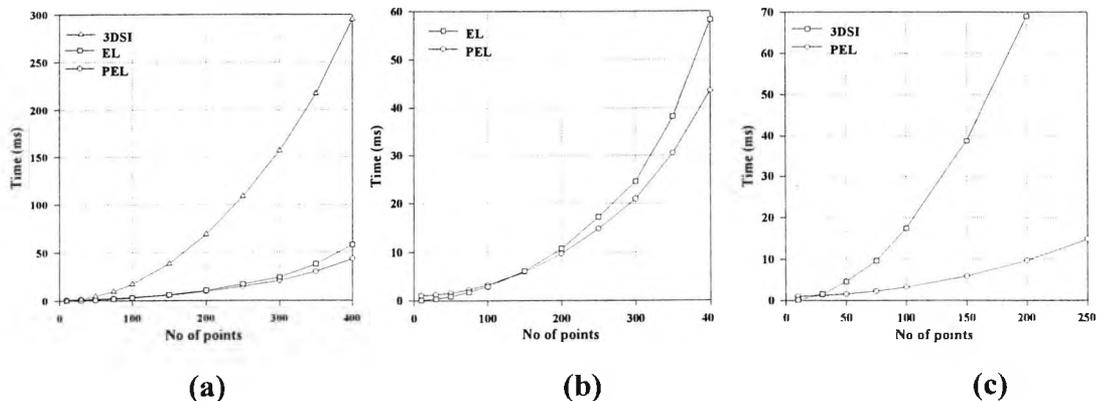
Figure 4.11 Speed performance of algorithms at stereo pair level

4.2.3 Use of parallel search band border lines in rectified image space

With the aforementioned knowledge of the behaviour of the search band border lines over the whole rectified image, the n^2 search can be avoided. This is achieved by replacing the sloped border lines with those that are parallel to the epipolar line in the rectified image space. Figure 4.12 illustrates a pair of rectified images I_{r1}, I_{r2} and a pair of corresponding points p_{r1i} and p_{r2i}. Lines l_{r1i} and l_{r2i} are conjugate epipolar lines. l_{r1bi} and l_{r2bi} are search band border lines which are parallel to the epipolar line l_{r2i}. d_1 and d_2 are the y - distances to each border line from the epipolar line. Provided the rectified image co-ordinates of both images are sorted in ascending order of the y-co-ordinate, an efficient search can be performed to find corresponding points.
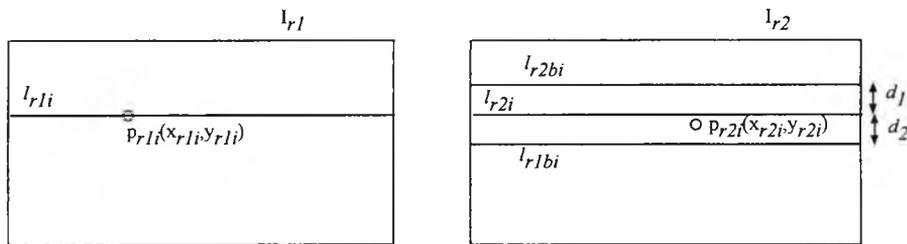

Figure 4.12 Parallel search band border lines in rectified images

The algorithm uses the knowledge about the co-ordinates of the transformed four corners of the convergent image on rectified image plane to determine distances d_1 and d_2. It was established that for a given stereo pair, the search band border line slopes and x-co-ordinate at the crossing point of border lines can be computed for any y-co-ordinate using the relationships developed. Figure 4.13 illustrates transformation of the four corner points cp_{r1i} and cp_{r2i} ($i=1,..,4$) on the rectified images I_{r1} and I_{r2}. The straight line DE represent the variation of the x-co-ordinates at the border line crossing points. At point C the border lines are AC and BC which intersect the line joining corner points cp_{r21} and cp_{r22} at points A and B. Hence, d_1 and d_2 are y-distances to the points of intersection from epipolar line which are simply the differences between y-co-ordinates at points A, C and B, C.
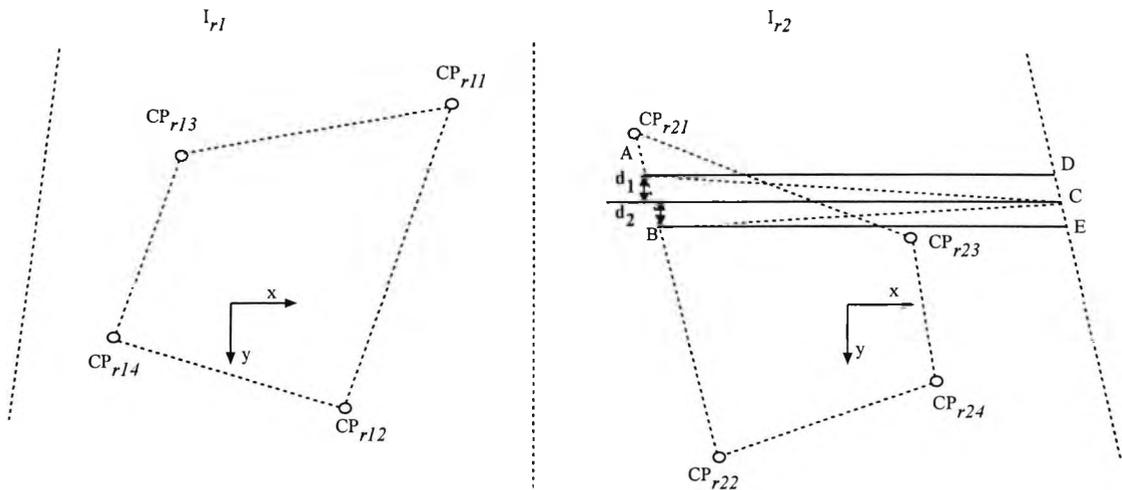
Figure 4.13 Transformed four corners and parallel border lined search band in rectified image

A possible concern is that area covered by the parallel border lined search band is larger than that covered by sloped border lined search band. The result will be that unwanted image points can also fall within the band. In order to avoid this problem, a simple check can be performed. In the first step, points that fall within the parallel search band area *ABED* can be identified efficiently. In the second step, the selected points that fall within the area *ACB* can be identified as potential candidates thus removing the additional ambiguities.

Having determined separations d_1 and d_2, stereo correspondences can be established searching through sorted image co-ordinates. Assuming *n* transformed points in each image, application of epipolar constraint only involves comparison of y-co-ordinates to check if a point falls within the band. With respect to reliability the *EL* and *PEL* algorithms should ideally have similar performance. The reason is that the search band in the rectified image is the projective transformation of the parallel band in the convergent image.

The reduced computations and searching due to the use of parallel border lines makes the correspondence process faster. The *PEL* algorithm is expected to be more efficient than *3DSI* and *EL* methods. Previous speed performance tests were repeated. The time taken by a stereo pair for a varying number of points was noted (Figure 4.14(a)).

(a)   (b)   (c)
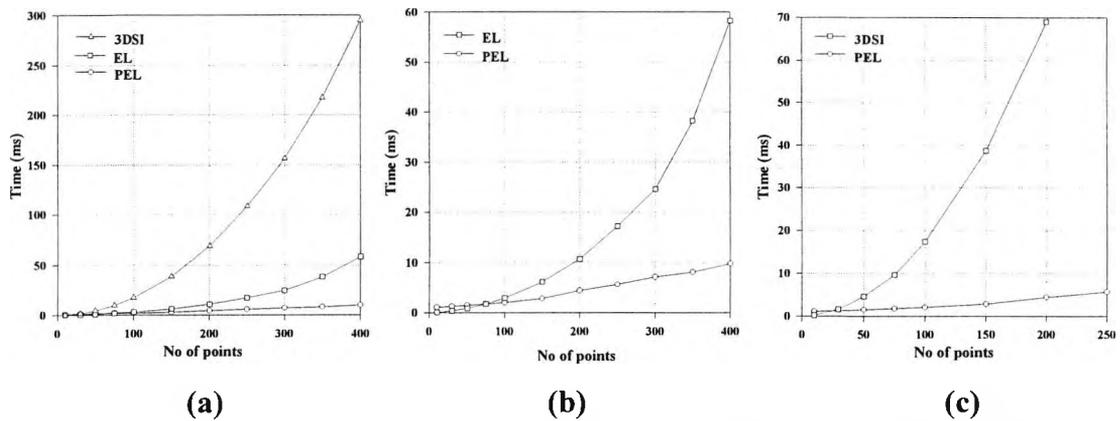
Figure 4.14 Speed performance of algorithms at stereo pair level with performance improved *PEL* algorithm

The *PEL* algorithm now out performs *EL* and *3DSI* algorithms when the number of points is greater than about *75* and *25* respectively ((b) and (c)). The algorithm now took *10ms* as opposed to *43ms* for correspondence of *400* points. The time taken now varies approximately linearly with the increasing number of image points.

4.3 *PEL* algorithm in multiple view point correspondences

The establishment of correspondences between a pair of views using the *PEL* algorithm was discussed. The implementation of the *PEL* algorithm for multiple view image point correspondences is discussed in this section. There are two steps. In the first step, a set of images acquired from multiple view points is considered as a combination of stereo pairs. The potential corresponding candidates are identified at stereo pair level as discussed previously. In the second step, sets of corresponding points in multiple views are identified by performing a search based on image point identities.

The epipolar geometric relationship between a pair of convergent axes images and the rectified images were discussed in section 4.1. Establishment of correspondences in the rectified image space was discussed in section 4.2. The purpose of the establishment of correspondences in the rectified image space is to avoid the searching that has to be performed in convergent image space. Once correspondences are established, the subsequent *3-D* location estimations are performed using the original convergent axis image co-ordinates. The geometry and the issues such as ambiguities can be conveniently discussed for a stereo pair in rectified image space but it is

inconvenient in multiple view situations. Hence, in the application of the *PEL* algorithm to multiple views, the discussions will be carried out in the convergent axes image space.

4.3.1 *PEL* algorithm in a 3-view configuration

Three convergent axis images I_1, I_2, an I_3 can be considered as three stereo pairs I_1/I_2, I_2/I_3, and I_3/I_1. Each stereo pair is rectified and the potential correspondences can be identified in the rectified image space. For the purpose of the investigation, correspondences are illustrated in equivalent convergent axes image space (Figure 4.15).
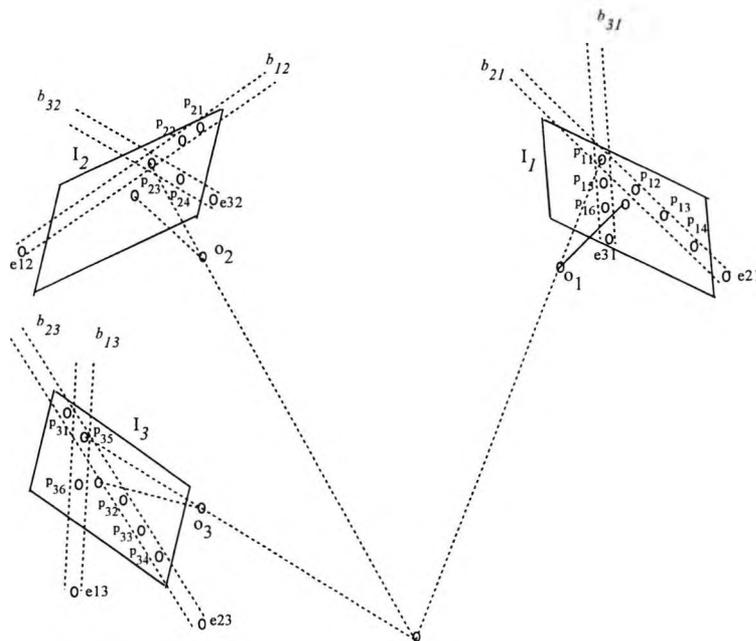

Figure 4.15 Three views geometry

Points p_{11}, p_{23}, and p_{35} in images I_1, I_2, and I_3 are projections of a *3-D* object point. Consider point p_{11} in image I_1 and stereo pair I_1/I_2. It can be seen that points p_{21}, p_{22}, p_{23} fall within the band b_{12} in image I_2 and hence are potentially corresponding to point p_{11}. Now considering point p_{23} in image I_2 and stereo pair I_2/I_3, it can be seen that points p_{31}, p_{32}, p_{33}, p_{34}, and p_{35} fall within band b_{23} in image I_3 hence are potentially corresponding to point p_{23}. Finally considering point p_{35} in image I_3 and stereo pair I_3/I_1, it can be seen that points p_{11}, p_{15}, p_{16} fall within band b_{31} in image I_1 hence are potentially corresponding to point p_{35}. The resulting correspondence information can be represented in the form of a tree (Figure 4.16) for the purpose of illustration. Points

that form a closed-loop in which a search path that begins and ends with the same point identity number can be taken as potentially corresponding points. It can be seen from *step-3* that one of the potentially corresponding candidates for p_{35} in image I_l is the same point that was initially considered *(p_{11})*. Hence, points p_{11}, p_{23}, and p_{35} form a closed loop and are taken as corresponding. If no ambiguities are present, all corresponding sets of points can be reliably identified.
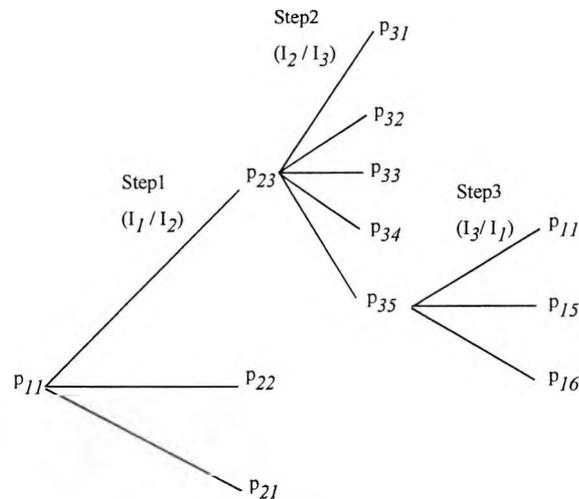

Figure 4.16 A tree of potentially corresponding image points

Consideration of ambiguities:

Ambiguities can be divided into two categories as *2-D* image space and *3-D* object space. The *2-D* category relates to image point location errors such as radiometric and geometric distortions. These are mostly due to systematic errors. In high precision applications, image data are corrected for systematic errors. The random errors such as are due to electronic noise are generally small. However, there are other factors of electronic errors due to camera "warm up" effects (Chen, 1995) which can be avoided by allowing for camera stabilisation prior to use. In the high precision environment, various *3-D* situations are more likely cause ambiguities. The reasons for the *3-D* category are due to the relative location of *3-D* points in object space with respect to imaging geometry causing occlusions and points being out of the field of view. In general, *3-D* ambiguities are unavoidable and may lead to incorrect correspondences. In order to understand these sources of potential errors, they are now analysed in detail.

It was discussed in section 3.3.1.4 how *3-D* ambiguities occur at stereo pair level. But most of these ambiguities can be solved if a third view is available. Some ambiguities are unsolvable without further constraints even when a third view is available. These can be categorised as follows.

- **When image points are too close together:**

This situation can occur due to *3-D* object space points that are close together. The resulting image points may lie close enough to fall within the intersecting area of epipolar search bands (Figure 4.17). It is not possible for the epipolar constraint to solve this ambiguity. Another reason for having multiple points within intersecting areas is when the selected search band width is larger than appropriate. The only solution for this type of ambiguity is to discard image points that are closer than the width of the search band.
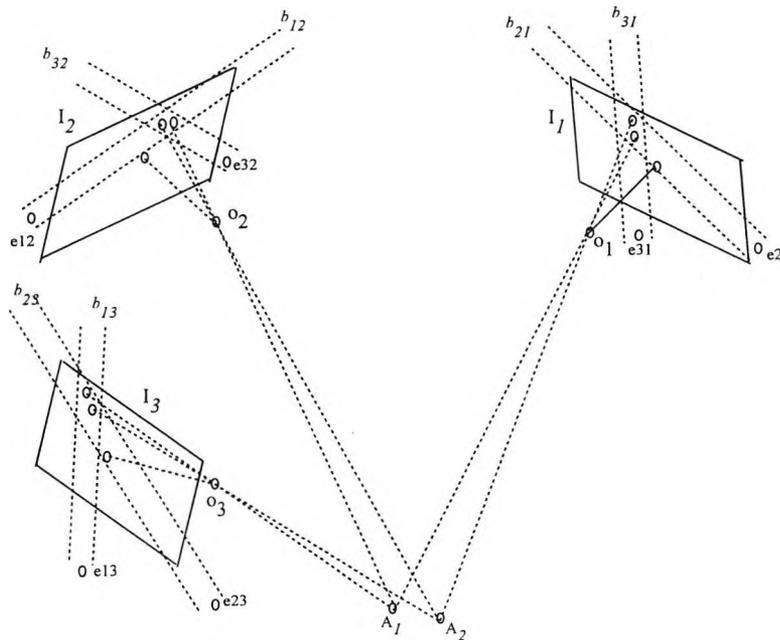

Figure 4.17 Ambiguities due to image points that are too close to each other

- **Due to multiple *3-D* points creating collinear image points in a view:**

When two *3-D* object points along with the optical centre of a view are collinear (Figure 4.18) two paths that begin and end with same identity number may be created. This situation can be further complicated if each view has many such collinear points. Collinear points are coincident in image I_1 but create two points in images I_2 and I_3 which results in two different paths that begin and end with the same point in image I_1.

Although both paths are correct, the detection of such two paths creates ambiguity if the two 3-D points are occluded by an object in between them. This can occur when points are on two surfaces of an object. However, this situation is unlikely to create any significant error in the computed *3-D* points. There is no solution for this type of ambiguity.
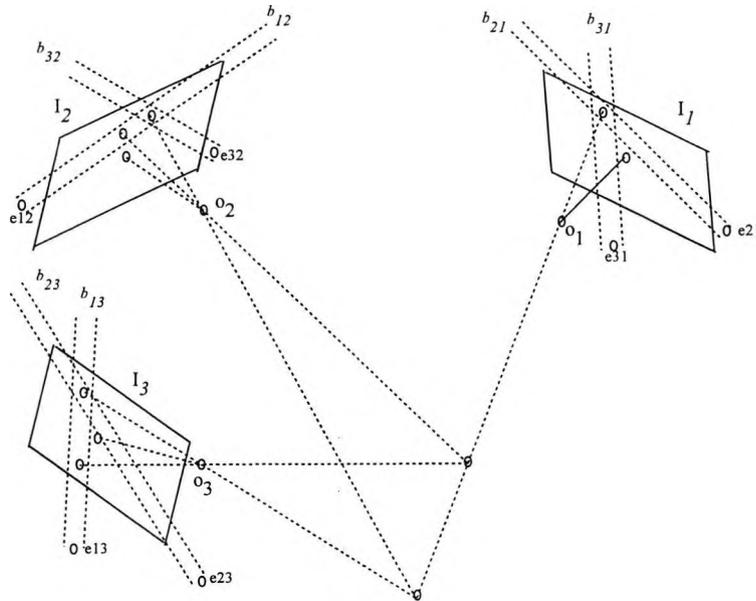

Figure 4.18 Ambiguities due to coinciding image points

- ***3-D* points that are out of the field of view of two images but appears on the third:**

Images of *3-D* point A on images I_1, I_2, and I_3 are shown at intersections of search bands (Figure 4.19). The *3-D* points B and C appear only on images I_2 and I_3 respectively. But rays AO_1, CO_2, and BO_3 intersect at point D hence, an unsolvable ambiguity due to two paths that begin and end with the same identity number in image I_1. This creates a *3-D* point that does not exist. Similar instances can occur when *3-D* point C is on the epipolar plane AO_1O_2 but within the area covered by points A,O_1, and O_2. Even if more views are available this ambiguity remains unsolvable using the epipolar constraint. However, if object space information such as structure is available, this type of non-existing *3-D* points may be detected.
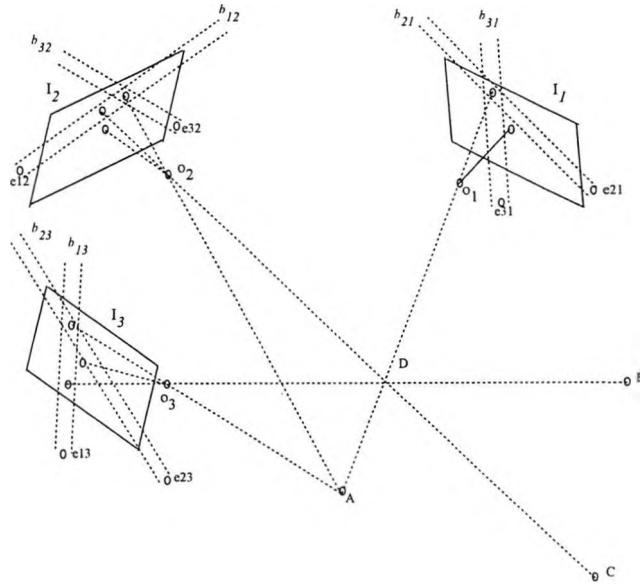
Figure 4.19 Ambiguities due to *3-D* points which are out of the field of view

4.3.2 *PEL* algorithm in 4-view configuration

Four views I_1, I_2, I_3, and I_4 can be considered, in the simplest way, as four stereo pair combinations I_1/I_2, I_2/I_3, I_3/I_4, and I_4/I_1. Each stereo pair is rectified and potential correspondences are identified in each rectified image space. For the purpose of this investigation, these correspondences are illustrated in equivalent convergent axes image space (Figure 4.20).
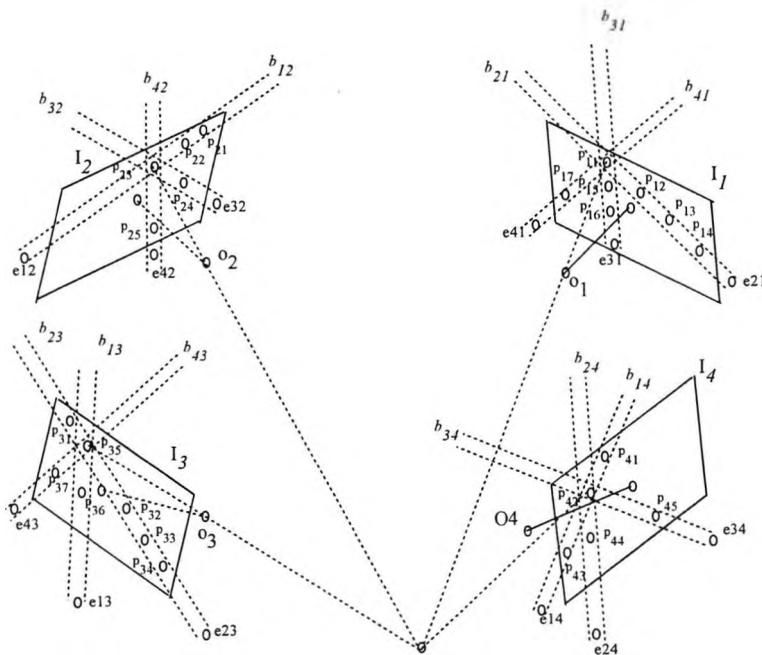

Figure 4.20 Four view geometry

As in the case of three views, the condition that four corresponding points in four views form a closed-loop can be used to establish point correspondences. A closed-loop is created by a path that begins and ends with the same point identity number. Images of a *3-D* point on images I_1, I_2, I_3, and I_4 are p_{11}, p_{23}, p_{35}, and p_{43} respectively. A search beginning with point p_{11} in image I_1 will enable the identification of corresponding point in each image.

Four views provide more geometric information than three views. The ambiguities discussed under three views can occur with four views as well. The availability of a further two combinations I_1/I_3 and I_2/I_4 provides information to cross check the validity of correspondences.

4.3.3 Speed performance comparisons with four views

As discussed previously, images acquired from multiple view points were considered as a combination of stereo pairs. Correspondences were established at stereo pair level and the geometric relationship discussed for the *EL* algorithm was used for the distribution of correspondences. Timing tests were carried out with the same set of simulated co-planar *3-D* points. For a varying number of image points, the time taken by each algorithm was noted (Figure 4.21(a)).

(a) (b) (c)
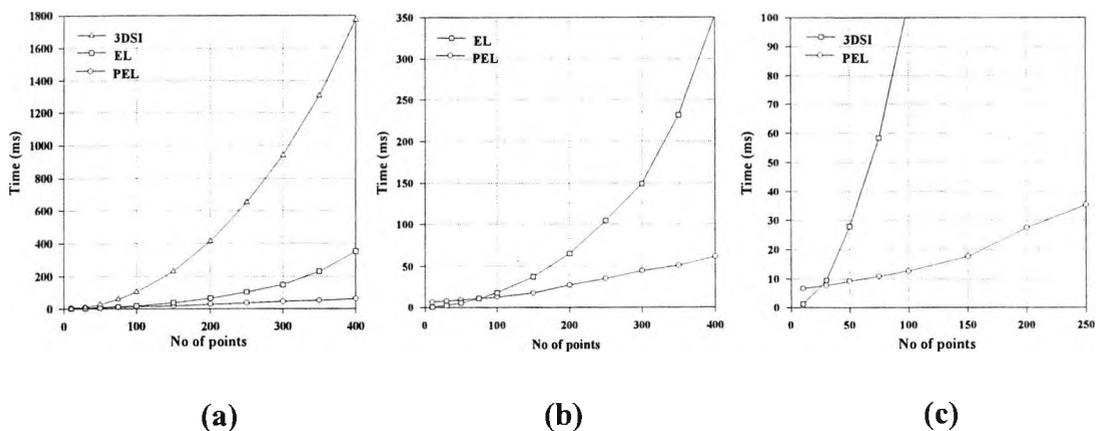
**Figure 4.21 Speed performance of algorithms with four views**

With this implementation, the *PEL* algorithm is faster than each algorithm when the number of points is more than about *75* and *25*. The time taken varies approximately linearly with increasing number of points. For *400* points, the *PEL* algorithm takes *60ms* where as the *EL* and *3DSI* algorithms takes *350ms* and *1.78s* respectively.

Hence, the *PEL* algorithm has the best correspondence speed performance. Approximately linear variation of the time consumption makes it possible to predict an approximate time the algorithm takes to establish correspondences for a given number of points.

4.4 Use of further geometric constraints

The epipolar constraint can be used to establish correspondences. However, it is possible that wrong correspondences are established which may lead to the creation of *3-D* points that do not exist. In such instances prior knowledge of the *3-D* object model may be used to discard incorrectly corresponded *3-D* points. Another method is to use the knowledge of image residuals. When corresponding points are identified, the *3-D* location can be estimated. If rays are reprojected into image space from this point, image residuals can be computed. For correctly corresponded points, the residuals should be within a small tolerance.

4.5 Summary of the chapter

The development of a technique for the establishment of multiple view point correspondences was discussed. The advantages of this algorithm over the others is its speed performance and the predictability of computational time. Assuming a stereo pair which has n points in each image the *EL* and *3DSI* techniques require approximately an n^2 search for identifying the corresponding points which is computationally expensive. The searching required by *PEL* algorithm is approximately proportional to n. This computational advantage is due to two factors. The first is due to the collinear epipolar lines in rectified image space. Hence, there is no computational work involved in determining the conjugate epipolar lines. The second is due to the new technique used to determine search band for each epipolar line in rectified image space. This new technique establishes a polynomial relationship between the slopes of search band border lines and the y-co-ordinate. The new technique also determines a straight line relationship between the x and y co-ordinates of the intersection point of search band border lines. Both of these relationships are functions of y-co-ordinate in the rectified image.

The *PEL* method was implemented in multiple view configurations. The multiple views are considered as a combination of stereo pairs. The stereo pair correspondences are established based upon which an efficient search is performed to distribute correspondences between multiple views.

Due to the computational effort required for the polynomial coefficient and rectification parameter estimation, the *PEL* algorithm more efficient than *EL* and *3DSI* methods when the number of points are larger than a certain value which is dependant on the implementation and the speed of the computer. For this implementation, *PEL* method is faster than *EL* and *3DSI* techniques for more than *75* and *25* points respectively.

Chapter 5

Simulation tests to evaluate the performance of *PEL* algorithm

In this chapter, results of the simulation tests are illustrated. First part of the tests analysed the effects of random errors of on the establishment of correspondences. For this purpose, random errors of varying magnitudes were introduced to image points and camera exterior parameters. Second part of the tests investigated the performance of the *PEL* algorithm under various operating conditions. Finally comparisons were made with the *EL* algorithm.

5.1 Description of the simulation system

The simulation system can be used to simulate camera, *3-D* objects and acquired images. Number of cameras can be generated in a circle configuration (Figure 5.1). The radius of the circle *(AC)* and the height *(AB)* can be specified. The initial orientation of cameras is such that all cameras point towards the origin *(B)* of the object space co-ordinate system. The exterior parameters of any camera can be altered to obtain new location and rotation parameters. Any number of randomly distributed *3-D* object space points can be generated which can be imaged onto the cameras. The sensor size and focal length of cameras can be specified. Image point correspondences using the *PEL*, *EL*, or *3DSI* algorithms can be established. *3-D* reconstruction can be performed using the non-rigorous direct intersection technique. Furthermore, random errors in *2-D* image point locations and camera exterior parameters can be generated.
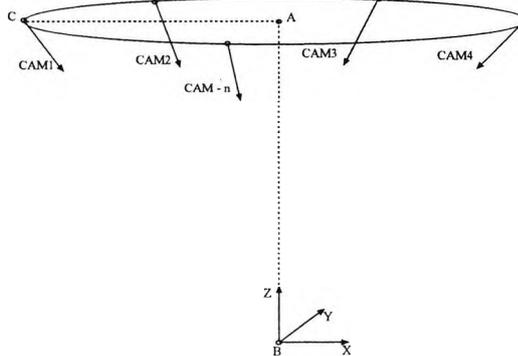

Figure 5.1 Simulation system configuration

A series of simulation tests were carried out. The initial tests analysed the effects of random errors in camera exterior parameters and *2-D* image point locations on the projective imaging geometric relationships. The performance of the *PEL* algorithm with image point location and camera exterior parameter errors were investigated. Finally, the *PEL* algorithm was tested with complex *3-D* object space situations.

5.2 Analysis of the effect of random errors in camera exterior parameters and *2-D* image point locations on the establishment of correspondences

In the high-precision close-range arena, the camera interior and exterior parameters are estimated with a high degree of accuracy. Image points of interest are located to sub-pixel accuracy and are corrected for systematic errors. However, there may still be errors in these quantities. The target image locations may have errors due to geometric and radiometric distortions and other random effects. Unknown lens distortion errors may exist due to modelling deficiencies. These *2-D* errors will be reflected in the adjusted camera exterior parameters. In certain continuous measurement applications, *3-D* location estimations may be required using non-rigorous direct intersection. In such a situation it is usual to obtain accurate exterior parameters using a set of accurately measured control points. Using these known exterior parameters, *3-D* measurements may be produced continuously. However, the exterior parameters may not be as rigorous as those produced by a continuous adjustment process. Another more coarse situation with respect to camera exterior parameters occur when resection is performed with poorly measured control points. The errors in control points will be reflected in exterior location and orientation parameters. However, a correspondence algorithm should be capable of performing correctly with these errors present up to a reasonable level.

In this chapter the effects of these errors on the establishment of correspondences are investigated in two stages. In the first stage, the errors due to coarse situations are investigated which encompasses the situations such as the use of exterior parameters obtained using coarsely measured control points. These errors are relatively large. In the second stage, the effect of smaller errors that occur during a continuous adjustment process are investigated.

5.2.1 Analysis of the effect of large errors on the establishment correspondences

In coarse situations with large exterior orientation errors the magnitudes of these errors are not known hence, it is not possible to determine the effect on the epipolar constraint. But in this section a series of simulations are carried out where magnitude of errors involved were known. Based upon this knowledge the effect on the epipolar constraint was investigated. Figure 5.2 illustrates a situation with large errors and their effect on the epipolar constraint. Points O_1 and O_2 are the correct optical centres of images I_1 and I_2. The projections of 3-D point A_i on images I_1 and I_2 are at points p_{1i} and p_{2i}. The epipoles are at points e_1 and e_2 and image points p_{1i} and p_{2i} lie on conjugate epipolar lines l_1 and l_2. If camera exterior parameters have errors, the optical centres may be considered to be at points O'_1 and O'_2 and the image planes may be slightly shifted (not shown). Hence, the epipoles may be computed as being at points e'_1 and e'_2. The new conjugate epipolar lines are l'_1 and l'_2. When random 2-D errors are present, rays $A_i p_{1i}$ and $A_i p_{2i}$ will not be straight lines. Hence, there will not be an epipolar plane consisting points A_i, O_1, O_2, p_{1i}, and p_{2i}. With these errors present, in order to correctly detect all potentially corresponding candidates on epipolar line l_2 for point p_{1i} in image I_1, a search band with parallel border lines l_{1bi} and l_{2bi} in image I_2 needs to be considered. The width of the band depends on the level of the errors.
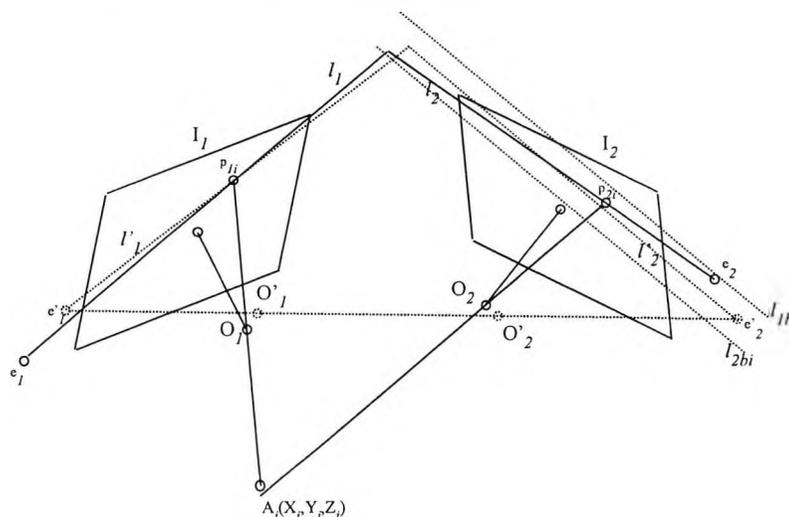

Figure 5.2 Errors in camera exterior parameters and epipolar constraint

A four camera network (Table 5.1) with random exterior parameters was used for a series of simulation tests. Over 1000 3-D object space points were created and imaged onto the cameras. Varying amounts of normally distributed random errors were

introduced to camera exterior parameters and to *2-D* image points separately and then simultaneously. The object space points were reconstructed and the image residuals were computed. In this analysis, image residuals means the difference between the original image point locations and those that are computed by the back-projections from the reconstructed *3-D* points. The effect of the errors on the epipolar constraint in convergent and rectified image spaces were analysed.

Camera	X (mm)	Y (mm)	Z (mm)	ω (deg)	φ (deg)	κ (deg)
1000	5030.7640	-4.3070	3063.0021	3.2558	60.7303	11.2267
1001	4774.2028	1549.9386	2987.0878	-25.3495	51.2202	-4.4002
1002	-1407.6632	4775.0368	3012.9010	-59.2601	-11.3313	-16.2021
1003	-4685.4662	-1528.9630	2645.0879	30.2271	-55.2077	0.1600

Table 5.1 Camera exterior parameters without errors

5.2.1.1 Effect of random errors in camera location parameters

Increasing amounts of random errors were added to camera location parameters only. A *3-D* reconstruction was made using direct intersection and the image residuals were computed. Figure 5.3 illustrates the effect of random errors on residuals.
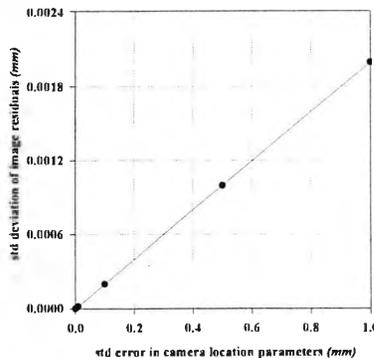

Figure 5.3 Effect of camera location parameter errors on image residuals

It can be seen that there is a linear relationship between camera location errors and image residuals. As the location errors increase, the image residuals increase linearly. The camera location error of *1.0mm* could produce image residual of approximately *0.002mm*.

5.2.1.2 Effect of random errors in camera orientation parameters

The same test was carried out with random errors added to the camera rotation parameters. It can be seen in Figure 5.4 that the effect on image residuals due to errors

in rotation parameters is larger than that due to the errors in location parameters. There is a linear relationship between image residuals and errors in rotation parameters. Standard error of *2.5 degrees* in rotation parameters can produce image residuals of about *1.0 mm*.
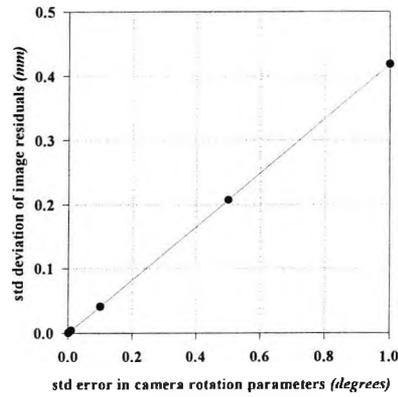

Figure 5.4 Effect of camera rotation parameter errors on residuals

5.2.1.3 Effect of random errors in both camera location and rotation parameters

In this test random errors were added to both location and rotation parameters. For a fixed value of rotation parameter error, the location parameter errors were increased and the residuals were noted (Figure 5.5). The test was repeated for a number of values of rotation parameter error values. An approximately linear variation of residuals is shown in all cases. For higher values of errors in rotation parameters, there is only a slight variation of residuals as the error in the location parameters increases. With standard errors of *1.0mm* and *0.1 degrees* in camera location and rotation parameters, there can be a standard deviation of *0.04mm* in image residuals.
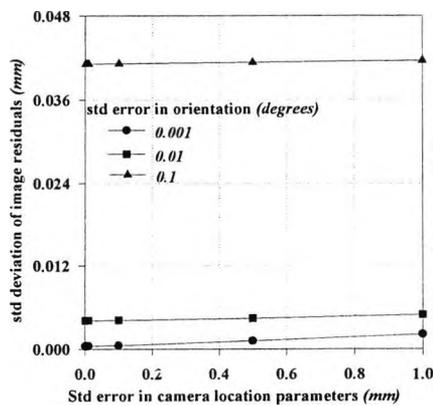

Figure 5.5 Effect of errors in camera exterior parameter on image residuals

5.2.1.4 Effect of random errors in camera exterior parameters on epipole locations

The errors in camera exterior parameters cause errors in computed epipole locations. Tests were carried out where increasing amounts of random errors were added (in the range *0-1.0 mm*) to location parameters of stereo pairs *1000-1001*, *1001-1002*, *1002-1003*, and *1003-1000* with a fixed error level in the rotation parameters and the errors in epipole locations were noted for each image of the stereo pairs. Tests were carried out for rotation errors of *0.001*, *0.01*, and *0.1degrees* (Figure 5.6). The stereo pairs *1000-1001* and *1001-1002* have the smallest and the largest angular separations.
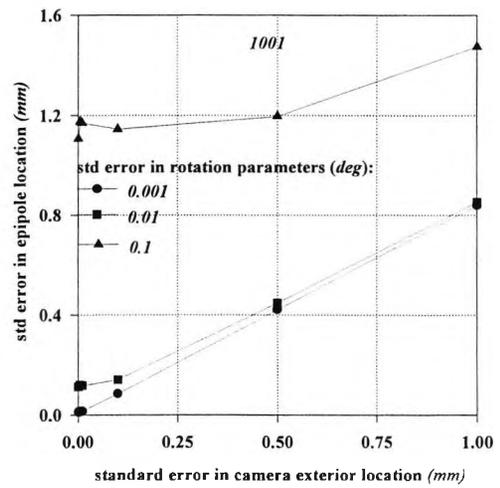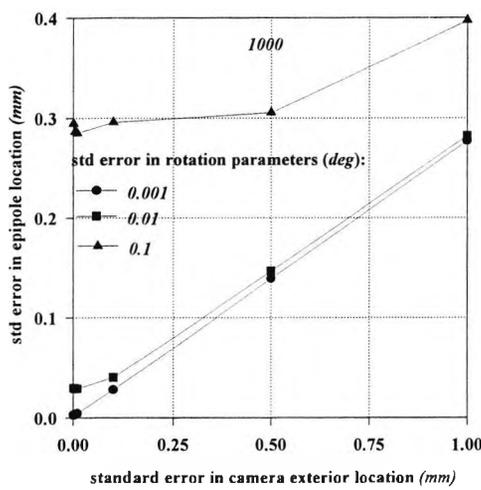

(a) *1000-1001*
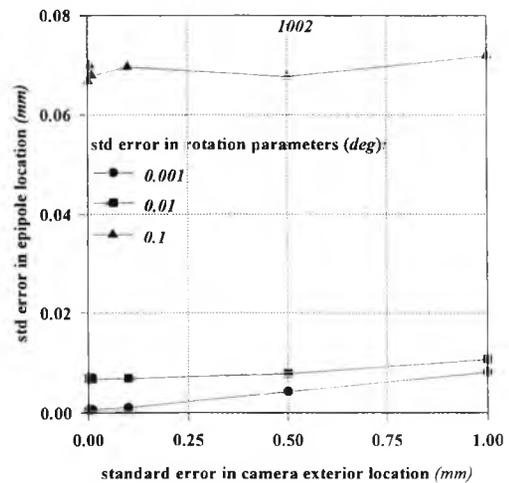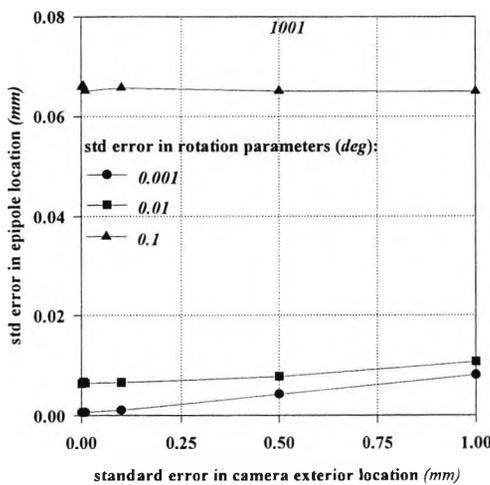
(b) *1001-1002*
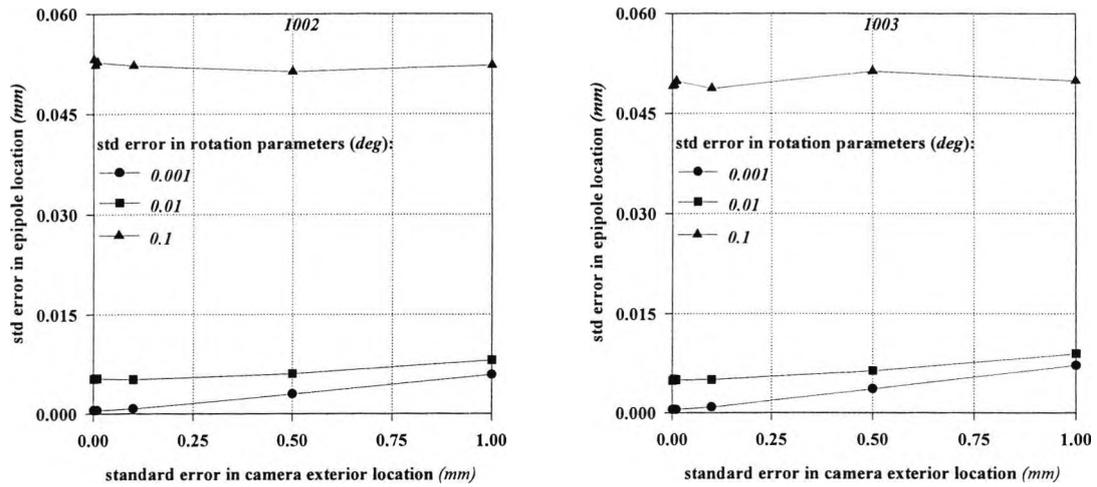
(c) *1002-1003*
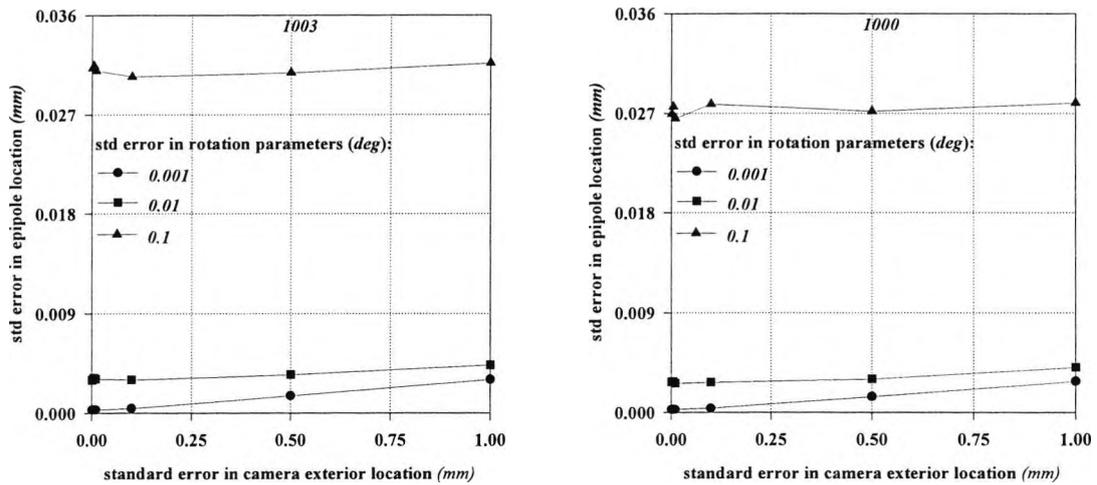
(d) *1003-1000*

Figure 5.6 Effect of errors in camera exterior parameters on epipole locations

It can be noted that when the angular separation is smaller, the epipole location errors are larger. In stereo pair *1000-1001*, for a location parameter error of *1.0mm* and a rotation parameter error of *0.1degrees*, the epipole location error can be about *0.4mm* in *1000* and *1.5mm* in *1001*. The stereo pair *1003-1000* with the largest angular separation has the smallest epipole location errors. For the same level of errors in exterior parameters, epipole location errors can be about *0.032mm* in *1003* and *0.028mm* in *1000*.

When the standard error in camera location errors are in the region of *0 - 0.1mm*, a tenfold increase in the standard error of rotation parameters causes approximately a

tenfold increase in the standard deviation of epipole location errors. This true for each stereo pair considered.

In certain stereo pairs, the standard errors in epipole locations are approximately equal but not in the others. For instance, the stereo pair *1001-1002* has approximately equal standard errors and in stereo pair *1000-1001* standard errors are significantly different. The reason is the angles that each image plane of a stereo pair makes with the base line. When these angles are approximately equal, the standard errors of epipole locations are approximately equal.

5.2.1.5 Effect of random errors in camera exterior parameters and *2-D* image point locations

The presence of relatively large random errors in camera location and rotation parameters and *2-D* image point locations is a realistic situation in practice during the initialisation process. Simulation tests were carried out with random errors added to these parameters and image point locations. Increasing amount of errors were added to camera location parameters with the fixed level of random errors in rotation parameters and *2-D* image point locations. The image residuals were noted. The tests were carried out for a number of values of random errors in rotation parameters (*0.01*, *0.1*, and *0.5 degrees*). The *2-D* image location errors were kept fixed at *0.0001mm* for all three tests. Figure 5.7 illustrates the variation of the standard deviation of image residuals for standard errors in the camera location parameters in the range *0 - 1.0 mm*.
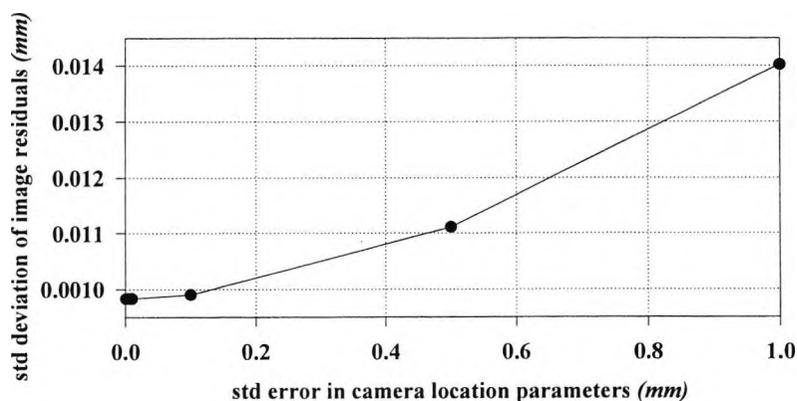
(a). std error in rotation parameters and *2-D* image points = *0.01 degrees*,

0.0001mm
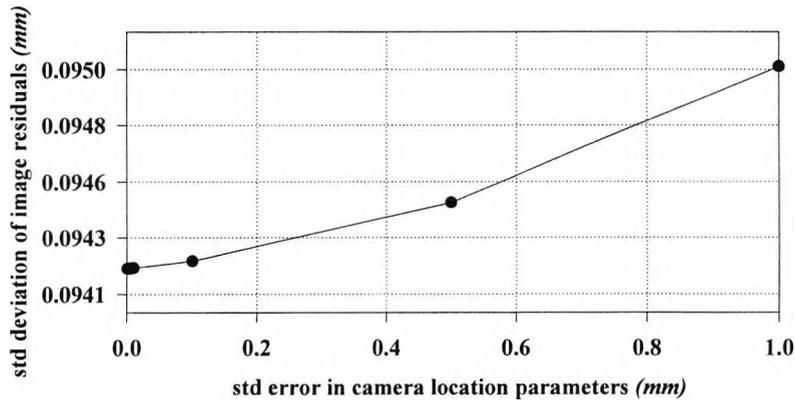
(b). std error in rotation parameters and *2-D* image points = *0.1 degrees,*

0.0001mm
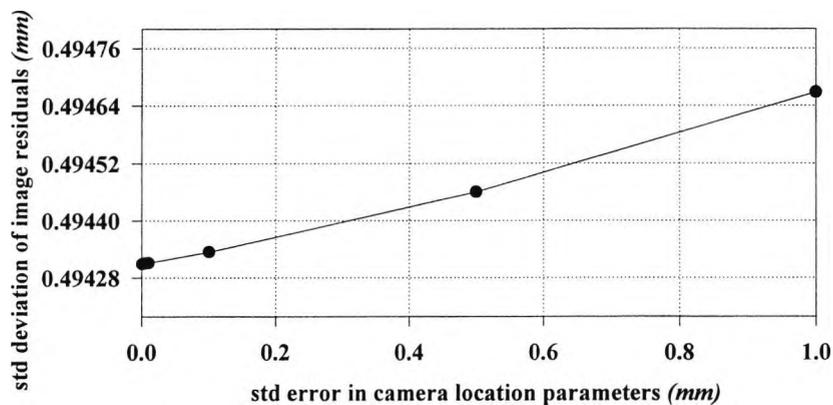
(c). std error in rotation parameters and *2-D* image points = *0.5 degrees,*

0.0001mm

Figure 5.7 Effect of camera exterior parameter and *2-D* errors on residuals

5.2.1.6 Estimation of search band width based on exterior parameter errors and image residuals

Using the above analysis of the standard errors in epipole locations and the image residuals, the search band widths for stereo pairs *1000-1001, 1001-1002, 1002-1003,* and *1003-1000* can be determined. An example on the computation of approximate band width based on these errors is illustrated here for the stereo pair *1000-1001*. When the standard error in camera location parameters are *0.1mm*, the standard errors in epipole locations of cameras *1000* and *1001* are approximately *0.04mm* and *0.14mm* (Figure 5.6 (a)). In this instance, the standard errors in rotation parameters were

0.01degrees (Figure 5.6 (a)). For these levels of error the standard deviation of image residuals was approximately *0.01mm* (Figure 5.7 (a)). The maximum possible epipole location errors and image residuals can be taken as three times their standard deviations. Therefore, the maximum errors of epipoles in images *1000* and *1001* are approximately *0.12mm* and *0.42mm*. The maximum possible image residual is approximately *0.03mm*. It can be seen that the possible error in epipole locations in each image is larger than the image residuals. Hence, the search band width in each image can be taken as twice the value of epipole location errors, *i.e. 0.24mm* for image *1000* and *0.84mm* for image *1001*.

Table 5.2 illustrates the search band width required for each image in stereo pairs *1000-1001*, *1001-1002*, *1002-1003*, and *1003-1000* for the same levels of errors in exterior parameters.

Camera pair	Camera	Std error in camera location (mm)	Std error in camera orientation (degrees)	Std dev of image residuals (mm)	Std error in epipole location (mm)	Required search band width
1	1000	0.1	0.01	0.01	0.0401	0.2406
	1001	0.1	0.01	0.01	0.1403	0.8418
2	1001	0.1	0.01	0.01	0.0066	0.0396
	1002	0.1	0.01	0.01	0.0069	0.0414
3	1002	0.1	0.01	0.01	0.0052	0.0312
	1003	0.1	0.01	0.01	0.0050	0.0300
4	1003	0.1	0.01	0.01	0.0030	0.0180
	1000	0.1	0.01	0.01	0.0028	0.0168

Table 5.2 Estimated search band width when standard exterior location and rotation parameter errors are *0.1mm* and *0.01 degrees* and standard deviation of image residuals are *0.01mm*

The preceding simulation test results provide useful information regarding the effect of camera exterior parameters and *2-D* location errors on search band in the convergent image space. It can be noted that as exterior parameter errors increase the epipole errors increases as well as the required band width. In general, for the same

level of errors in camera exterior parameters and *2-D* locations, the required search band width increases as angular separation decreases.

The other factor noted was that for stereo pairs with wide angular separations and smaller errors in exterior parameters, the image residuals become dominant error for determining the search band width. For instance, in stereo pair *1001-1002* when the standard errors in exterior location and rotation errors are *0.001mm* and *0.01degrees* the standard errors in epipole location are about *0.0064mm* and *0.0069mm* (Figure 5.6(b)). The standard deviation of image residuals are *0.0094mm* (Figure 5.7(a)) and is the dominant of the errors. Hence, the search band width would be *0.056mm.*

### 5.2.1.7 Effect of exterior parameter errors on epipolar lines in rectified images

In the ideal case, the conjugate epipolar lines in a pair of rectified stereo images are collinear. When there are errors in exterior parameters of each camera which are different from camera to camera, these ideal collinear epipolar lines do not exist. In Figure 5.2 where correct locations of an stereo pair are O_1 and O_2, the locations that are considered as correct are O'_1 and O'_2. Hence, the rectified image plane will be determined as parallel to the base line $O'_1O'_2$. The rectification parameters will be determined with errors in them. As a result the points that are on conjugate epipolar lines in convergent images will be mapped into locations which do not meet the collinear epipolar line criteria.

Simulation tests were carried out with a number of stereo pairs (Table *5.1*) to investigate this effect. A sample of over *1000* points was generated in the *3-D* space and imaged onto the stereo pair of interest. The points were transformed to rectified image planes and the ideal locations were noted. The random errors were introduced to image points. The image points were transformed to rectified image planes with random errors introduced to camera exterior parameters each time. The error in rectified image points were noted and standard error was computed for each image of stereo pairs. This test was repeated for standard errors of camera orientation parameters of *0.01* and *0.1degrees*. Figure 5.8 illustrates the results.
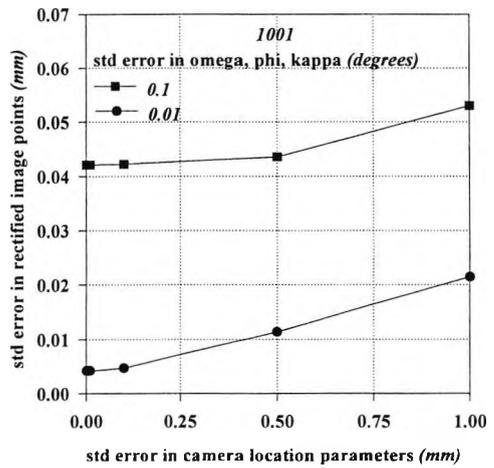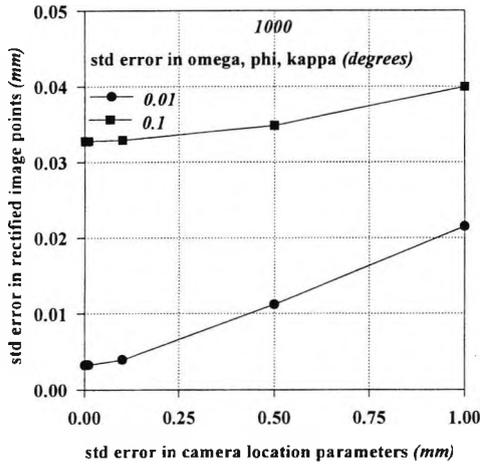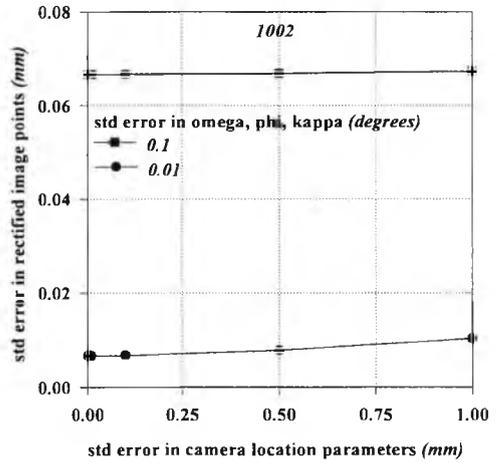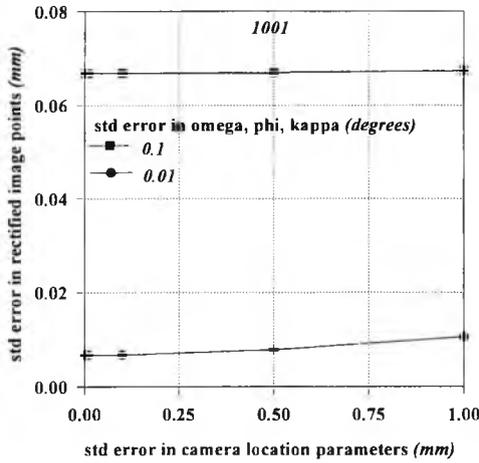
(a). Pair *1000 - 1001*

(b). Pair *1001 - 1002*
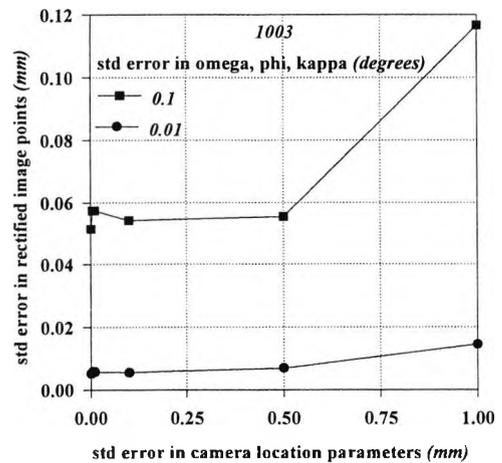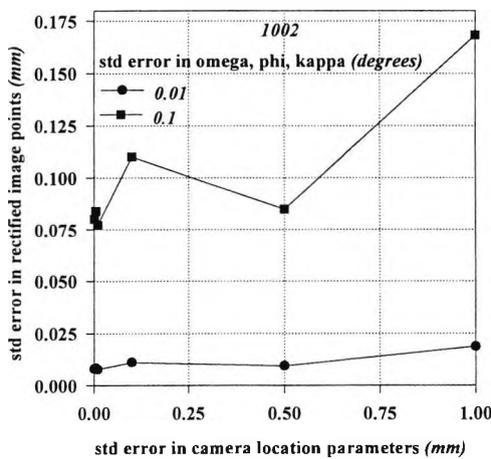
(c). Pair *1002 - 1003*
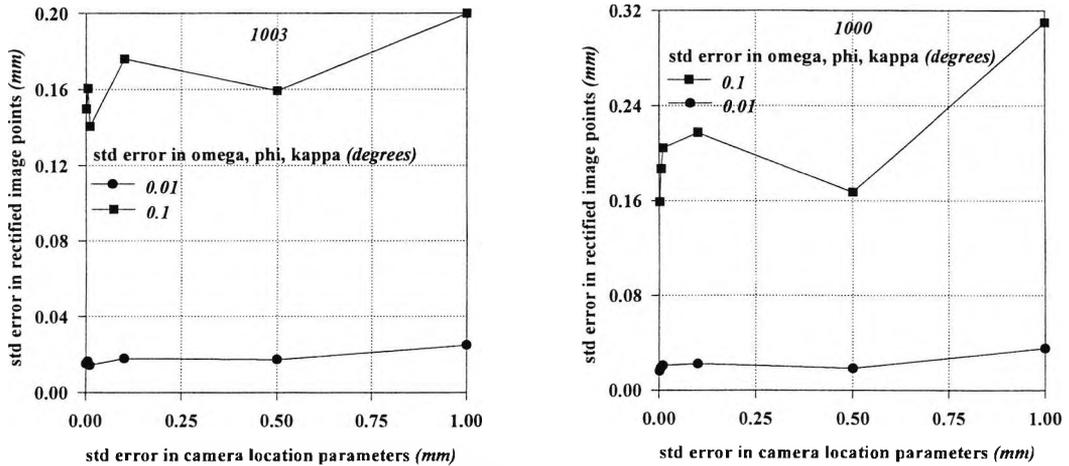
(d). Pair *1003 - 1000*

Figure 5.8 Effect of errors in camera exterior parameters on transformed co-ordinates

It can be seen that in each image of stereo pairs, for a given value of the standard error of orientation parameters, the variation of the standard error in rectified co-ordinates is different as the errors in the camera location are increased. There is an approximately linear variation in standard errors of rectified co-ordinates when errors in the orientation parameters are smaller and is non-linear when errors in the orientation parameters are higher. For a given value of standard error in location parameters in a the range *0-0.1mm* (approximately), a tenfold increase in rotation parameters causes a tenfold increase in the rectified image point location errors.

The most important is that for a given stereo pair and for a given standard error in camera location and rotation parameters, the standard errors in rectified image points are different. This causes conjugate epipolar lines to be non-collinear. In the case of stereo pair *1003-1000*, for standard errors in camera location and rotation parameters of *1mm* and *0.1degrees* the standard errors in rectified image point locations are approximately *0.2mm* and *0.3mm*.

In practice, there exist errors in image point locations too. But these tests illustrate the effect on the transformation of fixed points with the presence of errors in exterior parameters. In order to determine the search band in rectified image space fixed points are transformed from convergent to the rectified images. Two points that fall on each

border line of the search band are transformed into the rectified image space to obtain the equivalent search band. Hence, it is likely that there will be an error in the slopes of search band border lines.

5.2.1.8 Effect of errors in exterior parameters and *2-D* locations on transformed co-ordinates

This is a realistic situation in practice. The test carried out in section *5.2.1.7* was repeated with random errors , in the range *0.0001-0.0009mm,* added to the *2-D* image point locations. Results are illustrated in Figure 5.9 for the stereo pair *1000 - 1001* with different levels of rotation parameter errors.
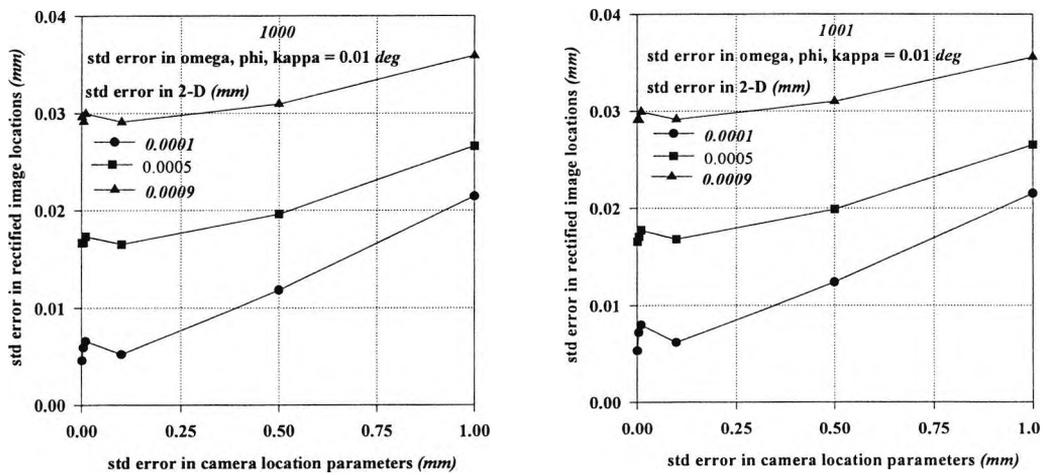
(a). Standard error in camera rotation parameters = *0.01 degrees*
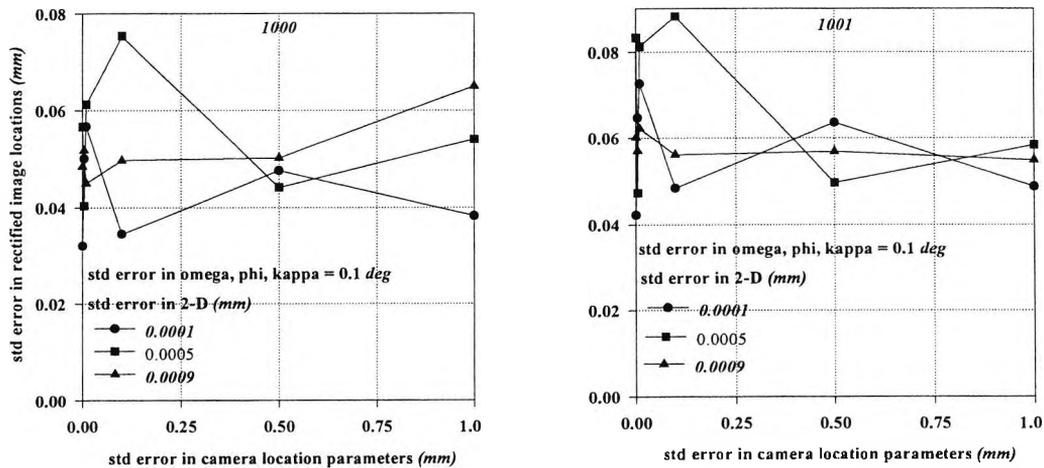
(b). Standard error in camera rotation parameters = *0.1 degrees*
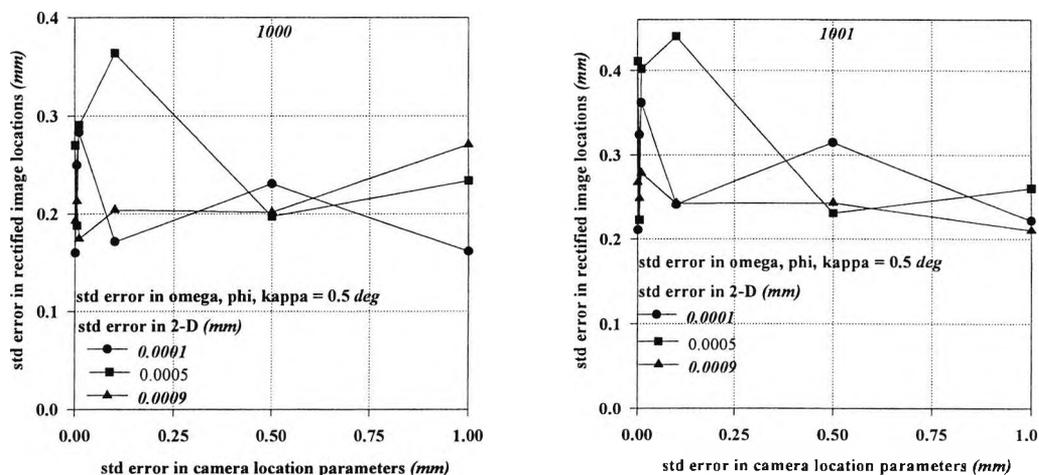
(c). Standard error in camera rotation parameters = *0.5 degrees*

Figure 5.9 Effect of camera exterior parameter and *2-D* errors on transformed co-ordinates

It can be noted that for smaller values of errors in rotation parameters the errors in transformed co-ordinates are approximately equal in each image (Figure 5.9(a)) and the variation is also approximately linear.

Although the search band in convergent images may be determined according to the errors in convergent image space, the transformation errors may cause errors in the transformed search band as discussed in section 5.2.1.7. Hence, potentially corresponding points might fall out of the search band which is most likely to occur when the errors are significant.

5.2.2 Analysis of the effect of small errors on the establishment of correspondences

In this section, the effect of small errors on the epipolar constraint is analysed. Small errors in these quantities occur during a continuous adjustment process and their magnitudes are known. It is assumed that the errors in camera exterior parameters are negligibly small. Figure 5.10 illustrates a stereo pair out of a multiple camera network. It is assumed that the network has been adjusted hence, exterior parameters are the most probable values. $A(X,Y,Z)$ is a most probable *3-D* point produced by the adjustment for a set of image measurements. p'_{1i} and p'_{2i} are the corresponding

measured points in images I_1 and I_2. If a ray into each image of the stereo pair is projected, image points p_{1i} and p_{2i} in images I_1 and I_2 can be found which are the most probable image points. Ideally, the most probable image points (p_{1i}, p_{2i}) and *3-D* point (A_i) and the most probable perspective centres (O_1, O_2) are in a single plane. The computed epipoles (e_1, e_2) are also on the same plane. In a continuous adjustment process, these seven points will jitter around but maintain the coplanarity.
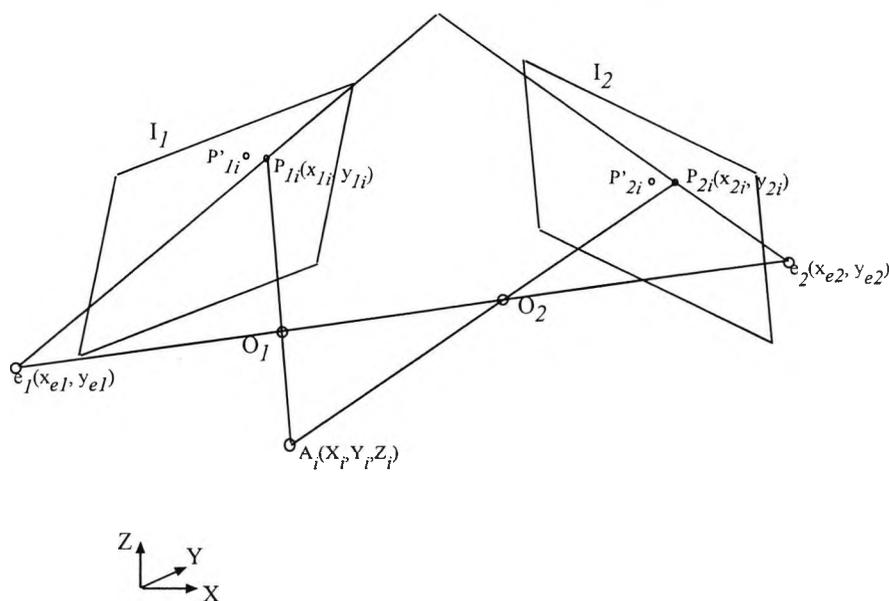
Figure 5.10 Coplanarity of adjusted points

The differences between the measured and computed image points are the residuals. *99.7%* of the computed most probable image points should fall within a circle of radius of three times the standard deviation (3σ) of image residuals. σ denotes the standard deviation of the image residuals. Figure 5.11 illustrates the effect of residuals on the application of the epipolar constraint. Assume that the imaged *3-D* point is on the *XY*-plane of the *3-D* world co-ordinate system. The most probable image points (p_{1i}, p_{2i}), *3-D* point (A_i), perspective centres (O_1, O_2) and computed epipoles (e_1, e_2) are on a single plane. In order to determine the conjugate epipolar line a ray from measured point p'_{1i} is projected onto the *XY*-plane of the world co-ordinate system. From this point a ray is then projected into image I_2 and a point p''_{2i} can be determined. The line (l''_{2i}) joining epipole e_2 and point p''_{2i} is taken as the conjugate epipolar line which does not coincide with the epipolar line (l'_{2i}) created by point p'_{1i} and epipole e_2. The most probable image point p_{1i} and p_{2i} will fall within the circles c_1 and c_2 whose radius is 3σ and the centres are p'_{1i} and p'_{2i}.
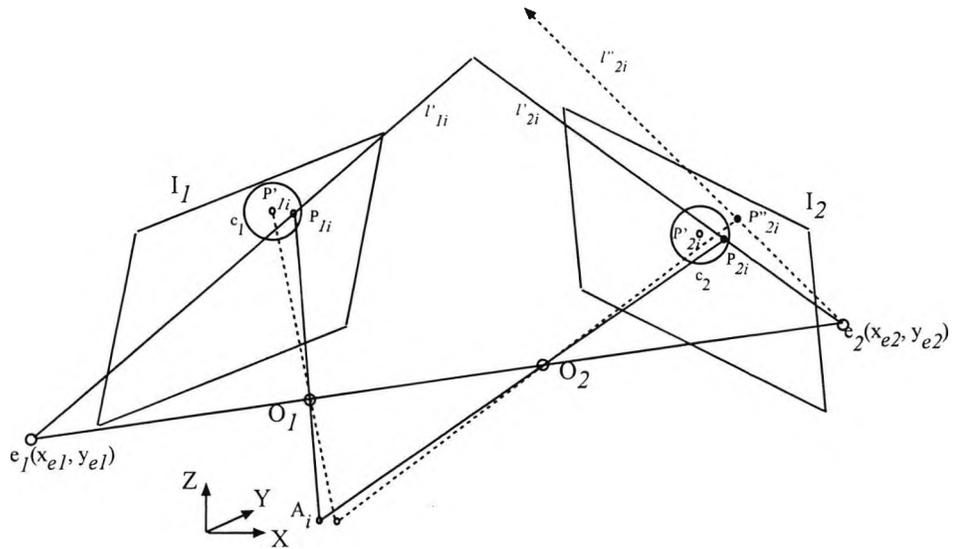
Figure 5.11 Effect of *2-D* residual on epipolar constraint

This information and some statistical analysis can be used to determine the search band width in image I_2. Figure 5.12 illustrates the worst case situation. A most probable point with the largest residual could fall on the perimeter of the circle c_1 whose radius is 3σ. The projected points onto XY-plane from points that fall on the perimeter circle c_1 may generate an ellipsoidal shape es_1. The points now projected onto image I_2 from ellipsoid es_1 may generate an ellipsoid shape es_2.
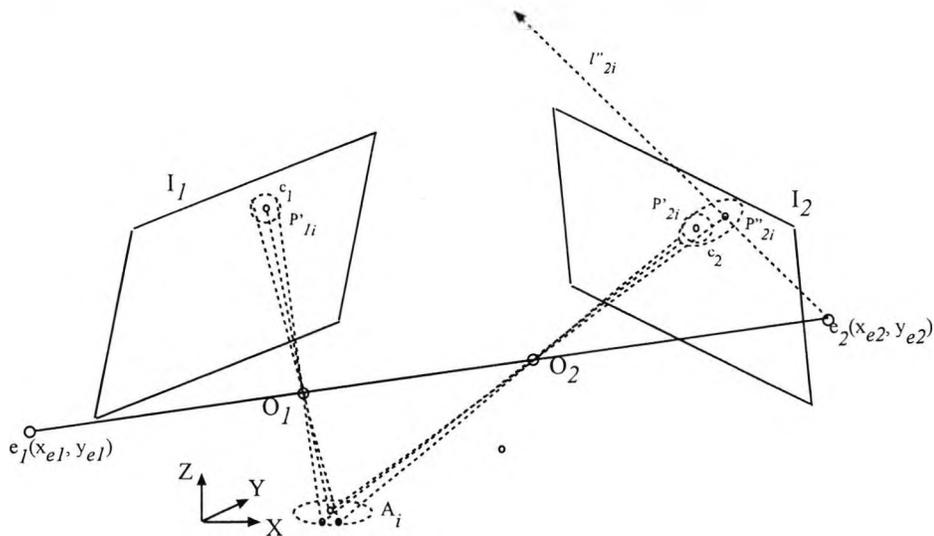
Figure 5.12 Illustration of possible most probable points around a measured point and their projections on *XY*-plane and image *I_2*

Although the radius of the circle in image I_1 for any measured point is 3σ, the ellipsoidal shapes on the XY-plane and on image I_2 may vary depending on the

location of measured points in the image I_1. Also ellipsoidal shapes may vary for each stereo pair and are dependent upon the orientation of image planes and the XY-plane.

A simulation test can be performed to find the width of the search band for the conjugate epipolar line l''_{2i}. A sample of points can be generated within the circle c_1 whose radius is 3σ. The corresponding points on image I_2 can be found and standard deviation (σ_l) for the spread around point p''_{2i} can be found. Hence, maximum distance that a most probable point could fall away from p''_{2i} is $3\sigma_l$ and the search bandwidth can be taken as $2 \times 3\sigma_l$ ($6\sigma_l$).

An example on the determination of σ_l is illustrated here using stereo pair *1002-1003* (Table 5.1). A random point p_1 on image *1002* was taken and its projected point p_2 on the image *1003* was determined. A sample of points were generated around p_1 within a standard deviation of $\sigma = 0.0001mm$. The projected points on image *1003* were determined and the standard deviation of the spread around point p_2 was computed. For various values of σ, the corresponding σ_l in image *1003* are illustrated in Figure 5.13.
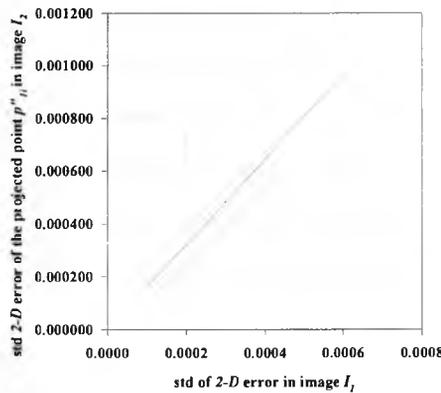
Figure 5.13 Variation of σ_l for various values of σ for a single point

A further test was carried out with the same stereo pair in which a point is moved along an epipolar line in image *1002* with random errors of standard deviation *0.0001mm* added to the point at each location. The standard deviation of their projections on image *1003* were determined. Table 5.3 illustrates the results. As the point in image *1002* is moved towards its epipole, the projected point in image *1003*

moves away from the its epipole. It was noted that the standard deviation of the spread around projection increases as point is moved away from the epipole in image *1003*. Hence, required search band width increases.

Point in image *1002* (*mm*)	Projected point in image *1003* (*mm*)	Standard error of projected point (*mm*)
(3.50, 3.50)	(0.48, 2.39)	0.0001613
(2.50, 2.35)	(0.58, 1.55)	0.0001763
(1.50, 1.20)	(0.71, 0.48)	0.0001979
(0.50, 0.05)	(0.88, -0.95)	0.0002306
(-0.50, -1.10)	(1.11, -2.93)	0.0002830
(-1.00, -1.68)	(1.27, -4.24)	0.0003225
(-1.50, -2.25)	(1.46, -5.87)	0.0003767

Table 5.3 Variation of the standard error of the projected point as it is moved away from epipole

Tests carried out for other epipolar lines on the same stereo pair shows similar results. As the image point is moved towards the epipole in image *1002*, the projected point moves away from the epipole in image *1003* and the standard error increases hence the required search band width must be larger. These results shows that the closest point to the epipole in image *1002* will produce the largest standard error and hence the largest band width. A global search band for this stereo pair can be determined by considering the corner that is nearest to the epipole in image *1002*. Considering a sensor of dimensions *9x9mm*, simulation results show that the maximum standard error in image *1003* was *1.1 μm*. Hence, required global search band with is *6.6μm*.

A simulation test was carried out to test the above results. One thousand *3-D* points were created and imaged onto images *1002* and *1003*. Random noise with standard deviation *0.0001mm* was added to the points in each image. Using the *EL* method, the stereo correspondences were established. In some cases ambiguities were present but the correct point always fell within the search band. Hence, this method of estimating the search band width proved to be correct.

5.3 Performance of *PEL* algorithm in multiple view configurations under ideal conditions

The algorithm was tested in various four view configurations. The purpose of the tests was to ensure that the algorithm performed correctly under each situation. The results from two of these tests are illustrated for which the four camera configuration given in Table 5.1 was used. The focal lengths of convergent axis images were *16mm*.

Test-1:

Over *1000* targets were randomly generated on a plane which is in the field of view of all cameras. Due to the planar target field ambiguities due to occlusions were not present. Targets were imaged from all four cameras. Figure 5.14 illustrates four images.
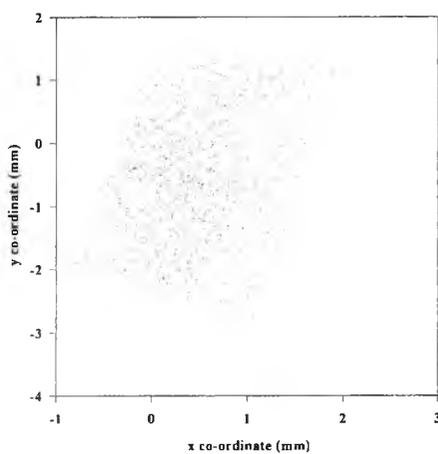
(a). view - 1000
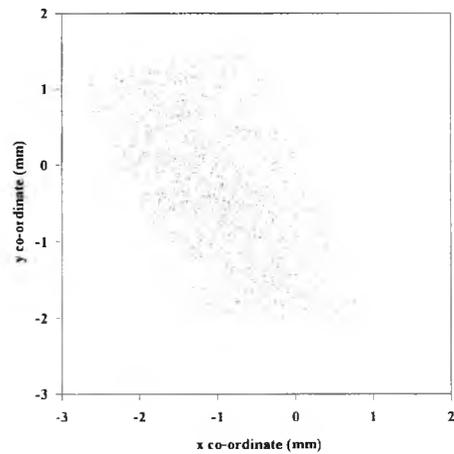
(b). view -1001
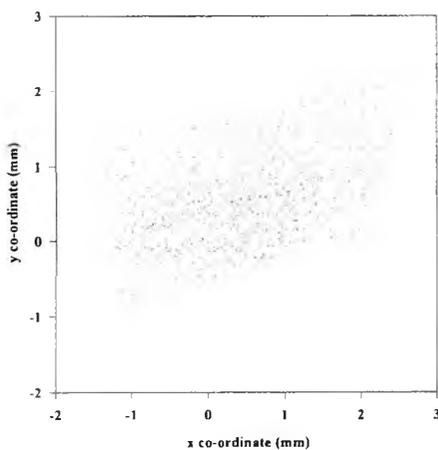
(c). view - 1002
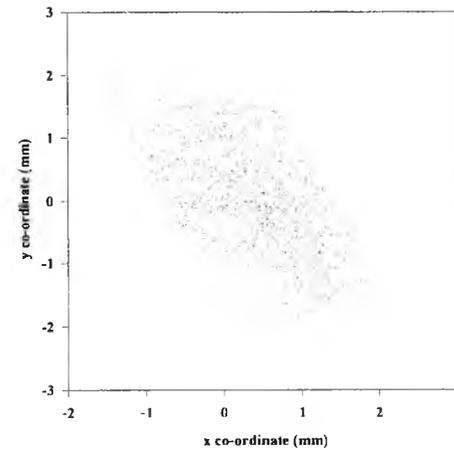
(d). view - 1003

Figure 5.14 Four views of a planar target field

These image data were analysed. The smallest distance between image points in each image were computed (Table 5.4). When there are errors in neither camera exteriors nor image point locations, a search band width which is less than the smallest distance between two points can be used. This ensures that there will not be ambiguities due to multiple points within an intersection of search bands. Hence, a band width of $0.2\mu m$ was selected.

Image No	Smallest distance *(mm)*
1000	0.00021
1001	0.00025
1002	0.00038
1003	0.00024

Table 5.4 Smallest distances between two points in each image

Using this value of search and width, polynomials representing border line slope and the straight lines representing variation of the x-co-ordinate at crossing points of border lines were determined as follows,

pair-1:

$$s(y) = -0.0000000017y + 0.0000012035$$

$$s(y) = 0.0000000017y - 0.0000012035$$

$$x(y) = -0.2405y + 162.7051$$

pair-2:

$$s(y) = 0.0000000005y^2 + 0.0000000043y + 0.0000035275$$

$$s(y) = -0.0000000005y^2 - 0.0000000043y - 0.0000035275$$

$$x(y) = 0.085y + 46.7339$$

pair-3:

$$s(y) = 0.0000000009y^2 - 0.0000000010y - 0.0000042634$$

$$s(y) = -0.0000000009y^2 + 0.0000000010y - 0.0000042634$$

$$x(y) = -0.016y + 34.2886$$

pair-4:

$$s(y) = 0.0000000016y^2 + 0.0000000012y + 0.0000052193$$

$$s(y) = -0.0000000016y^2 - 0.0000000012y - 0.0000052193$$

$$x(y) = 0.0194y + 21.1705$$

The stereo pairs *(1,2)*, *(2,3)*, *(3,4)*, and *(4,1)* were rectified. The rectified image focal length was *32mm*. Figure 5.15 illustrates four pairs of rectified images.
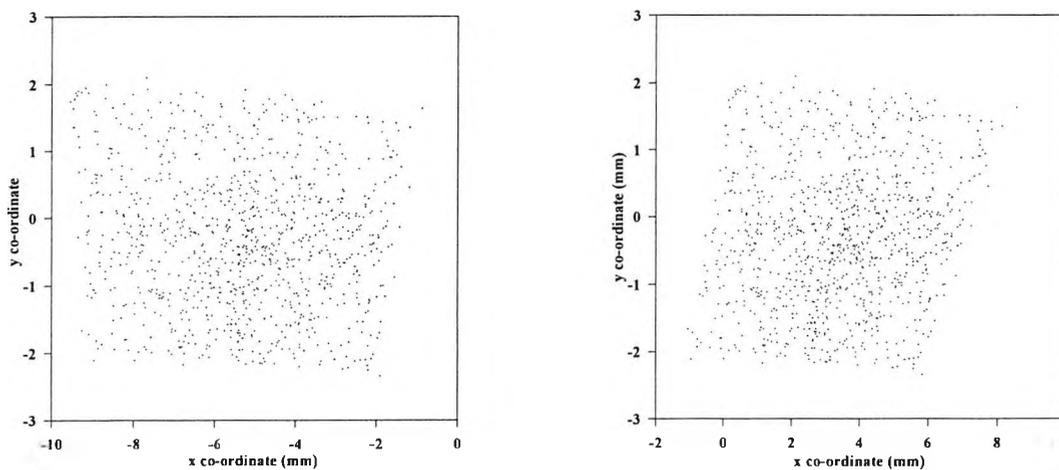
(a). Rectified image pair (1000,1001)
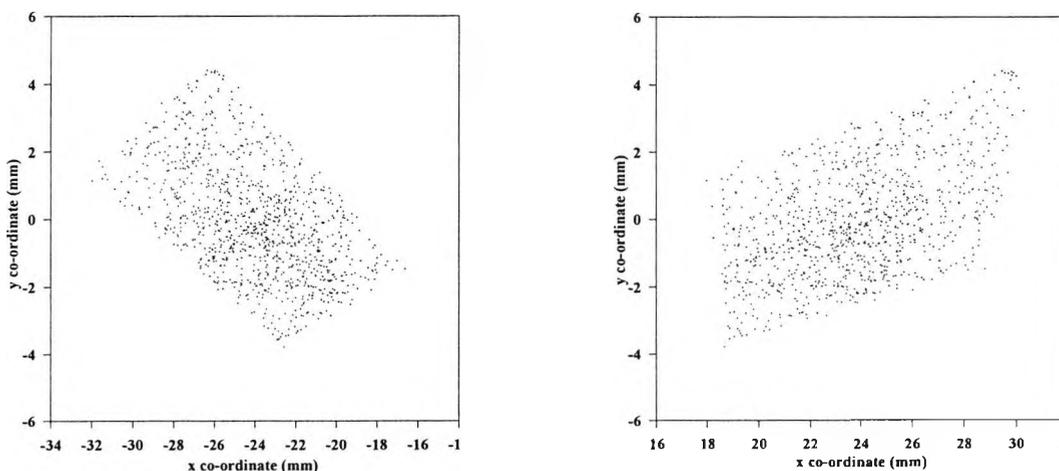
(b). Rectified image pair (1001,1002)
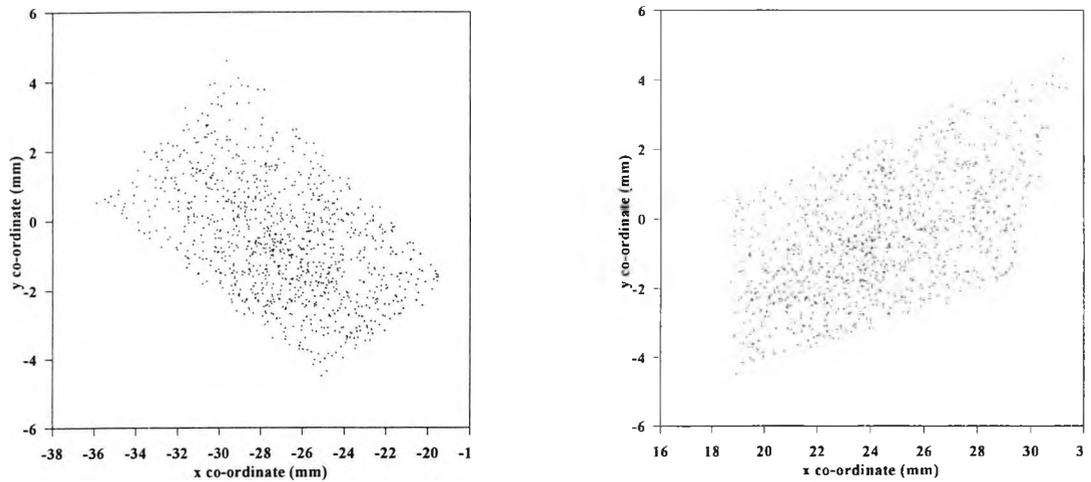
(c). Rectified image pair (1002, 1003)
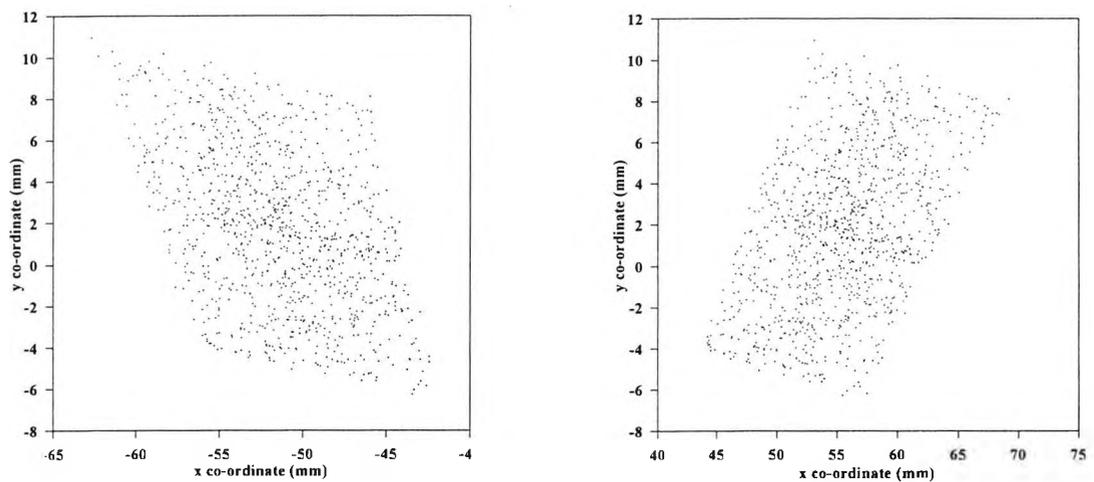
(d). Rectified image pair (1003, 1000)

Figure 5.15 Rectified stereo pairs

Correspondences were established using the *PEL* algorithm. As expected there were no ambiguities and correspondences were established between all image points. The *3-D* locations of points were computed using direct intersection. For each computed *3-D* point, image residuals were also computed. The correct correspondences were confirmed by zero image residuals in this case. Figure 5.16 illustrates the reconstruction of the planar target field.
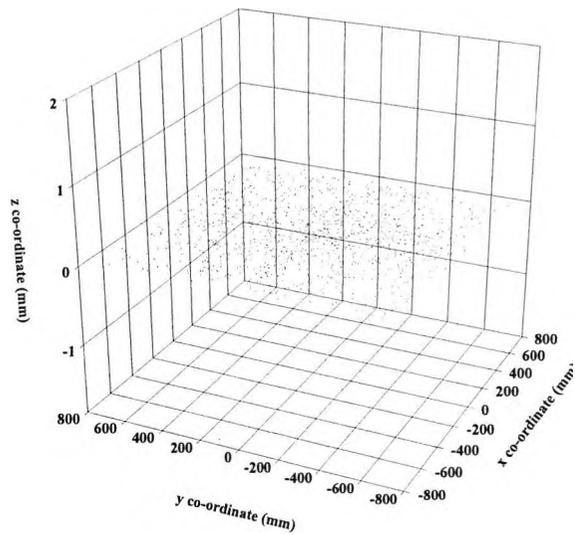
Figure 5.16 *3-D* **reconstruction of planar target field**

Test-2:

In this test, the performance of *PEL* algorithm was evaluated with the presence of a complex *3-D* object space situation. Ambiguities that occur due to occlusions or points being out of the field of view of some images and appearing in others were present. The camera network given in Table 5.1 was used here. A set of *1500* points that are randomly located within an area of (*1.5x1.5x1.5*) m^3 in the field of view of cameras was created (Figure 5.17).
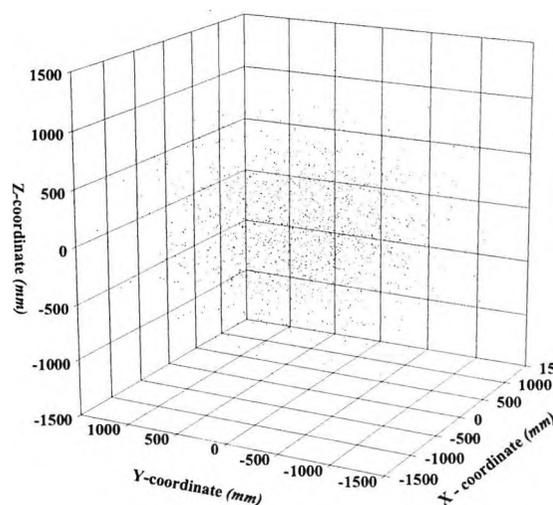
Figure 5.17 Randomly distributed *3-D* targets

Images of these points were obtained from four views (Figure 5.18). Table 5.5 contains number of points in each image and the distance between the closest two points in each image. The number of points in each image was different. One of the reasons was due to the some points appearing in certain images but not in others because of *3-D* points being out of the field of view. The other reason was due to occlusions. The distances between the closest points was considered to determine the search band width which was set to be *0.001mm.*
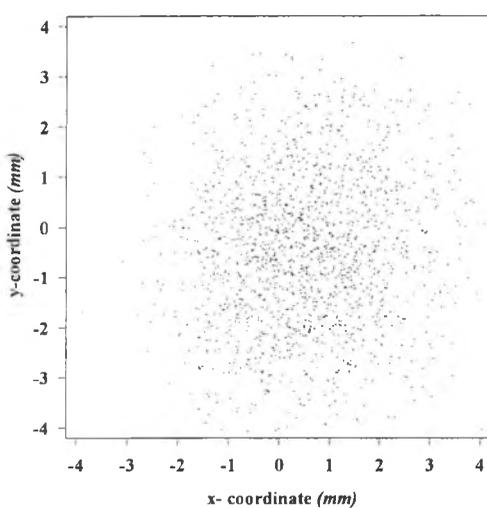
(a). Camera - 1000
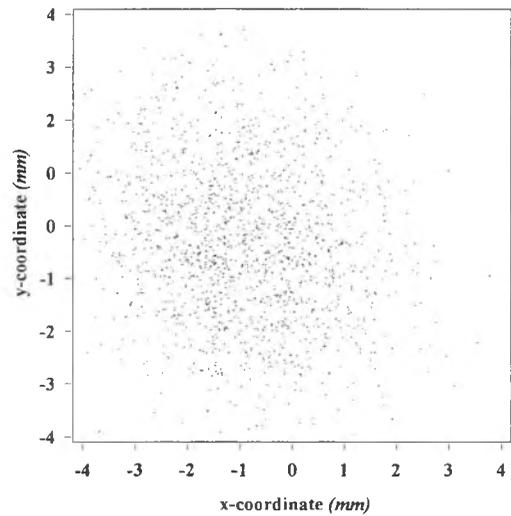
(b). Camera - 1001
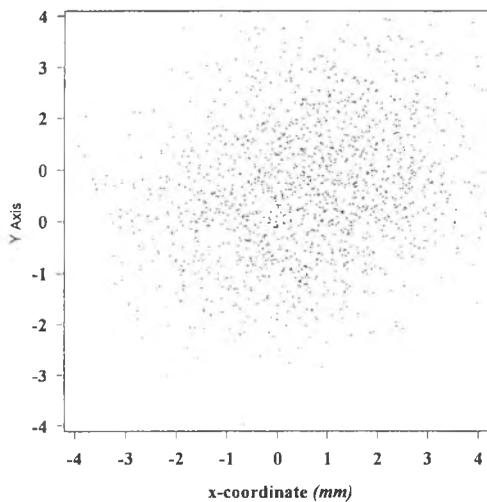
(c). Camera - 1002
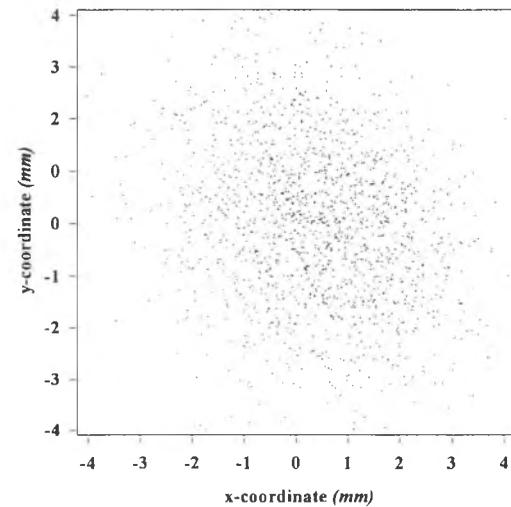
(d). Camera - 1003

Figure 5.18 Images of randomly distributed *3-D* object space points

Camera	1000	1001	1002	1003
Number of points	1484	1476	1482	1488
Distance between closest pair of points *(mm)*	0.002214	0.001205	0.002144	0.001164

Table 5.5 Number of image points and distance between closest pair of points

Correspondences were established using the both *PEL* and *EL* algorithms for the purpose of comparison. The results of correspondences were analysed. Table 5.6 illustrates details of established correspondences for each algorithm in which correspondences between all four views, any combination of three views, any combination of two views, and finally any non-corresponding points or single views are given. The results show that both *PEL* and *EL* algorithms have the same performance in this ideal situation where there are no errors in exterior parameters and image point locations.

Algorithm	*4*-views correspondences	*3*-views correspondences	2-views correspondences	single views or non-corresponding
PEL	1444	44	10	2
EL	1444	44	10	2

Table 5.6 Details of correspondences

5.4 Performance of *PEL* algorithm with errors in camera exterior parameters and image point locations

Test-1:

In this test the same camera network given in Table 5.1 and a lower density target field with *100* targets was used. With errors in rotation parameters and image point locations set to fixed values the errors in exterior location parameters were increased. The test was repeated for standard errors in rotation parameters of *0.001*, *0.005*, and *0.01degrees*. The standard error in image point locations was set to *0.0001mm* in all tests. Figure 5.19 (a) and (b) illustrate percentage of correspondences established in four and three views.

(a) Four views   (b) Three views
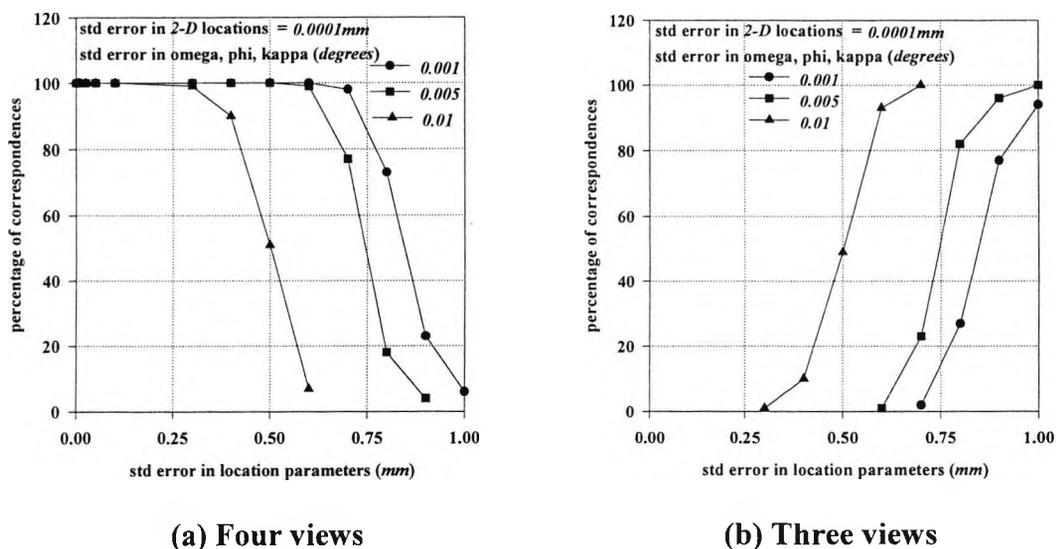
Figure 5.19 Percentage of correspondences with a given search band width and varying errors in camera exterior parameters

When the target field was imaged the shortest distances between two image points was *0.019mm*. Hence, the maximum possible search band width should be less than *0.019mm* and it was set to *0.018mm*. It was noted that as errors are increased beyond a certain value, the number of correspondences in four views decreases and those in three views increases. This test illustrated the establishment of correspondences with a search band of given width under number of levels of errors.

Test-2:

A randomly distributed set of *50* points were created in the field of view of the cameras (Table 5.1). As discussed in section 5.2.2, global search band widths for the second image of the stereo pairs *1000-1001*, *1001-1002*, *1002-1003*, and *1003-1000* were determined assuming *2-D* random error of *0.0001mm*. For this purpose, it was considered that the dimensions of the sensors were *8x8mm*. In the first image of each pair, one of the corner points produced the largest search band width requirement for the second image. The results are illustrated in (Table 5.7). Images were acquired and random noise of standard deviation of *0.0001mm* was added to image points.

Image pair	Epipoles	Global search band width for second image *(mm)*
1000, 1001	(16.76, 79.70), (67.60, -127.48)	0.0002076 x 6 = 0.0013
1001, 1002	(-21.02, 10.21), (23.31, -0.09)	0.0003424 x 6 = 0.0021
1002, 1003	(-10.98, -13.17), (-1.70, 20.86)	0.0007340 x 6 = 0.0044
1003, 1000	(10.16, 2.95), (-8.47, -1.99)	0.0070900 x 6 = 0.0425

Table 5.7 Global search band widths for each image pair

In order to establish the correspondences, a global search band width common for any stereo pair which is the largest of the global band widths for stereo pairs was taken as *0.0425mm*. Correspondences were established between four views for all *50* points. Further tests were carried out with increasing random errors added to image points. It was noted that as the image point errors were increased, the number of correspondences established between all four views decreased.

5.5 Summary of the chapter

This chapter described the simulation tests that were carried out. The tests included the analysis of the effect of errors in camera exterior parameters and image point locations on the application of epipolar constraint. Errors were of two categories. Large errors that could occur due to obtaining camera exterior parameters with poorly measured control points. Small errors occur in a continuous adjustment process. The performance of *PEL* algorithm under a number of situations were investigated and comparisons were made with *EL* algorithm.

For the purpose of analysis, known large errors were introduced to camera exterior parameters. Based on the knowledge of these errors possible search band widths in convergent images were determined for the purpose of illustration. However, these errors are not known in practice hence band widths cannot be determined. The discussion is common to any camera network, the quantities that were noted belong to the network that was used for these tests. For other networks the errors are different.

It was noted that when exterior parameters have large errors, the conjugate epipolar lines in rectified image space were no longer collinear. Hence, it possible that corresponding points may fall out of the search band depending on the magnitude of errors.

The effect of small errors on the application of epipolar constraint was investigated. It was assumed that in a continuous adjustment process, the camera exterior parameter errors were negligible. A method of establishing search band width was developed and tested. It was proved that the method of determining search band widths in a continuous adjustment process as correct using simulation tests.

The performance of the *PEL* algorithm was tested under error free conditions with two object space point distributions. It demonstrated similar performance to that of the *EL* method. This illustrated the accuracy and the stability of the search band estimations using the polynomials.

Chapter 6

Development of a real-time close range photogrammetric system

Many new industrial measurement tasks demand a real-time response. For instance, the robotic assembly of industrial components to the correct geometry is a such task. Each camera viewing the parts of interest provide video frames every *40ms (CCIR standard)*. A frame consists of approximately *438k pixels*. A *10*-bit *A/D* converted video frame is equivalent to *548kbytes*. If a system consists of four cameras the total amount of data is *2.2Mbytes* or *25x2.2Mbytes* of data to be processed in every second. This still represents a requirement for specialist equipment. The system described in this thesis acquires each word of intensity data and the *2-D* processing system extracts and processes information of target images but discards unnecessary background data. Finally correspondences are established and *3-D* location estimations are computed. The speed requirements of the measurement task determine the rate of the supply of processed data which could be less than the video frame rate for most measurement tasks. However, in many new applications such as high precision assembly the frame rate functionality may be essential.

Conventional close-range systems use frame-grabber hardware which is normally a plug-in card to the back-plane bus system of a *PC*. Usually, a frame grabber consists of electronics for acquisition and storage of video data from single or multiple cameras. Multiple cameras may be driven by a single clock source for synchronous acquisition of video data. The currently available frame grabbers are now capable of acquiring video data at the frame rate. Even if it is possible to acquire at frame rate, it may not be possible for a *PC* to process the data at that rate. Another feature of a frame grabber based systems is that video data from number of cameras is often only available sequentially. Due to these reasons, it may not be possible to provide processed data at frame rate by a frame grabber based system with conventional hardware.

In order to overcome the speed of processing problems of frame grabber systems, an intelligent camera based system was developed to enable processing of video data in real-time (Clarke *et al*, 1998; Clarke *et al*, 1997; Gooch *et al*, 1996; Pushpakumara *et al*, 1996). An intelligent camera consists of video processing and data communication hardware and a *CCD* camera. A multiple intelligent camera system exploits the idea of distributed processing. Locally processed video data at each camera are transferred via Ethernet to a central processing computer *(CPC)* for further processing. The distinction between the frame grabber and this system is the method of processing of individual images and the method of data transfer from cameras to *CPC*. The method of establishing target image correspondences between multiple images and the method of *3-D* location estimation are the same for both systems which are performed in the *CPC*. The advantages of this system are multitude. Video data are processed simultaneously at each intelligent camera hence, avoiding the need for sequential processing of images in *CPC*. Assuming an image consists of *100* target images the amount of data transferred to the *CPC* by an intelligent camera is approximately *600* bytes. This is significantly lower amount of data than transferring an image as in the case of frame grabber based systems. As a result, a significant amount of time is saved in processing individual images and data transfers. A further advantage of the systems is the convenient interface to the *CPC* via Ethernet which is a commonly available facility in most *PCs*.

6.1 Intelligent camera system

Each intelligent camera is a modular system with video data acquisition, target recognition, location, and data communication functionalities. Hardware consists of five stackable modules which are connected physically and electrically via stack-through headers. The data bus and control signals run through one *32*-pin header and the power supply runs through another *12*-pin header. The five main modules of an intelligent camera are,

1) Video processing module

2) Input/Output module

3) DSP module

4) Ethernet communications module

5) Power supply module

Figure 6.1 illustrates an intelligent camera hardware, an external Ethernet transceiver, and a *CCD* camera.
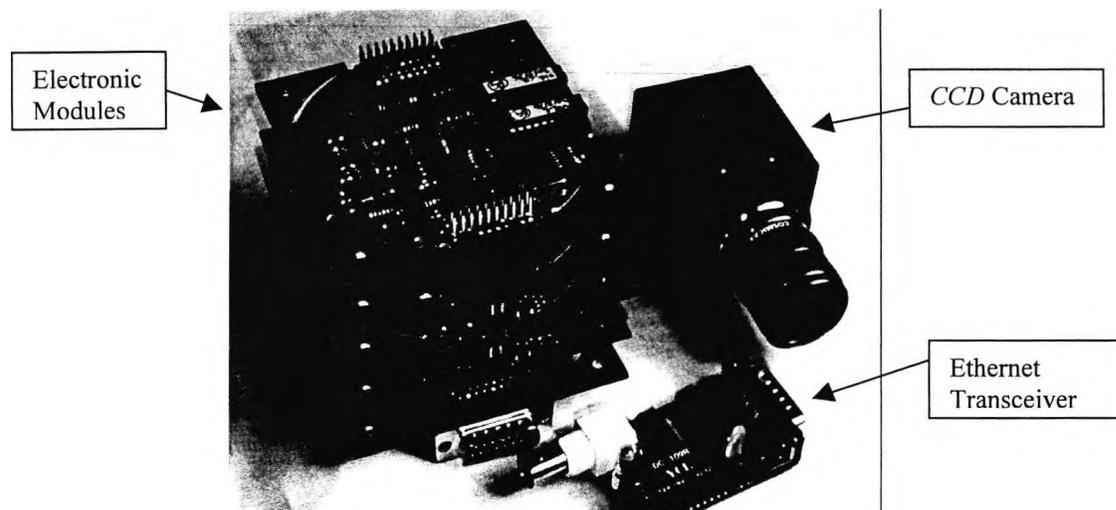
Figure 6.1 Intelligent camera

The history of the development of intelligent camera dates back to November, 1994 when the author started his *MSc* project. A wire-wrap prototype of the *DSP* module was developed and successfully tested (Pushpakumara, 1995) with a *PC* interface. In this prototype the Input/Output hardware, which is currently a separate module, and *DSP* hardware were developed as a single module. The object recognition and location algorithm was developed, implemented, and successfully tested with real and simulated image data. A *PC* version of the algorithm was also implemented for use with the frame grabber. Concurrently, Mr. Stuart Cowhig developed the first wire-wrap prototype of the Video Processing Module.

In late 1995, development of the current modular intelligent camera system commenced. The *DSP* and Video Processing hardware were further developed (Gooch *et al*, 1996). The author did further work on the prototype version of *DSP* hardware and the object processing algorithm implementation was further developed. The author was given the responsibility for the development of Ethernet communication module. A *PC* version of the packet driver software was developed and successfully tested for data transfers between *PCs*. The Ethernet hardware was developed and the *DSP* version of the packet driver was implemented. Ethernet communication between a *PC* and a *DSP* module was fully tested by September, 1996 (Pushpakumara *et al*, 1996). Gooch *et al*, 1996 developed the video initialisation software routine. The author

produced the first version of the complete software for the intelligent camera which performed video data acquisition, object image processing, and Ethernet communication tasks. In January 1997, real-time operation of a complete intelligent camera system was demonstrated with real video data output to a *PC* where target location data was displayed graphically.

Clarke *et al*, 1998 modified and further refined software of the intelligent camera system. The packet driver and communication software on the *PC* side were replaced by a version of Windows Sockets based software. The *DSP* version of the packet driver was modified accordingly to arrive at the currently available four-camera system where a *PC* acquires target location data from each intelligent camera system and performs correspondence establishment and *3-D* location estimation tasks. The rest of this chapter describes the hardware and software development.

6.2 Object image reconstruction, recognition, and location algorithm

The intelligent camera system performs the detection, recognition, and sub-pixel location of object images. Conventional methods require a complete image prior to beginning the recognition and the location of the target images which is not desirable in a real-time system. At any given time, the object processing algorithm developed for this system requires object edge pixel locations and the internal intensity data belonging to two consecutive lines of an image to begin processing (Ellis, 1987; Pushpakumara *et al*, 1996; Pushpakumara, 1995). Hence, object image reconstruction can be performed as each line of edge and intensity data becomes available. As reconstruction proceeds in a progressive manner, the algorithm accumulates the sum of the product of intensities and pixel locations and the sum of intensities for centroid location computation and determines other necessary parameters such as area and peak intensity for recognition for each object image.

Although the target images are approximately of circular or ellipsoidal shape, the algorithm considers object images of all possible shapes to ensure reliable operation and the isolation of good target candidates. Figure 6.2 illustrates part of an image which has object images of various shapes, the edge pixel locations, and internal intensities for each object segment in each line. The algorithm uses a number of logical tests to ascertain the connectivity of object segments found in a pair of

consecutive lines. For this purpose, an object segment encountered in the current line can be categorised with respect to a segment in the previous line as *new*, *finish*, *continue*, *split*, and *merge*. Most of the shapes that are found in practice can be separated into these categories and can be reconstructed.
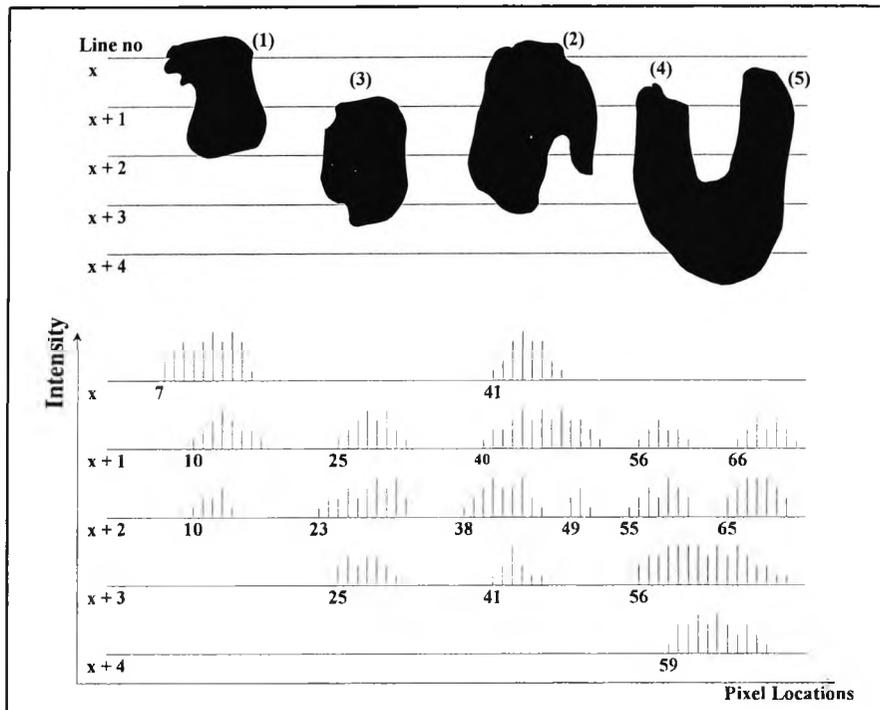
Figure 6.2 Section of an image (top), detected edge locations and *A/D* converted intensities (bottom)

If the pixel locations of starting and finishing edges of a segment are defined as the *starting_edge_location* & the *finishing_edge_location*, five logical tests used for connectivity establishment can be stated as follows,

1). New object test

If previous *starting_edge_location* >= the current *finishing_edge_location*, object segment under consideration in the current line belongs to a new object.

2). Finishing object test

If previous *finishing_edge_location* <= the current *starting_edge_location*, the object segment under consideration belongs to a finishing object.

3). Continuing object test

If tests (1) & (2) are not true *and* the object segments overlap, they belong to the same object, hence are part of a continuing object.

4). Splitting object test

If the current *starting_edge_location* is < previous *finishing_edge_location* the object segment under consideration in the current line belongs to a splitting object image. Every continuing object image is tested for splitting.

5). Merging object test

Every continuing object which does not satisfy test (4) is tested for merging. If the current *finishing_edge_location* is > the previous *starting_edge_location* then the object segment under consideration in the current line belongs to a merging object.

Every object image is reconstructed and is subjected to tests for identifying whether it is an image of a valid target or a spurious object in the background clutter. This test uses the parameters such as area and peak intensity. The centroid is computed for valid target images. The algorithm was tested on various shapes and densities of object images in simulated and real data. Current implementations in the intelligent camera system and *PC* are being used successfully in practice for various day-to-day target recognition and location tasks.

6.3 Hardware development

The intelligent camera system uses a combination of hardware and software to process video data (Clarke *et al*, 1998; Clarke *et al*, 1997; Gooch *et al*, 1996; Pushpakumara *et al*, 1996). The video processing module performs the edge detection of object images and *A/D* conversion of intensity values which are transferred to *DSP* module via the Input/Output module. Using the edge data, the periphery of each object is reconstructed and the gray scale centroid is computed within the *DSP* module. The Ethernet module transfers centroid and other details such as peak intensity and the area of target images to the *CPC*.

6.3.1 Video processing module

95mm

95mm
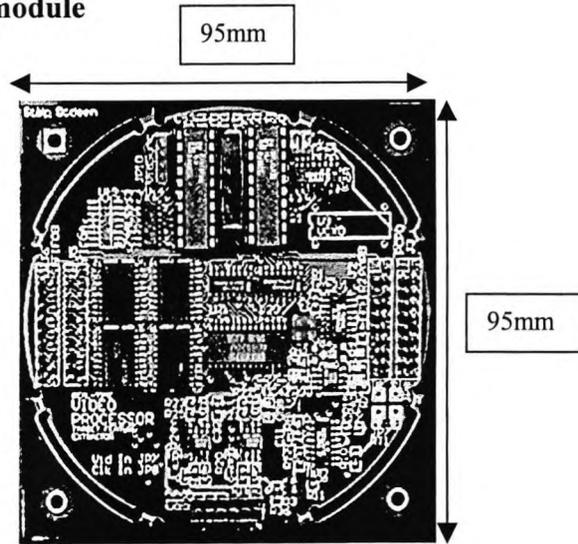
Figure 6.3 Image of the Video Processing Module

The video processing module (Figure 6.3) performs the extraction of object edge pixel locations and the *A/D* conversion of intensity values (Gooch *et al*, 1996). The extracted data are transferred to the *FIFO* buffers in the Input/Output module. Figure 6.4 illustrates a block diagram of this module.
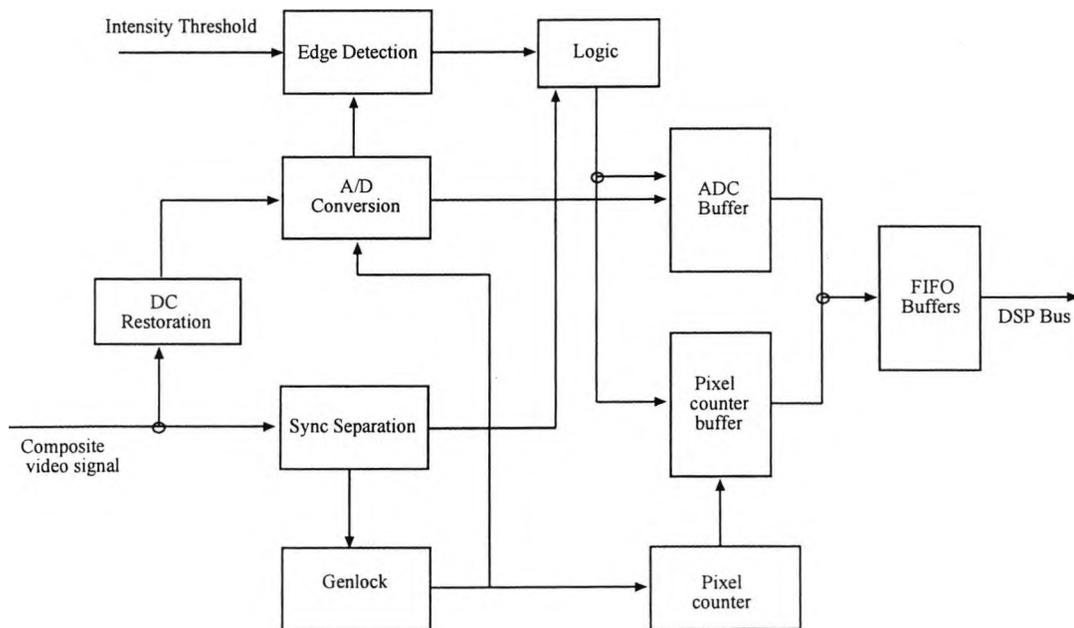
Figure 6.4 Block diagram of the video processing module

Video signal:

The intelligent camera system is based on cameras which comply with the *CCIR* television standard. Table 6.1 illustrates the specification of a Pulnix *TM-6CN* camera.

The *CCD* sensor of the camera converts the light energy received by the sensing elements (pixels) into electrical energy in the form of electrical charge. On board electronics output the collected charge belonging to each array of pixels at a time in the form of an electrical signal. In addition to the intensity information, signal synchronisation information are also encoded in the signal which is termed the Composite Video Signal (Figure 6.5).

Parameter	Specification
Imager and transfer mode	1/2 inch interline transfer *CCD*
Pixels	752(H) x 582(V)
Cell size	8.6(H) x 8.3(V) *μm*
Sensing area	6.41(H) x 4.89(V)
Dynamic range	*67dB*
Scanning	625 lines, 2:1 interlace (*CCIR*)
Clock	28.375*MHz*
Pixel clock	14.1875*MHz*
Horizontal frequency	15.725*KHz*
Vertical frequency	50*Hz*
Video output	1.0V pk-pk composite video, 75Ω o/p impedance
S/N ratio	50*dB*
Minimum illumination	10.0 *lux*
Shutter speed	40*ms*
Power requirements	12*V*, 2.5*W*

Table 6.1 Specification of Pulnix TM-6CN camera

A video frame consisting of *625* lines is output every *40ms*. Hence, the line period is *64μs* or line frequency is *15.625kHz*. Each line of video is separated by a line blanking period which is *12μs*. Line blanking period consists of a front porch (*1.5μs*), a back porch, and a horizontal or line synchronisation pulse (*4.7μs*) which is used to identify a new line. The falling edge of a horizontal synchronisation pulse indicates the start of a new line. The *0V* and *0.7V* voltages indicate black and peak white levels.
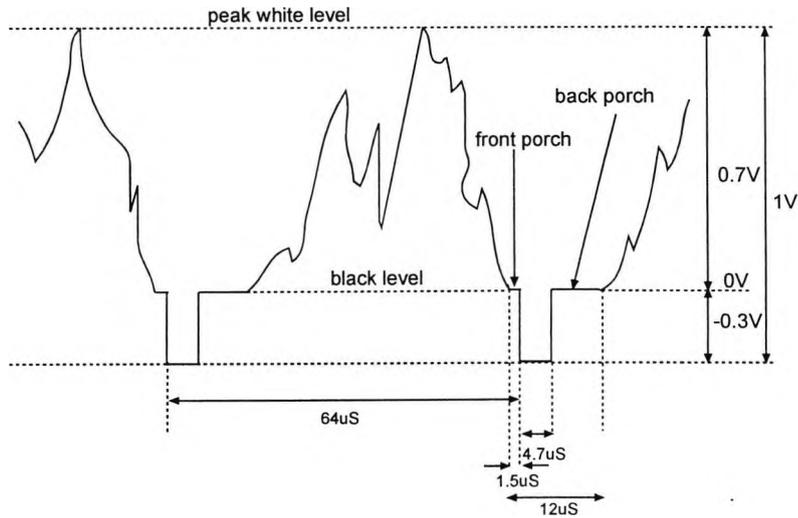
Figure 6.5 Composite video signal (*CCIR*)

A video frame is divided into two fields, namely odd field and even field each of which consists of *312.5* lines of active video and each takes *20ms*. Figure 6.6 illustrates field synchronisation timing for both fields. The odd field begins with a complete line and ends with a half line and it is opposite for the even field. Each field is separated by a field blanking period which is *25* lines of video in length. Field synchronisation pulse sequence is *7.5* lines of video which consists of five pre-equalising pulses (*2.3μs* each), five broad synchronisation pulses or serration pulses (*27μs* each), and five post equalising pulses (*2.3μs* each).
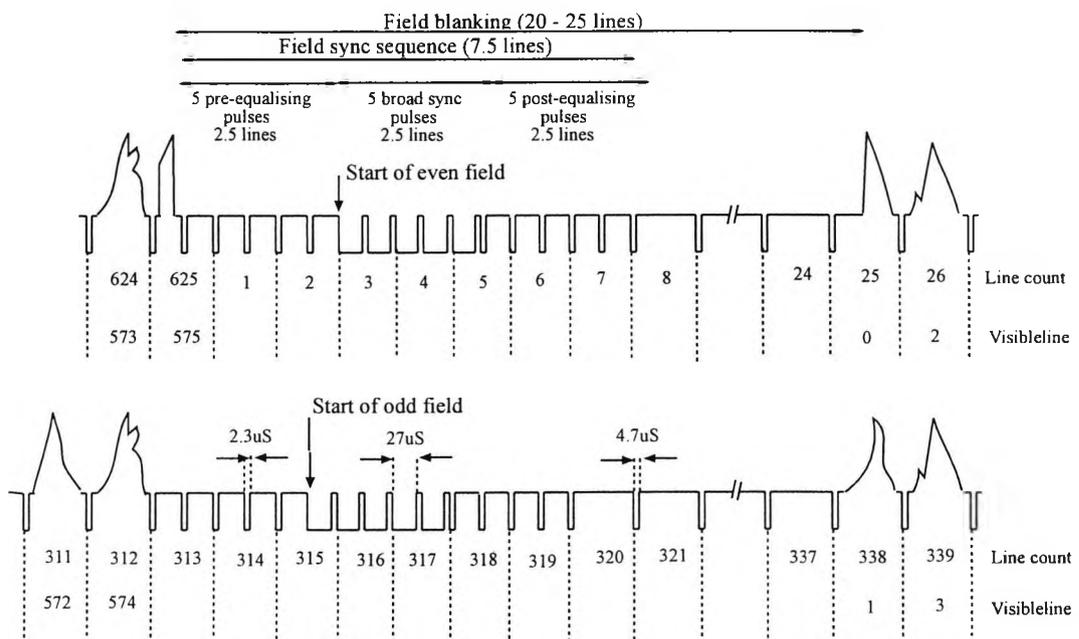
Figure 6.6 Odd and even field synchronisation (*CCIR*)

DC restoration:

The first task of the video module is the *DC* restoration of the video signal. The *DC* restoration process should ideally ensure that the back porch is at *0V DC* level (Figure 6.7) or at another reference level. However, there may be a *DC* restoration error in practice which causes back porch to shift. The conditioned signal is input to the *A/D* converter.

Video signal before DC restoration

Video signal after DC restoration
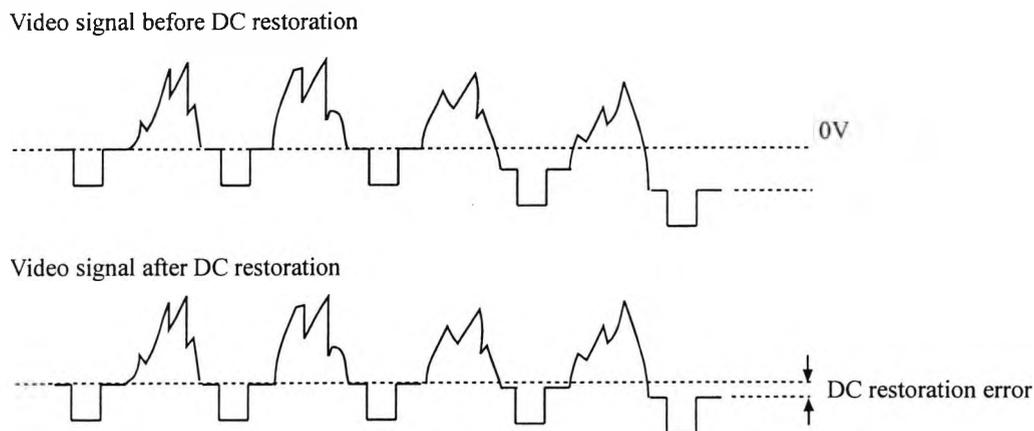
Figure 6.7 DC restoration

Sync separation:

The composite raw video signal is input to the video sync separator for separation of synchronisation pulses which are subsequently used for controlling the operation of the module. The separated signals include horizontal line synchronisation signal, odd/even field output, vertical synchronisation pulse signal, and no signal or black level detection output.

Pixel clock recovery:

The pixel clock is recovered by a genlock chip which is driven by a voltage controlled oscillator operating at a free-running frequency of *14.1875MHz*. The separated horizontal line synchronisation signal is used as input to the phase locked loop which outputs recovered pixel clock and the phase locked/non-locked status signal.

An issue related with pixel clock recovery is line jitter. When the input horizontal synchronisation signal is not in phase with the that generated by the feedback of the phase-locked-loop there will be a displacement in sampling which is called the line

jitter. The typical variation in line-to-line synchronisation is in the order of *2-10ns*. The field-to-field synchronisation is normally larger (Beyer, 1992).

A/D conversion of intensities:

The *DC* restored video signal is fed into the *10*-bit *A/D* converter via a reference voltage level setting circuit. The operation of the *A/D* converter is controlled by the pixel clock and the each converted intensity value is temporarily stored in a buffer which gets overwritten by the next converted value.

Pixel counters and intensity buffering:

In addition, the video processing module has a pixel counter which increments count at rising edge of each pixel clock. It resets at the beginning of each new line and increments the count until end of a video line is detected. Each counter value is also temporarily stored in a buffer which gets overwritten by the next counter value.

A threshold is used to distinguish background noise from the useful object intensity information. When an *A/D* converted intensity is greater than the set threshold, the current pixel counter value and the intensity value followed by a number of subsequent intensity values are transferred to the appropriate *FIFO* buffer in the Input/Output module. These transfers are continued until the next converted intensity value is less than the intensity threshold. The bottom part of Figure 6.2 illustrates an example of pixel location and intensity data that would be produced for the section of the image shown. At the end of each video line, the maximum pixel count is output to the *FIFO* which is used as end of line marker.

Video data are output to the *FIFO* buffers in a special *16*-bit format (Figure 6.8). The first *12* bits of the word (b_0-b_{11}) are allocated for pixel location or intensity data which can be distinguished by the bit in b_{15} where *'1'* represents pixel count and *'0'* represents intensity data. In the case of intensity data only *10* bits are valid where as for pixel count all *12* bits are valid. New field is represented by *'0'* in bit b_{12}. Value in b_{13} represent status of genlock where *'0'* represent locked state. Value in b_{14} denotes whether the video signal is present.

pixel count or intensity

b15	b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	b0
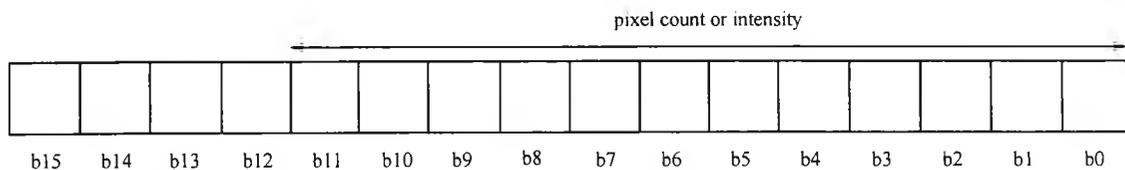
intensity or pixel count data: b_0 - b_{11};

field status: b_{12}: '0' - new field detected;

genlock status: b_{13}: '0' - locked;

no signal: b_{14}: '1' - signal present;

data type: b_{15}: '0' - intensity data, '1' - pixel count data.

Figure 6.8 Data format

The separated synchronisation signals and the pixel clock are used by the on-board programmable logic devices to derive the necessary control signals to drive the data transfers into the appropriate *FIFO* buffer.

6.3.2 Input and Output module

95mm

95mm
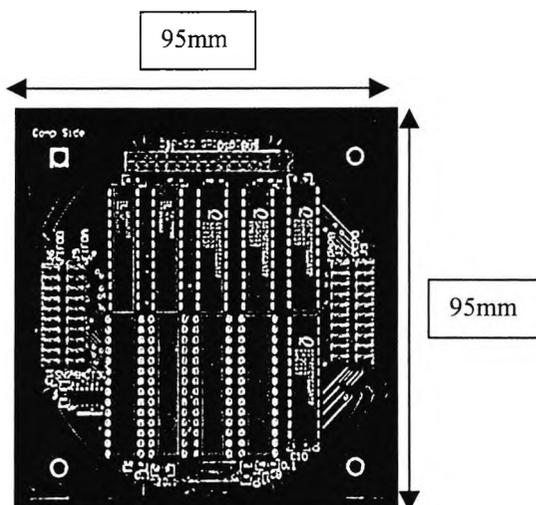
Figure 6.9: Image of Input/Output module

The Input/Output module (Figure 6.9) function as a data buffer for encoded video data. It also function as a buffer for control data (Gooch *et al*, 1996). The module consists of two pairs of *8x16kbits FIFO* buffers and four *8*-bit bi-directional ports. *FIFO* buffers are used to store 16-bit words encoded with pixel count or intensities, and other control information. The first pair is used for buffering the data from odd

field and the second is used to buffer data from even field. The status of the *FIFO* capacity at any instance is indicated by *full* and *empty* flags. One of the *8-bit* ports is used for holding the intensity threshold value which is set by the *DSP* module (section 6.4.3).

6.3.3 Digital Signal Processor module

95mm

95mm
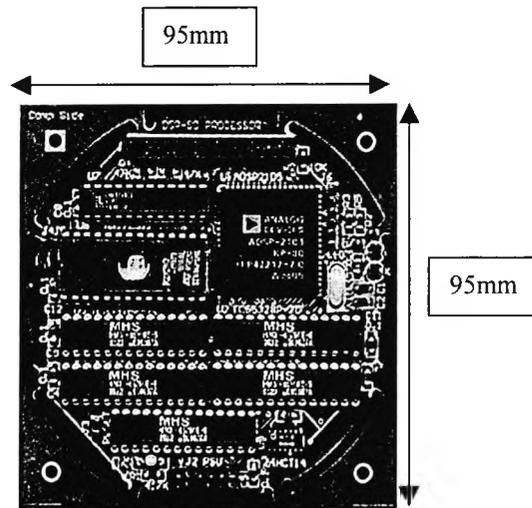
Figure 6.10: Image of *DSP* module

The *DSP* module (Figure 6.10) is the central processing and controlling unit of an intelligent camera system (Gooch *et al*, 1996). It performs three key functions. First, it reads encoded *16*-bit words from the *FIFO* buffers. Second, it performs the target image reconstruction, recognition, and computation of the gray scale centroid. Third, it controls the Ethernet communication module which transfers computed centroid data to and receive commands from the *CPC*.

The Analog Devices *ADSP2101* Digital Signal Processor (*DSP*) which has a modified Harvard architecture was used. The *DSP* architecture is optimised for computationally intensive tasks and has an on-chip arithmetic and logic, multiplier, and barrel shifter units. The on-chip multiple bus system is multiplexed off-chip. The instruction execution time is one clock cycle (*50nS*) except for the division instruction. It has *16x16kbits*, *24x16kbits*, and *24x16*k*bits* of data, program, and boot memory spaces. The data memory space is occupied by internal registers (*1x16kbits*), on-chip *SRAM* (*1x16kbits*), external *SRAM*, and the rest is used for memory mapping other peripherals of the system. Boot memory space is occupied by *EPROM* memory where

the program are stored. Upon the power up, *DSP* loads program memory with the specified contents of *EPROM* and begins execution.

The operation of this module is independent from the video processing module. The synchronisation between the two is achieved using *FIFO full* and *empty* flags. Upon the power up, the *DSP* initialises *FIFO* buffers for a new video frame. It then reads data from *FIFO* buffers and performs the object processing task. If the *FIFO full* status is detected, the *DSP* performs *FIFO* initialisation for a new video frame. If the *empty* status is detected *DSP* waits until data arrives in *FIFOs*. The object processing (section 6.2) is performed and the target locations and the attributes are stored in a transmit ring buffer (section 6.4.3) which has sufficient space for storing data belonging to two video frames assuming a maximum of 500 targets images in a video frame.

On the Ethernet communication side, when the *DSP* receives a command requesting data, the latest data from the ring buffer are transferred to the Ethernet module and is instructed to transmit. The *DSP* is interrupted when an Ethernet packet has received by the Ethernet module upon which *DSP* services the interrupt and reads received data into *DSP* memory. Currently, the *DSP* receives commands from *CPC* for setting the intensity threshold and for requesting target location data. Each receiving command has a unique encoding.

6.3.4 Ethernet communication module
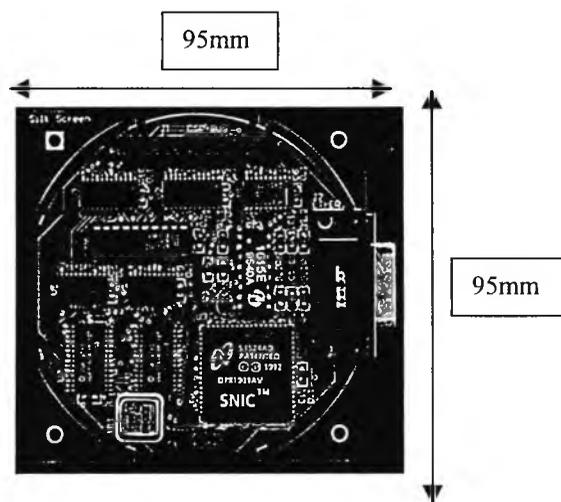
Figure 6.11: Ethernet module

The Ethernet communication module (Figure 6.11) enables the intelligent camera to communicate with the *CPC* (Pushpakumara *et al*, 1996). It is implemented using a National Semiconductors *DP83901AV* Network Interface Controller (*NIC*) device which is *NE2000* compatible and complies with *IEEE802.3* Ethernet standard. An external transceiver unit is required for the Ethernet module to make a physical connection with Ethernet co-axial cable.

In this module, *I/O* port based architecture was used where data transfers between the *DSP* and the Ethernet modules is via a bi-directional *I/O* port. The transfer of control information between the two modules is via a bi-directional latch. The *NIC* has *32k* words of address space for the local packet buffer which is occupied by *16x8*kbits of *SRAM*. The packet buffer is divided into transmit and receive sections. The size and the boundaries of which can be specified in the control registers.
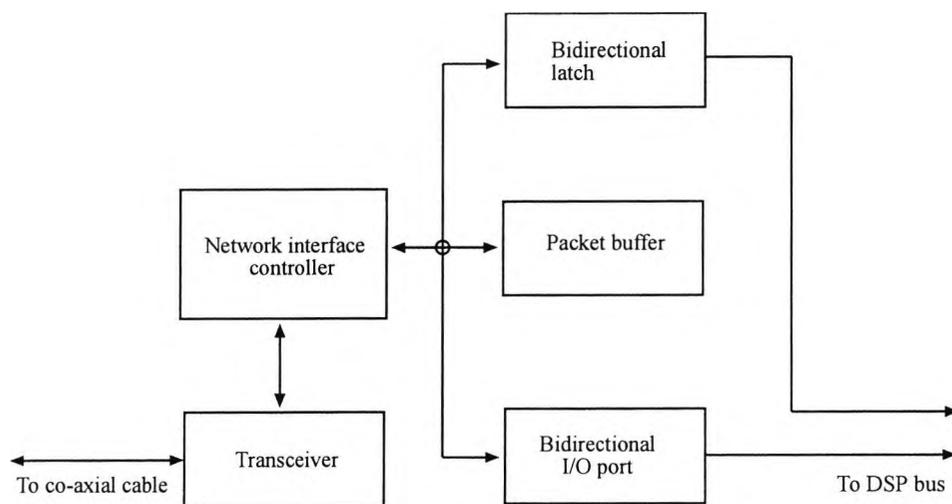
Figure 6.12 Block diagram of Ethernet module

The *NIC* has *3* pages of on-chip *8*-bit registers which are mapped into *16x16bits* memory locations in the *DSP* memory map. The *DSP* selects the appropriate register and the *8*-bit bi-directional latch provides a means of reading from and writing to these registers. These registers must be initialised with appropriate values prior to any communication operations.

The *NIC* has a *16*-byte *FIFO* memory for buffering data transfers between the communication media and the packet buffer. A programmable threshold facilitates

filling and emptying the *FIFOs* at different rates. A dedicated on-chip Direct Memory Access (*DMA*) controller provides local *DMA* and remote *DMA* facilities. Local *DMA* facilitates data transfers between *FIFO* and packet buffer. Remote *DMA* facilitates data transfers between packet buffer and the *DSP* memory. The former has priority over the latter. On-chip Manchester encoding/decoding and *CRC* generation/checking functions are also provided by the *NIC*. The multiplexed address/data bus of the *NIC* is available externally for interfacing.

6.3.5 Power supply module
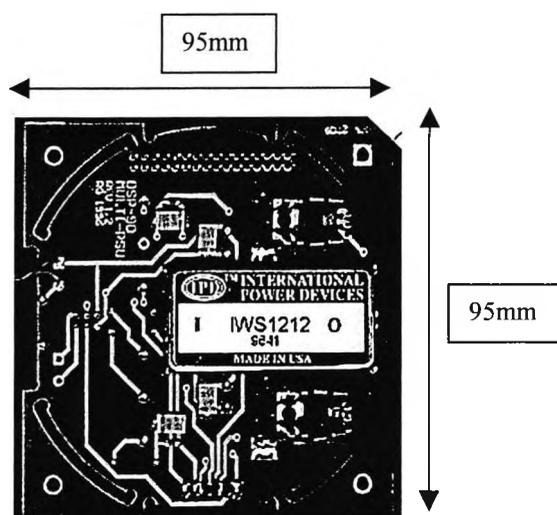
Figure 6.13 Image of Power Supply Module

The Power Supply Module (Figure 6.13) provides power to the complete intelligent camera system (Gooch *et al.*, 1996). This module has three *DC-DC* converters to enable a wide band input (*9-18V* or *18-36V*) to produce the necessary output voltages.

6.4 Software development

The software for the intelligent camera system was developed using structured methodologies. The software comprises of three parts,

> 1) Video data acquisition
> 2) Object processing
> 3) Ethernet communications.

Each part consists of a number of sub modules. One of which is the initialisation sub module. The video initialisation part may require a significant amount of work reading

and discarding *FIFO* data for the alignment of a new video frame. The initialisation sub modules in object processing and the Ethernet communication parts require setting certain registers and variables to specific values. Each part is initialised upon the power up. Video acquisition may need re-initialisation after the startup depending on the occurrence of *FIFO* buffer overflow condition.

Each part of software performs a time critical task. A complete video frame worth of object edge location, intensity, and control information are continuously output by the video processing module into the *FIFO* buffers during each *40ms* period. The video data acquisition and image processing parts work together. The data belonging to a line are acquired from appropriate *FIFOs* (*i.e.* odd or even) and object processing is performed. This is repeated until a frame is completed and the processed data are stored in the transmit ring buffer waiting to be transferred to the Ethernet module. Once a frame is completed, processed data are transferred to the Ethernet communication part to be transferred to *CPC*. The amount of useful data in a frame depends on the number and the size of the objects. Hence, the time required is data dependent.

6.4.1 Video data acquisition

There are two tasks for this part of the software.

> 1) Initialisation of the *FIFO* buffers for alignment of a new video frame
> 2) Reading the pixel count and intensity data from *FIFOs*

Upon power up, the *FIFOs* are hardware reset and the encoded data are input by the video processing module. The incoming data to the *FIFOs* are most likely to belong to one of the four states which can be determined by checking *full* and *empty* flags.

> 1) Odd *FIFOs* filling and no data in even *FIFOs*,
> 2) Even *FIFOs* filling and no data in odd *FIFOs*.
> 3) Odd *FIFOs* are filling and some data is already in even *FIFOs*
> 4) Even *FIFOs* are filling and some data is already in odd *FIFOs*

Hence, an initialisation of the *FIFOs* is required for the alignment with a new video frame. A new frame begins with an odd field. If the *FIFOs* are in the first state, data should be read from odd *FIFOs* until empty and then from even *FIFOs* until empty. At this point the data that begin to flow into the odd *FIFOs* belong to a new frame. If the *FIFOs* are in the second state, data should be read from even *FIFOs* until empty. At this point the data that begin to flow into the odd *FIFOs* belong to a new frame. If in either the third or fourth state, which is indicated by both odd and even *FIFOs* not being empty, the odd *FIFOs* are read until empty and then even *FIFOs* are read until empty. The new data that appears in odd *FIFOs* belong to a new frame. Also if the *FIFOs* are *full*, hardware resetting is required.

Once initialised, the system can operate in either field or frame modes. In the field mode, data from only one field is used for object processing hence, resolution is halved in the *y* direction. The *DSP* reads and stores data from the selected *FIFOs* in *data_register1* or *data_register2* (Figure 6.8). Once data belonging to a complete line has been read, the object processing is performed. The data reading and processing is continued until the end of field.

In the frame mode, the data from both fields are used. The *DSP* reads and stores data of a complete line from odd *FIFOs* first and then from even *FIFOs*. The object processing is then performed. This process is continued until the end of the frame. In either mode of operation, a new field is detected by the new field output of sync separator. The trailing edge of the new field pulse from sync separator coincides with the first serration pulse during the vertical drive. Hence, the first end-of-line after the detection of new field is the end of third line of the new field (Figure 6.14). From this point onwards, *22* video lines should be read and discarded prior to the beginning of the first active line of the field. The end-of-line is indicated by the maximum pixel count value which is output to *FIFOs* at the end of each line.

It was determined during the development stage (Gooch *et al.*, 1996) that a complete video line, that is from a trailing edge of a horizontal sync pulse to a same moment of the next sync pulse, consists of *907* pixels. A line period is *64μs* and the active line period is *52μs*. Hence, there is an offset of *10.5μs* from the beginning of the line until

the active period begins which is equivalent to about *152* pixels. Therefore, *152* pixels should be deducted from every counter value to get the correct active pixel count.

6.4.2 Object processing

Object processing software performs reconstruction, recognition, and centroid location. The details of the algorithm was discussed in section 6.2. For the progressive reconstruction, the object segments belonging to two consecutive lines are required. The implementation is discussed in this section. It has special data structures which include a number of buffers and a stack (Figure 6.14).
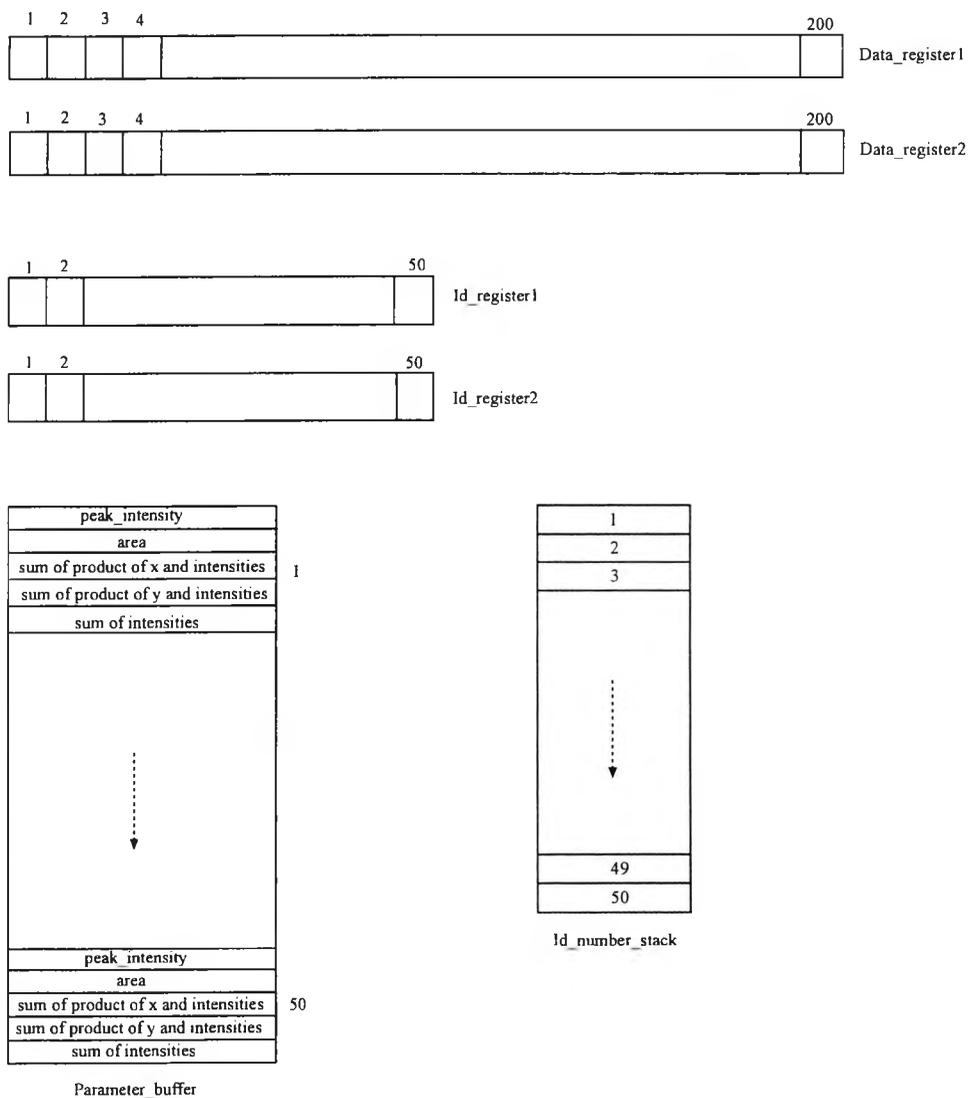

Figure 6.14 Data structures for object processing

A simple initialisation process is required which sets buffers and the stack with appropriate values. For each new video frame these buffers and the stack require

initialisation. The *data_register1* and *data_register2* are one dimensional arrays of *200 16*-bit words each. An end-of-line marker (*i.e.* maximum pixel count) is written to the first location of *data_register1* to enable the identification of object segments belonging to the first line of a video frame as new objects. The *id_number_stack*, an array of fifty *16*-bit words, is used for storing object identity numbers. It should be initialised with a sequence of numbers (*i.e. 1....50*). Hence, fifty objects can be dealt with at a time.

The buffer pair *data_register1* and *data_register2* are used to store pixel count and intensity information of object segments in current and previous lines. When a new object segment is encountered a unique identity number is popped from the *id_number_stack*. A pair of buffers *id_register1* and *id_register2* are used for storing the identity numbers of the object segments in current and previous lines. The *parameter_buffer* is used to store parameters of objects being reconstructed. When a new object is encountered five consecutive memory locations within the buffer are assigned. The actual locations within the buffer is dependant upon the object identity number. The two attributes area and peak intensity, the sum of the product of pixel locations and intensities, and the sum of intensities are stored in each location. These parameters are updated until the object is completely reconstructed (finished). Parameters for objects that cannot possibly be targets (objects that are large or of irregular shape) are not updated. However, each object encountered is completely traced to retain the integrity of the recognition process but any parameters pertaining to these objects are discarded when the object is completely reconstructed. When a completely reconstructed object passes the criteria for selection as a target the identity stack is pushed and grey-scale centroid is calculated, the area and peak intensity are stored in transmit ring buffer (section 6.4.3) waiting to be transferred to Ethernet module. The locations in the *parameter_buffer* are freed and are made available for a new object.

6.4.3 Ethernet communications

The initial version of the Ethernet software for the *CPC* was developed for *NE2000* compatible hardware. The *CPC* has off-the-shelf Ethernet hardware installed and intelligent camera system has dedicated Ethernet hardware (section 6.3.4) both of which are *NE2000* compatible. Hence, apart from the language of coding, the Ethernet

software at either end is essentially the same. The *DSP* and the *CPC* versions were coded in *ADSP2101* assembly and '*C*' languages. The software comprises,

1) a packet driver

2) a simple protocol

Packet drivers:

The packet driver communicates with the hardware. Each packet driver has a hardware initialisation routine, a transmit routine, and a receive interrupt service routine. The hardware initialisation routine configures the *NIC* to operate in the current system which is performed at the system start up. The three pages of on-chip registers are initialised.

The transmit routine performs the data transmission to the *CPC*. The location, area, and peak intensity for target images are stored in a transmit ring buffer as objects are processed (Figure 6.15). The transmit ring buffer has pointers to start, end, current write, and current read locations. As data are read from, or written to, the buffer, the current write and read pointers are updated. The first seven words of the buffer are used to store destination address (*DA0-DA5*), source address (*SA0-SA5*), and byte count (*BC0-BC1*). During the initialisation source address is written to *SA0-SA5*. When it is required to transmit, *DA0-DA5* are filled with appropriate destination address. The total data byte count (*TBC*) belonging to each frame are stored in *TBC* which is read and written to *BC0* and *BC1*. The *DSP* transfers the header and data to the packet buffer by Remote *DMA* transfer. Then instructions are given to the *NIC* to transmit the packet to the destination. If a packet is larger than *1500* bytes, the destination will know it by the difference of values in *TBC* and *BC0/BC1* and will wait for the rest of the data to arrive which is looked after by the protocol. Similarly, the protocol continues transmission until complete frame of data are transmitted.

The receive interrupt service routine performs packet reception. The highest priority interrupt *(IRQ2)* of the *DSP* is used to activate the receive routine. An incoming packet whose destination address is the same as that of the local Ethernet module is detected by address recognition logic of the *NIC*, it Local *DMA* transfers the complete packet to the receive section of the local packet buffer. Then *NIC* informs of the

packet reception to the *DSP* by means of an *IRQ2* interrupt. The service routine then Remote *DMA* transfers the packet to the receive buffer (Figure 6.15) in *DSP* memory. The receive buffer keeps the source address, byte count, and data of incoming packets. Currently, the data sent to the *DSP* are intensity threshold and command messages. The most significant bit of the *TBC* field is used to make the distinction between the data and the commands (*0* - data, *1*- command).

On the *CPC* side, a packet driver of the *CPC* version was present and the operation was the same. The same protocol controls the data and command transfers and the reception of data.

This Ethernet software was tested for data transfers between *PCs*. Four *PCs* installed with Ethernet software successfully simulated the data transfers between a *CPC* and three intelligent camera systems. Finally, the real-time data transfer between a *PC* and an intelligent camera system was demonstrated.

| Transmit buffer		Frame 1	Receive buffer	
Flag, Total byte count				
Data			Message	
Flag, Total byte count			SA5	SA4
Data			SA3	SA2
Flag, Total byte count			SA1	SA0
BC1	BC0		Message	
SA5	SA4		SA5	SA4
SA3	SA2		SA3	SA2
SA1	SA0	Header information	SA1	SA0
DA5	DA4			
DA3	DA2			
DA1	DA0			
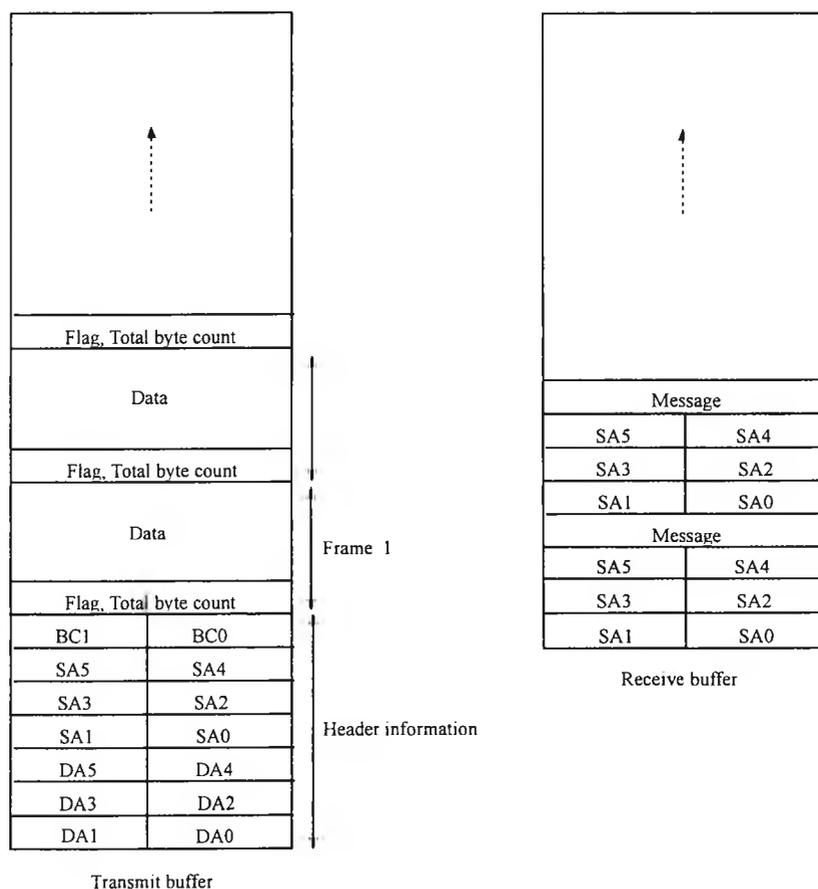
Figure 6.15 Data structures for Ethernet communications

The basis of the above mentioned software paved the way for further developments. Clarke *et al*, 1998 further developed the video data acquisition and Ethernet software. The field mode video data acquisition was implemented which facilitated low accuracy *2-D* but better speed performance. The packet driver on the *PC* side was replaced by Windows sockets based Internet Protocol (*IP*) Datagram software. The *DSP* packet driver was modified accordingly. A simple message passing scheme was implemented for data and command transfers. In real-time tests where varying number of targets were processed, the Ethernet took some *5%* of the total processing time. A four intelligent camera and a *PC* based system was implemented where each intelligent camera was able to produce *170 2-D* target locations at frame rate (*i.e.* every *40ms*).

6.5 Summary of the chapter

This chapter described the development of a real-time close-range photogrammetric system. The system exploits the concept of distributed processing of video data to gain the speed performance over frame-grabber based systems in which feature extraction, correspondences, and *3-D* computations are performed in a central computing unit. In this system the feature extraction is performed in parallel by each intelligent camera system. This saves a considerable amount of processing time. The hardware edge detection and the object processing algorithm are central to the speed performances of feature extraction which requires only a pair of consecutive lines of video to reconstruct the object images in a progressive manner. Also the amount of data transferred between the cameras and the central computing system are much lower with this system due to the transfer of only processed data. For each recognised target its location, area, and peak intensity are output. If there are *100* targets in each image, only *800bytes* will be transferred by each camera. A single Ethernet packet is sufficient for this purpose. In the case of a frame grabber system, transfer of a complete image (approximately *440kbytes*) will require about *294* Ethernet packets of *1500bytes*. Hence, intelligent camera solution saves time in transferring data.

Chapter 7

Experiments with real image data

The performance of the real-time system comprising intelligent cameras and *PEL* correspondence establishing technique was assessed using a series of tests. Initially, lens systems were calibrated and target image location accuracy performance of the intelligent cameras were analysed. Tests on the establishment of real-time correspondences were carried out with moving objects. Further tests on *PEL* technique were carried out with frame grabber acquired data which covered situations such as occlusions and the establishment of correspondences using the locations of radiometrically distorted target images. Some of these tests and results are illustrated here. in this chapter.

7.1 Real-time system set-up

A network comprising four intelligent cameras was set up for the purpose of experiments. Pulnix *TM-6CN* monochrome interline transfer *CCD* cameras with *16mm* C-mount lenses were used. Each intelligent camera was connected to a *PC* (central processing computer-*CPC*) via an Ethernet link. The *2-D* target location data were transferred to the *CPC* where multiple view point correspondences and least squares adjustment functions were carried out.

7.1.1 Setting up illumination

Appropriate illumination of the object(s) being imaged is an integral part of a vision system. In high precision measurement applications, it is necessary to use special lighting and targeted object points. In order to gain maximum light reflection from retro-reflective targets that are placed on an object, *LED*-ring illumination was set up which illuminates the targeted object close to the optical axis of the camera. Due to the characteristics of this configuration, a higher percentage of the reflected light is imaged by the sensor. Hence, target images can be located accurately.

7.1.2 Lens system calibration

Chapter 2.0 discussed lens distortion and modelling. The calibration process estimated the following distortion parameters for each lens.

1) Principal point shift xp, yp

2) Correction to the focal length df

3) Radial distortion parameters k_1, k_2, k_3

4) Decentering distortion parameters p_1, p_2

5) Geometric distortion parameters A,B,C

The distortion parameters were estimated using a self-calibrating adjustment program. Images of eight different views of a calibration target field (Figure 7.1) were obtained from each camera by moving the target field into eight different positions. The purpose of having various views by a single camera was to ensure strong network geometry for the self-calibrating adjustment process and also to have a wider coverage of the area of the lens. Using independently identifiable coded targets placed on the target field, the cameras were resected and approximate exterior parameters were obtained. For each camera, the image point correspondences were established in eight views without using the camera interior parameters. Only a *40%* of the targets were corresponded with these coarse exterior parameters. The approximate *3-D* locations of the corresponded targets were computed. A self-calibrating least squares adjustment was carried out which resulted in improved camera exterior parameters, improved *3-D* co-ordinates of target field points, and approximate values for camera interior parameters. Using these improved parameters correspondences were established again which resulted in about *95%* of targets corresponded. The self-calibration was carried out again and further improved values of the camera interiors were obtained. Table 7.1 illustrates camera interior parameters for each camera system.
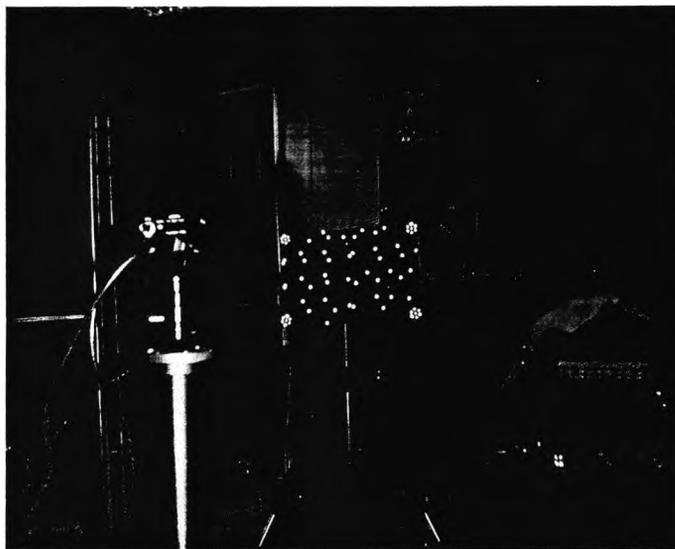

Figure 7.1 Image of test field

Parameter	Camera 1	Camera 2	Camera 3	Camera 4
xp (mm)	-8.8175e-002	1.3359e-001	-1.6191e-001	-1.2940e-001
yp (mm)	2.2715e-001	-2.2270e-001	2.9299e-002	-2.9601e-001
df (mm)	-7.2234e-002	-1.8857e-002	1.7841e-002	1.8749e-001
k1	1.8483e-003	-1.3226e-003	8.1297e-004	-2.9951e-003
k2	-4.6867e-004	4.0711e-004	-1.2063e-004	8.3418e-004
k3	3.3478e-005	-2.6166e-005	8.3108e-006	-6.8741e-005
p1	5.3565e-005	-4.3899e-005	3.2586e-004	6.1739e-005
p2	-3.5832e-004	2.4073e-004	-1.9958e-004	-1.8659e-004
A	2.4964e-004	1.0217e-002	7.7636e-003	8.4946e-003
B	1.1460e-005	-2.1801e-004	1.6234e-003	7.3428e-005
C	-5.6341e-005	-6.7743e-005	-6.4007e-005	-5.9205e-004

Table 7.1 Camera interior parameters

7.2 Correspondences between four views

Four views of the *3-D* calibration target field were obtained (Figure 7.2). Target image point locations were corrected for systematic lens distortion errors using camera interior parameters given in Table 7.1. In each view there are different number of target locations although all the *3-D* targets were in the field of view of each camera due to the occlusions.

view - 1

view - 2

view - 3   view - 4
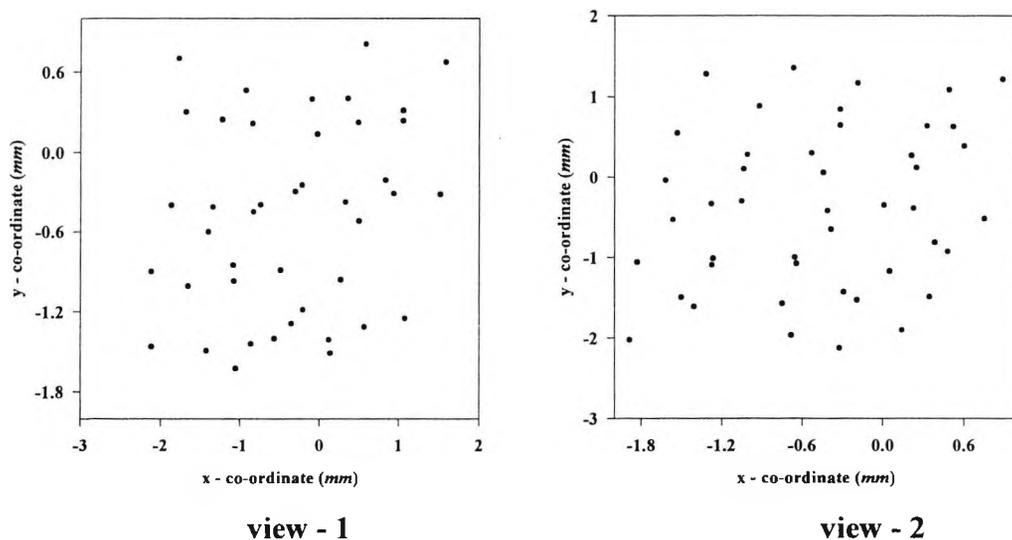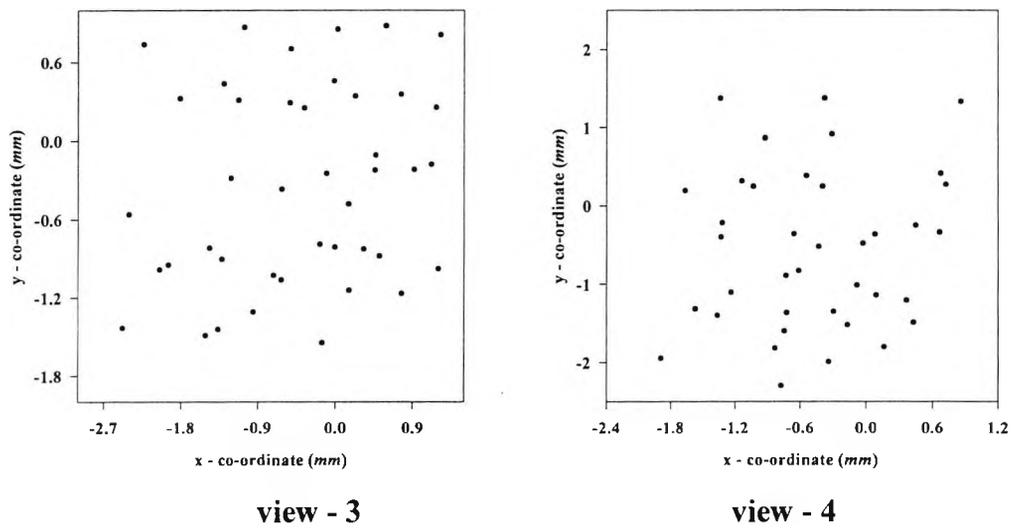
Figure 7.2 Four views of *3-D* calibration target field

Using four coded targets approximate exterior orientation of the views were obtained as discussed in section 7.1.2 and refined camera exterior parameters were obtained iteratively. Table 7.2 illustrates camera exterior parameters and effective focal length of each camera lens. The standard deviation of image residuals was *1.07μm*.

Camera no	X0 (*mm*)	Y0 (*mm*)	Z0 (*mm*)	Omega (*degree*)	Phi (*degree*)	Kappa (*degree*)	Effective focal length (*mm*)
1000	1013.2937	-1133.2219	1898.9145	28.2076	25.9665	178.7695	15.9278
1001	1080.1937	974.2855	1898.0216	-29.4981	24.7967	-77.5299	15.9811
1002	-974.8957	1080.0551	1889.8619	-29.0854	-25.7232	-2.7346	16.0178
1003	-1124.6651	-1014.8325	1898.9858	28.9467	-25.4029	102.0058	16.1875

Table 7.2 Camera exterior parameters

Correspondences were established with improved camera exterior parameters for a total of *42* points. A search band width of *6.42μm* was used. Table 7.3 illustrates some details of the correspondences. Certain targets were corresponded in all four images and some others were corresponded in three images. The correctness of correspondences was checked using the image residuals. For this purpose each set of

corresponded targets, either from three or four views, *3-D* location was computed by intersection. Image residuals were then computed by back-projecting rays into each view from the computed *3-D* location. If the residuals were larger than a pre-determined threshold it was considered that the corresponded set was incorrect. A threshold of *1.95μm* was used.

Correspondences in 4-views	Correspondences in 3-views
42	5

Table 7.3 Details of correspondences

Figure 7.3 illustrates the *3-D* reconstruction of the target field. The standard deviation of estimated *X*,*Y*, and *Z* were *0.2492*, *0.2424*, and *0.3451mm*.
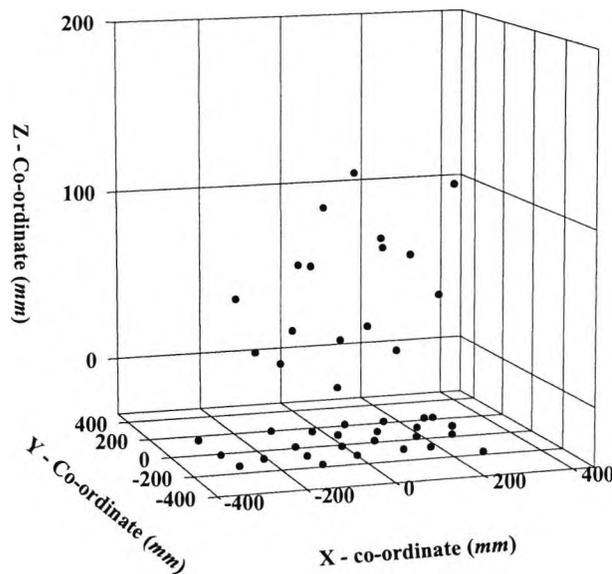
Figure 7.3 *3-D* reconstruction of the calibration target field

7.3 Real-time data flow from a single intelligent camera

This set of tests were carried out to investigate the stability and correctness of the *2-D* co-ordinates from an intelligent camera. Three sets of tests were carried out.

Test 1:

In this test three retro-reflective targets were placed in the field of view of a single intelligent camera. The camera optical axis was approximately perpendicular to the

target surface. One thousand images of these three targets were acquired in real-time. Figure 7.4 illustrates the location variation of a targets over the period of time of this test. Standard deviation of the location variation for this target in the *x* and *y* directions was *0.025* and *0.0217* pixels.
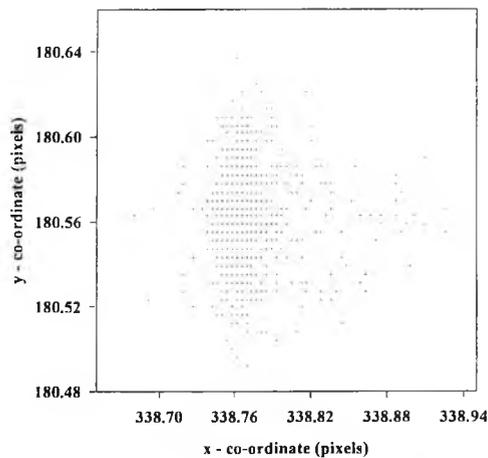
Figure 7.4 Variation of a single target location over *1000* images

Test - 2:

This test involved collection of images of a moving target. The movement of the target was controlled by a stepper motor to ensure smooth motion. One thousand images of this target was acquired over a period. Figure 7.5 (a) and (b) illustrate the location profile and residuals from straight line fit. Area of this target was approximately *24* pixels and the peak intensity was about *747*. The standard deviation of the location residuals was *0.022* pixels.

(a)   (b)
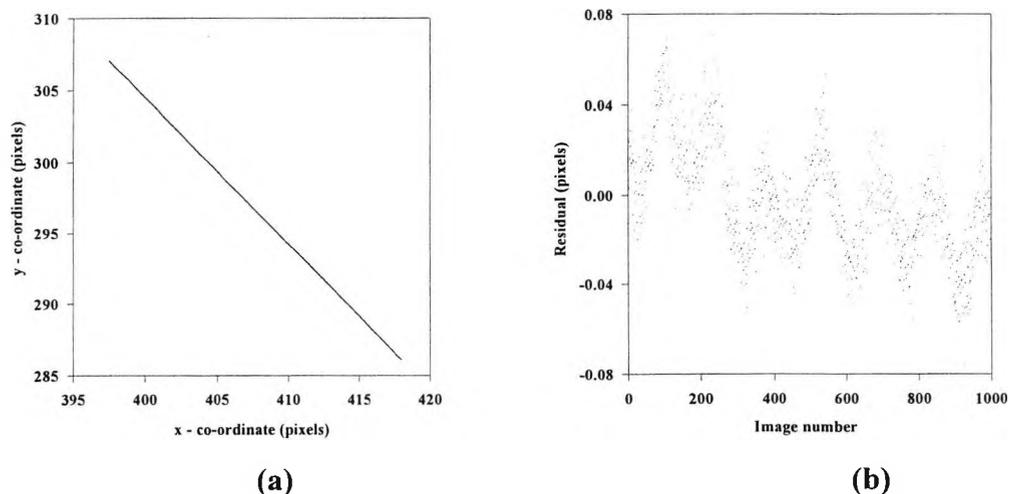
Figure 7.5: (a). Location profile of a moving target over *1000* images (b). image residuals from straight line fit

Test - 3:

In this test three targets were placed on a surface of an experimental apparatus which has three degrees of movement freedom (X, Y, and Z). The targeted surface was moved in a direction that is approximately parallel to the surface of the image plane and over *1000* images were acquired in real-time. Figure 7.6 illustrates the movement profile for each target.
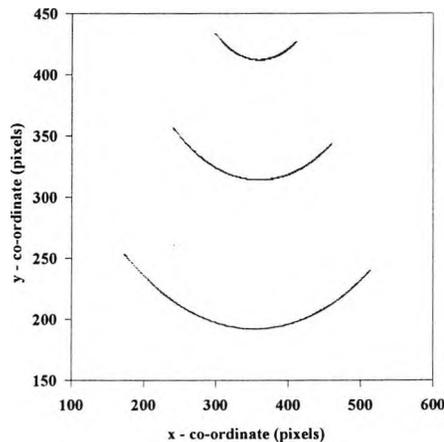
7.6 Location profile of three moving targets

7.4 Correspondences with real-time data flow from *4*-cameras

In industrial assembly processes, it is required to position assembly components together to the correct geometry with tight tolerances. This test attempts to illustrate a simple situation where two objects are involved. The four camera system was resected and the exterior parameters were obtained. A least squares adjustment was carried out and refined camera exterior parameters were found. Each object used was independently identifiable in images by the number and the pattern of the placement of retro-reflective targets. The first and the second objects have *15* and *5* targets on their surfaces. A sequence of *100* sets of images were acquired from four views in real-time while the second object was moved towards the stationary first object. The point correspondences were established and a sequence of *3-D* reconstructions were made. Figure 7.7 illustrates a set of four views from the sequence and the *3-D* reconstruction. Figure 7.8 illustrates the movement of five target images in the subsequent sets of the four views and the sequence of *3-D* reconstructions.
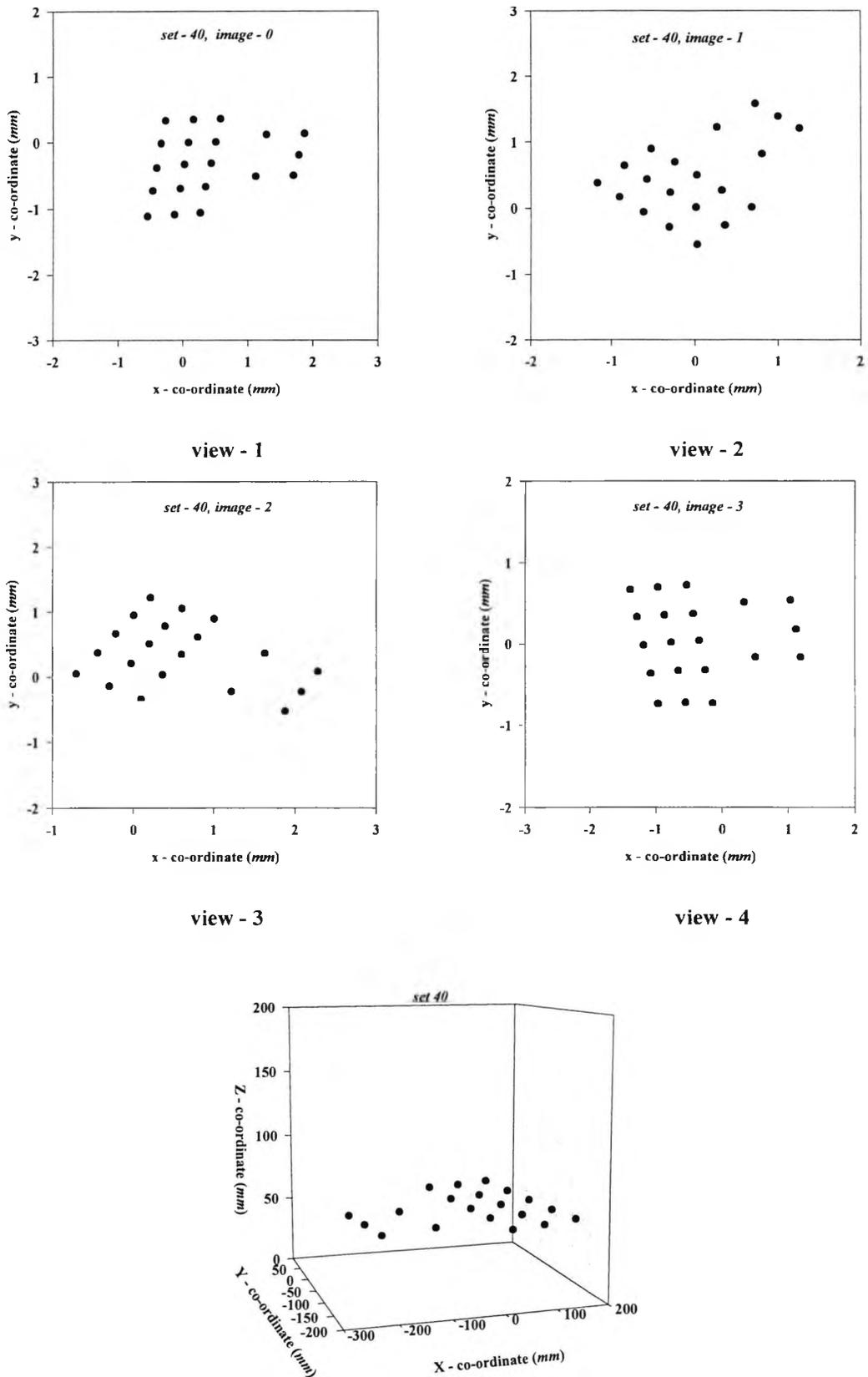
Figure 7.7 Four views of the targeted objects and the *3-D* reconstruction
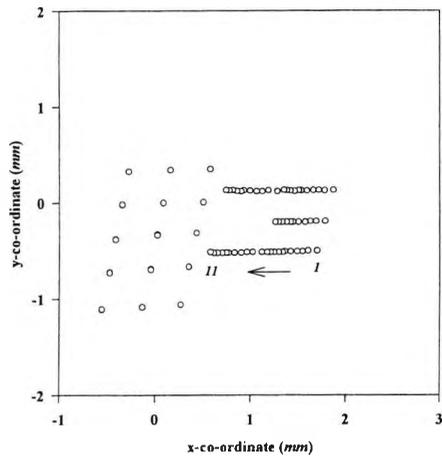
view - 1
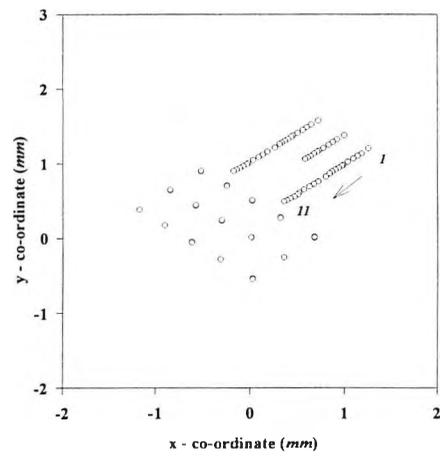
view - 2
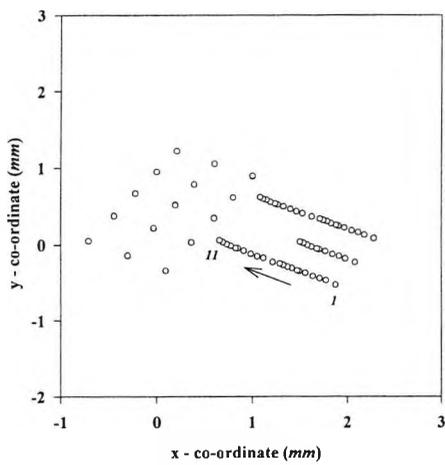
view - 3
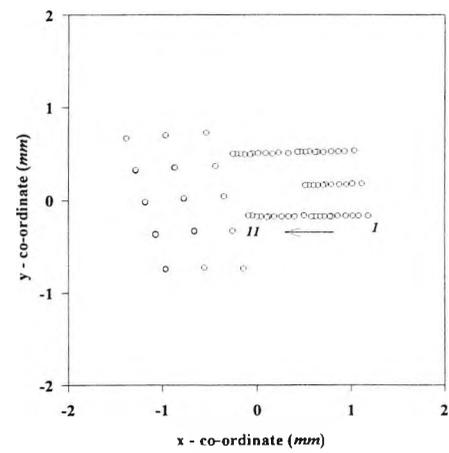
view - 4
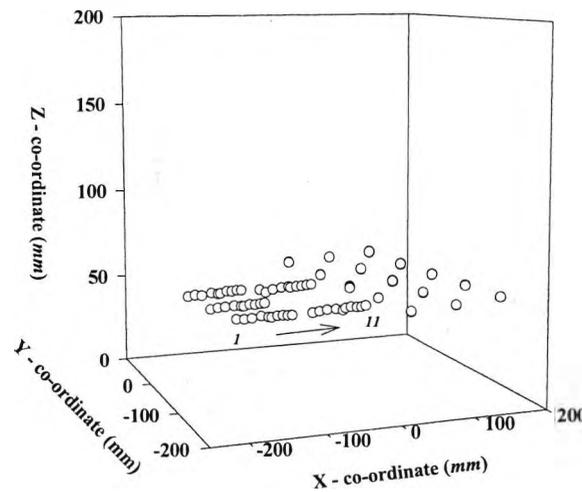
Figure 7.8 Sequence of four views and the *3-D* reconstructions

For this test search band width of *0.01mm* was used. Since there were fewer targets which were well spreaded, a wider band width did not have effect on the

correspondence accuracy or speed. It was noted that in each stereo pair combination, one-to-one correspondences were established. It took *0.9ms* for the intelligent camera/*CPC* data transfer and *7.37ms* to establish correspondences using the *PEL* method.

7.5 Correspondences using *PEL* algorithm with frame grabber acquired image data

Two tests were carried out to investigate the performance of the algorithm using a target field with (a) a dense distribution of targets, (b) low intensity targets

Test - 1:

A target field which consisted of about *500* densely distributed targets were used for this test. Figure 7.9 illustrates four views of the target field.

view - 1   view - 2

view - 3   view - 4
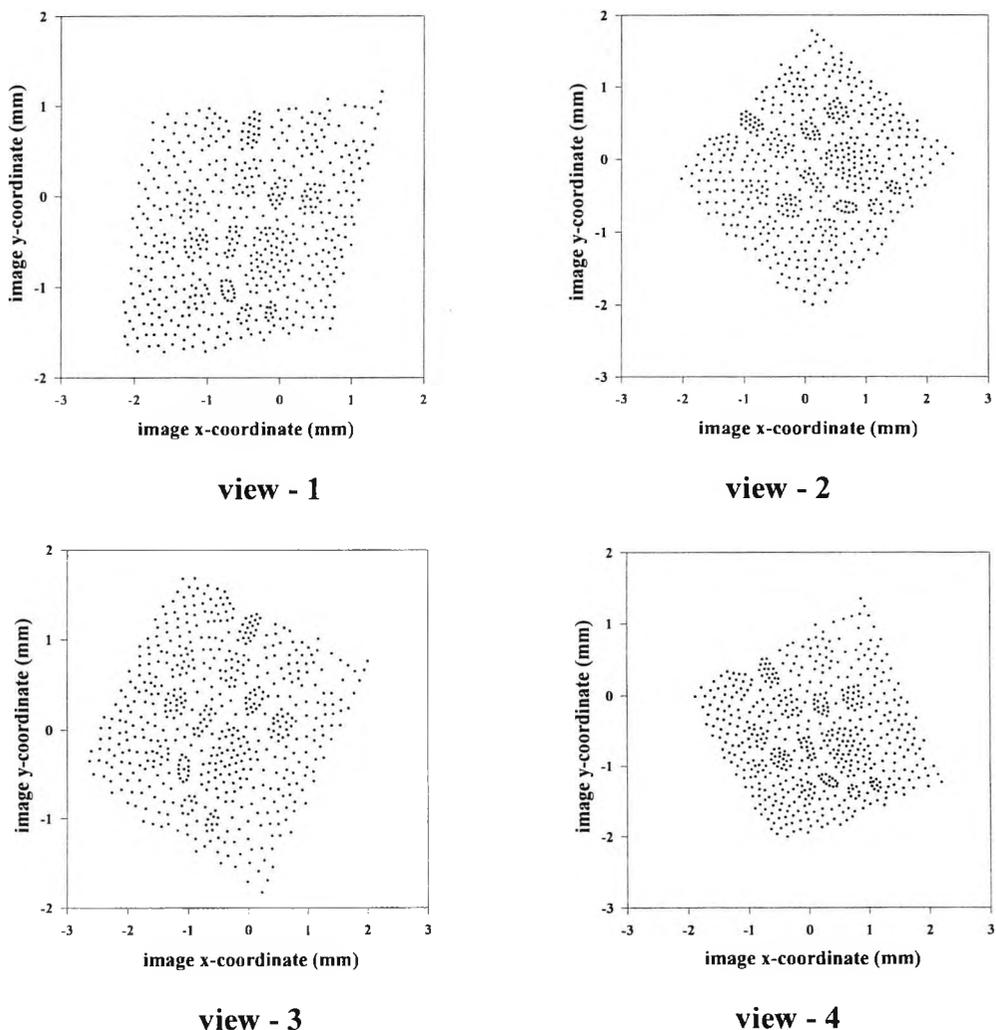
Figure 7.9 Four views of dense target field

Target image locations were corrected for lens distortions. Image data were analysed and it was found that the shortest distance between two points in images was about $68.8\mu m$. The four camera network was adjusted using a calibration target field and accurate camera exterior parameters were available. The standard deviation of the image residuals was $0.27\mu m$ as given by the adjustment. Hence, a search band of $2\mu m$ was used and correspondences were established for all points between four views. The correctness of the correspondences were checked using the intersection/back-projection technique. Figure 7.10 illustrates the *3-D* reconstruction.
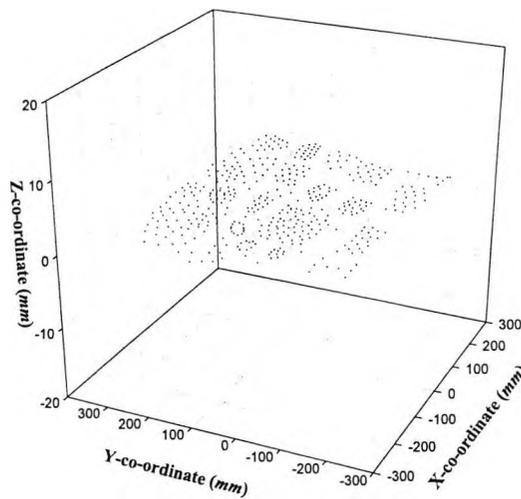
Figure 7.10 *3-D* **reconstruction of target field**

Test - 2:

In this test a perfectly flat target plate was used in which there were about *150* targets uniformly distributed with equal spacing between them. Images from four views were acquired from an adjusted network (Figure 7.11). Hence, accurate camera exteriors were available. Each view had a different number of target images (*143,145,142,136*) as not all the targets on the plate reflected sufficient light to the cameras. Most targets were bright and occupied a larger area but some targets were very dim (peak intensity of *3*) and occupied an area of one, two or three pixels. As a result these targets were located very poorly in the image space which could make it difficult to establish the correspondences. Since the camera exterior parameters were accurate and target image locations were corrected for lens distortion errors, a wider search band width of $0.05\mu m$ was used. Figure 7.12 illustrates the *3-D* reconstruction.

view - 1

view - 2

view - 3

view-4
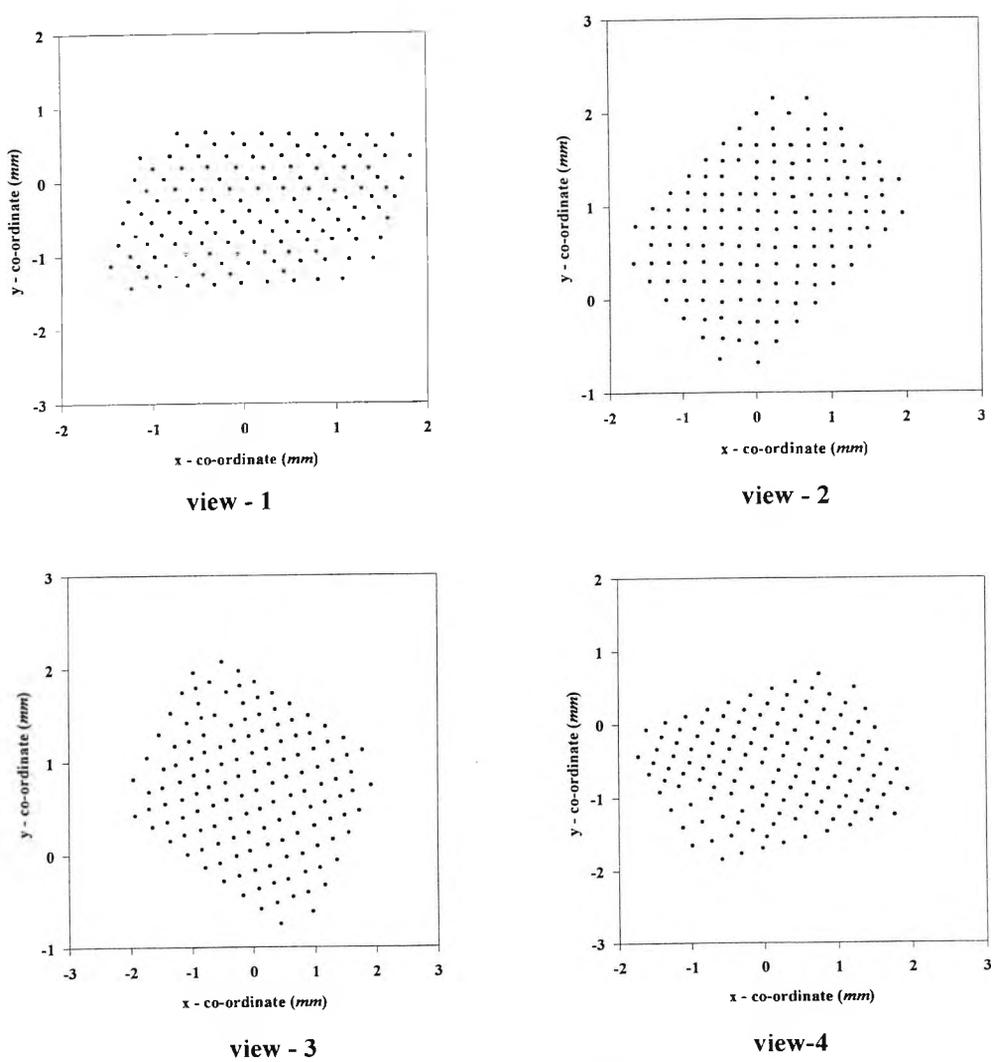
Figure 7.11 Four views of target plate
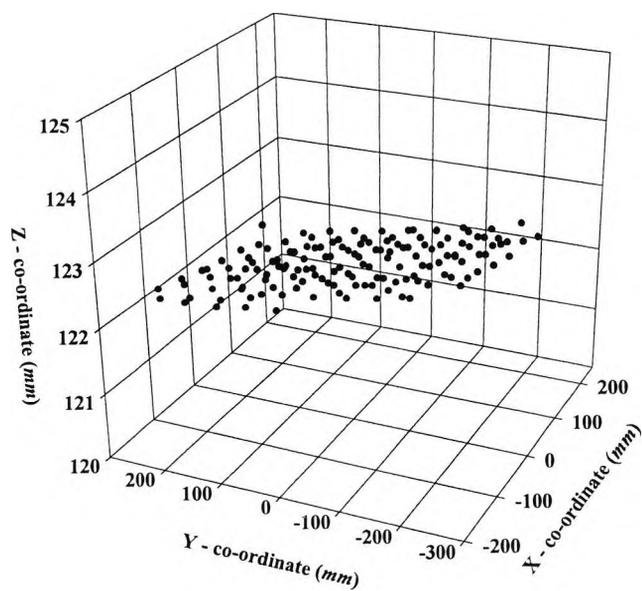
Figure 7.12 *3-D* reconstruction of test field

7.6 Summary of the chapter

In this chapter, the performance of the real-time system comprising intelligent cameras and the *PEL* correspondence establishment technique was evaluated. The real-time system setup was described. The lens calibration process and the estimated lens calibration parameters were illustrated.

The stability and the correctness of the image data acquired from the intelligent cameras were analysed. Using the repeated localisation of an image of a stationary target in real-time, the stability of the target image location in the x and y directions were estimated as standard deviation of *0.025* and *0.0217* pixels. Further tests illustrated the stability of the location of a moving target as standard deviation of *0.022* pixels. These stability values are comparable to those obtained from a frame grabber acquired data.

The real-time image data acquisition/processing/transfer and the establishment of correspondences using four views of a moving and stationary objects was successfully carried out. It was noted that for the intelligent camera/*CPC* data transfer and the establishment of correspondences for a set of four views it took approximately *0.9ms* and *7.37ms*. The target image processing time taken by each intelligent camera system is not estimated in this experiment. It can be approximately *2.8ms* according to previous estimations (Clarke *et al*, 1998).

Further tests were carried out to assess the performance of the *PEL* algorithm with frame-grabber acquired data. For these tests accurate camera exterior and interior parameters were available. In the first test, four views of a dense target field was acquired and correspondences were established. The minimum distance between the two closest points was approximately *8 pixels* and the standard deviation of the image point location error was *0.27μm*. A search band of *2μm* in width was used.

The next test on correspondences was carried out on images acquired from four views of a target test plate. In this case, the accurate camera exteriors were available. However, not all the targets in images were located accurately due to the poorly formed images. Certain target images of few pixels in size and had lower peak

intensities hence were poorly located. A search band of larger width was used still poorly located target were not corresponded in all four views.

Chapter 8

Achievements, suggestions for further work, and summary

The work carried out for this thesis contributed towards the development of a real-time *3-D* measurement system based on close-range photogrammetry and computer vision principles. The previous chapters described the work carried out. In this chapter the major achievements of the work, the suggestions for further work, and the summary of the thesis are given.

8.1 Achievements of the work carried out for the thesis

The achievements of the work comprises the contribution to the development of intelligent camera system, the development of *PEL* algorithm, and the *3-D* measurement system integration. The remainder of this section provides further details of these achievements.

Contribution to the development of intelligent camera system

This part of the work contributed towards the development of an intelligent camera system for real-time video data processing and data communications. The first version of the video data acquisition, processing, and Ethernet communication software were developed. In addition, Ethernet hardware was also developed.

Intelligent camera system has real-time video processing and data communication capabilities. The real-time performances are due to the unique architecture of the system and use of efficient video processing and data communication techniques. An intelligent camera is capable of producing about *170* target image locations and transferring these location data to a *CPC* at *CCIR* video frame rate (*25Hz*). The *2-D* target image location accuracy is comparable to that of frame grabbers.

Development of *PEL* algorithm

This part of the work resulted in the development of *PEL* algorithm for the establishment of fast multiple view point correspondences. It was shown that the *PEL* technique can establish faster multiple view point correspondences than *EL* and *3DSI*

methods. This is mainly due to the use of parallel epipolar lines, novel technique for determining the search band in stereo pairs, and target image identity based distribution of stereo pair correspondences. However, due to the initialisation step required for the pre-estimation of parameters, this method is faster than *EL* and *3DSI* techniques when the number of points are greater than a certain value. The timing tests showed that this technique is faster than *EL* and *3DSI* methods when the number of points are greater than *75* and *25*.

Integration of the real-time system

This part of the work contributed to the integration of multiple intelligent cameras and the *PEL* technique to perform real-time target location data acquisition and multiple view point correspondence establishment. Experiments were successfully conducted for the establishment of real-time correspondences of target images acquired from moving objects.

8.2 Suggestions for further work

In practice, establishment of point correspondences between images that are acquired from arbitrary number of cameras is required. Various multiple view configurations that may be used in practice can be imagined. The placement of cameras depends on the application. For instance, in *3-D* modelling the cameras may be placed in random locations to view the object from different view points. In a high precision *3-D* measurement applications, cameras may be positioned to obtain the optimum network with respect to the measurement accuracy. In a low precision autonomous vehicle navigation task, cameras may be positioned in a very specific configuration to obtain speed performance in establishing correspondences which may exploit geometric relationships between two or three views. Hence, the selection of positions for the cameras need to be considered with respect to obtaining *3-D* information as well as establishing correspondences.

In this chapter various ways of using *PEL* algorithm with an arbitrary number of views is discussed. The arbitrary number of views can be divided into two categories where the number of views may be positioned in two configurations,

1) Systematic configuration

2) Random configuration

Systematic configuration:

Figure 8.1 (a) and (b) illustrate two examples of systematically positioned n views in a circular configuration. Many other configurations where cameras are systematically positioned can be imagined. The positioning cameras on the vertices of a polygon are seen in many close range modelling applications. In most of these cases, the same points on object(s) are not viewed by all the cameras.

(a)

(b)

Figure 8.1 Systematically positioned n cameras

For the establishment of correspondences, these configurations can be considered as similar to those discussed in section 4.3. The multiple views can be considered as combinations of stereo pairs and the stereo pair correspondences can be established. Then the corresponding points between all views can be picked by searching through the established stereo pair correspondences. It is expected that not all the targets will be seen by all views. Hence, it is required to consider subsets of the views. Initially, targets that can be seen by all n views are corresponded. Subsequently, combinations of $(n-1)$, $(n-2)$, $(n-3)$, ..., and 3 views can be used to establish correspondences.

Random configuration:

Establishment of point correspondences between arbitrarily positioned n views is more complex. Figure 8.2 illustrates arbitrarily positioned cameras. In this situation it is not possible to distribute the correspondences simply as in systematic configuration. A technique is suggested here which uses subsets of three views at a time and the correspondences are distributed in a progressive manner.
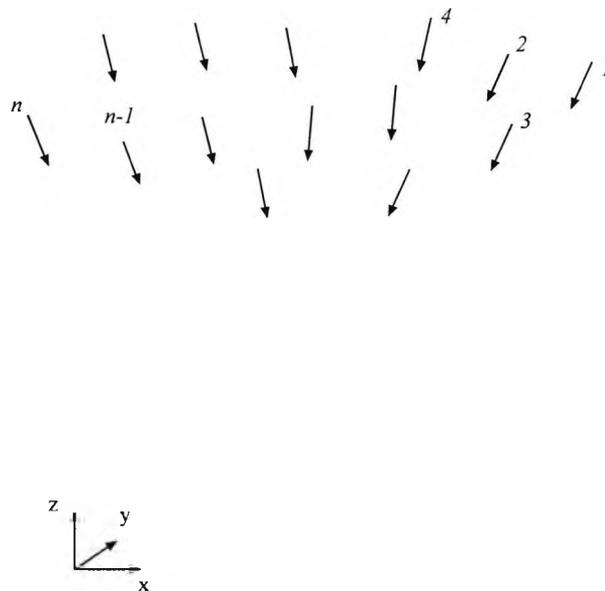

Figure 8.2 Arbitrarily positioned n views

The steps involved could be as follows,

1. From a given set of n views, a view is selected that is approximately at the centre among other views. Such a camera can be identified by using the exterior location parameters.

2. Select a view that is nearest to the one selected in step-*1*.

3. Select a view which is closest to the baseline formed by stereo pair selected in step-*1* and step-*2*. The criteria for the selection is that the two base lines of triangle formed by a third camera should have angles that are less than or equal to *90°* with the base line of the stereo pair selected in step-*1* and step-*2*. If there is no camera that satisfies this criteria then there will not be a triangle formation. The other criteria is that a camera that is selected once will not be selected again as the third camera of a triangle.

4. Establish correspondences between three views selected in steps *1 to 3*.

5. Repeat step-*3* for each stereo pair combination of the three views and establish correspondences.

An example is illustrated as in Figure 8.3 where each node represents a camera. Assuming that camera-*1* as the seed, the nearest camera to that is camera-*2*. The first triangle is formed by adding camera-*3* that satisfy the criteria in step-*3*. The correspondences are established and a further three triangles are formed by considering cameras-*4*, *5*, and *6*. This process is repeated until all cameras are used and any number of cameras can be connected.
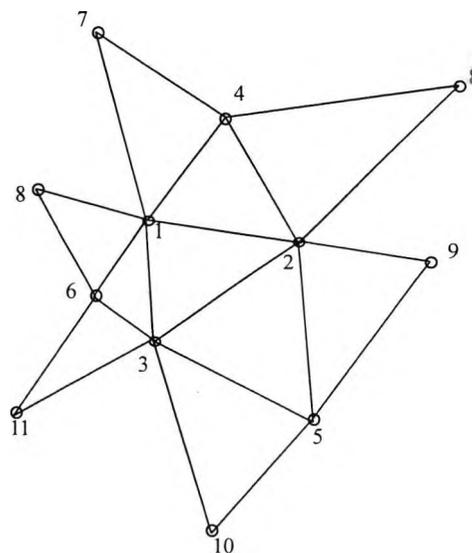

Figure 8.3 Selection of three views

8.3 Summary

In close range measurement applications it is required to establish image point correspondences between multiple views. Due to the accuracy requirements targets that are made of material with special reflective characteristics are placed on the object points of interest. Special illumination techniques are used to obtain the maximum reflected light from targets onto the image planes. Images of the targeted object(s) are acquired from multiple view points and the accurate target image locations are obtained. Normally, the background clutter in images are removed by setting the camera aperture smaller. For the establishment of correspondences image point locations and accurate or approximate camera exterior parameters are available. Hence, geometric techniques can only be used for establishing image point correspondences.

This thesis was devoted to the subject of geometric techniques for multiple view point correspondences. The conventional and new geometric techniques were surveyed. The *EL*, *3DSI*, back projection, and epipolar line slope techniques were reported in the literature. The *EL* and *3DSI* methods were implemented for the purpose of performance comparisons. A common characteristic of these methods is that they require a significant amount of time for searching. Considering a stereo pair with n points in each image, for a given point in one image the corresponding or potentially corresponding point(s) in the other image may be obtained by searching through all the points. Hence, requires an n^2 search which is not favourable for real-time applications involving large number of points.

In this thesis, a technique was introduced which uses the image rectification. This technique uses parallel epipolar lines in rectified images, hence it is termed the *PEL* algorithm. Stereo pair rectification is a technique used for many years. Normally, stereo images acquired with narrow angular separation are rectified to gain the advantage of parallel epipolar lines for matching. The narrow angular separation is used to avoid matching problems which occur with wide angular separation due to radiometric effects. The *PEL* algorithm uses the rectification for establishing point correspondences between multiple views that have wide angular separations. In this algorithm only the target images that are located in convergent image space are

transformed into the rectified image space but not the complete image. Hence, there will not be any radiometric or geometric distortions due to the transformation.

The other distinguishing feature of this *PEL* algorithm is the establishment of correspondences in the rectified image space. When image points are transformed, their location errors in convergent image space are also transformed. If camera exterior parameters have errors the rectified images will not have collinear conjugate epipolar lines. Therefore, to find the corresponding point in one image for a selected point in the other image a search band is required. Normally, a parallel search band that covers either side of the epipolar line is used. In the *PEL* algorithm a new technique is introduced to determine the search band in rectified image space.

In convergent image space, a search band is used to establish correspondences. If this search band is transformed into rectified image space, the corresponding points should ideally fall within the transformed band. The search band border lines that are parallel to the epipolar line in convergent image space are sloped with the epipolar line in the rectified image space. It was shown that the slope of these transformed border lines follow curves which can be represented by 2nd order polynomials as a function of the y-co-ordinate in the rectified image space. For a selected point in one rectified image the collinear epipolar line in the other rectified image is known automatically. Knowing the y-co-ordinate, the search band border line slopes can be directly computed using the polynomials. Another characteristic of the border lines is that they intersect at a point which lies on the epipolar line. The x-co-ordinate of this intersection point follows a straight line as the y-co-ordinate varies from top to bottom of the image hence it can be represented by a straight line. As a result of these two relationships, for a given y-co-ordinate in the rectified image, the slopes of the relevant border lines and the co-ordinates of the intersection point can be computed. Hence, points that fall within the band can be found with less searching provided that the rectified image points are sorted in the ascending order of the y-co-ordinate.

The *PEL* algorithm requires an initialisation step to compute rectification parameters, and polynomial coefficients. With the current implementation this requires *1.0ms* for a stereo pair. This type of initialisation step is not required by *3DSI*, *EL*, or back

projection methods. Hence, the *PEL* method is faster than these methods when number of points are larger than a certain number which is dependant upon the implementation.

References and Bibliography

Abdel-Aziz, Y. and Karara, H.M., 1971. Direct linear transformation from comparator co-ordinates into object space co-ordinates in close-range photogrammetry. American society of photogrammetry symposium on close-range photogrammetry, Urbana, Illinois. pp. 1-18.

Advanced Micro Devices, 1993/1994. EPROM products hand book. Advanced Micro Devices Inc., U.S.A.

Albert, A.A., 1949. Solid analytic geometry. University of Chicago Press. Phoenix science series. 158 pages.

Analog Devices, 1995. ADSP-2100 Family user's manual. Analog Devices Inc. U.S.A.

Analog Devices, 1995. ADSP-2100 Family EZ-Kit Lite reference manual. Analog Devices Inc. U.S.A.

Ariyawansa, D.D.A.P. and Clarke, T.A., 1997. High speed correspondences for object recognition and tracking. Videometrics V. SPIE Vol. 3174. San Diego, U.S.A., pp. 70-79.

Ariyawansa, D.D.A.P. and Clarke, T.A., 1999. High speed multiple view image point correspondences using rectification. Videometrics VI. SPIE Vol. 3174. San Jose, U.S.A., pp. 70-79.

Atkinson, K.B. (Editor), 1996. Close-range photogrammetry and machine vision. Whittles Publishing, UK. 362 pages.

Atkinson, K.B (Editor), 1980. Developments in close-range photogrammetry-1. Applied Science Publishers, London. pp. 1-80.

Atkinson, K.B. (Editor), 1977. Selection of papers by E.H. Thompson. Photogrammetric society, London. pp. 33-102.

Ayache, N. and Hansen, C. 1988. Rectification of images for binocular and trinocular stereovision. Proc. 9th International conference on pattern recognition, Beijing, China. Vol. 1, pp. 11-16.

Ayres, F. (JR), 1967. Theory and problems of projective geometry. McGraw-Hill Book Company. 229 pages.

Baltsavias, E.P. 1991. Multiphoto geometrically constrained matching. PhD thesis. Institute for Geodesy and Photogrammetry, ETH, Zurich. 221 pages.

Beyer, H.A., 1992; Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system. PhD thesis. Institute for Geodesy and Photogrammetry, ETH, Zurich. 177 pages.

Boufama, B, Mohr, R., and Morin, L., 1998. Using geometric properties for automatic object positioning. Image and vision computing. Elsevier Publishers. Vol. 16. pp. 27-33.

Brey, B.B., 1991. The Intel microprocessors: 8086/8088, 80186, 80286, 80386, and 80486. Architecture, programming, and interfacing. Maxwell Macmillan International editions, 616 pages.

Candocia, F. and Adjouadi, M., 1997. A similarity measure for stereo feature matching. IEEE Transactions on image processing, Vol. 6, No. 10. pp. 460-464.

Chen, J. 1995. The use of multiple cameras and geometric constraints for *3-D* measurement. PhD thesis, City University, London. 245 pages.

Chen, J., Clarke, T.A., and Robson, S., 1996. An alternative to the epipolar line method for automatic target matching in images for 3-D measurement. Optical 3-D measurement techniques II, Published by Wichmann, Karlsruhe: pp. 197-204.

Chen, J. Clarke, T.A. Cooper, M.A.R. Grattan, K.T.V.G. 1995. An optimised target matching based on a 3-D space intersection and a constrained search for multiple camera views, Videometrics III. SPIE Vol. 2350. Boston, U.S.A., pp. 324 - 335.

References and bibliography

Chen, J., and Clarke, T.A., 1992. The automatic recognition, location and labelling of targets in digital photogrammetric engineering measurement. International archives of photogrammetry and remote sensing. Vol. XVII(5): pp. 686-693.

Clarke, T.A., Cooper, M.A.R., Chen, J., and Robson, S., 1995. Automated three dimensional measurement using multiple CCD camera views. Photogrammetric record, Vol. 15(85), pp. 27-42.

Clarke, T.A., Cooper, M.A.R., and Fryer, J.G., 1993. An estimator for the random error in subpixel target location and its use in bundle adjustment. Optical 3-D measurement techniques II, Published by Wichmann, Karlsruhe: pp. 161-168.

Clarke, T.A., and Wang, X., 1996. An analysis of subpixel target location accuracy using Fourier Transform based models. Videometrics IV. SPIE Vol.2598. Philadelphia. pp. 65-76.

Clarke, T.A., Gooch, R.M., Ariyawansa, D.D.A.P., Wang, X., 1997. 3D-Net: The development of a new real-time photogrammetric system. Videometrics V. SPIE Vol. 3174. San Diego, U.S.A., pp. 222 - 233.

Clarke, T.A., Li, L., and Wang, X., 1988. *3D-NET* - A new real-time photogrammetric system. ISPRS commission V symposium on real-time imaging and dynamic analysis. Vol. XXXII(5). Hakodate, Japan. pp. 29-34.

Collins, R.T., 1996. A space-sweep approach to true multi-image matching. Proc. IEEE conference on computer vision and pattern recognition. Vol. 16, No. 2. pp. 385-363.

Dhond, U.R., Aggarwal, J.K.,1989. Structure from stereo - A review. IEEE Transactions on systems, man and cybernetics, Vol. 19, No 6. pp. 1489-1510.

Dhond, U.R. and Aggarwal, J.K., 1991. A cost-benefit analysis of a third camera for stereo correspondence. International journal of computer vision, 18. Kluwer Academic Publishers. pp 5-19.

Elantec, 1994. High performance analog integrated circuits, Video-Instrumentation-Power control. Elantec Inc. pages 4-153.

Ellis, T.J., 1987. A programmable processor for real-time feature extraction. SPIE Intelligent robotics and computer vision VI. Boston, U.S.A., Vol. 848, pp. 594-599.

Faugeras, O. and Mourrain, B., 1995. On the geometry and algebra of the point and line correspondences between N images. INRIA research report no. 2665. INRIA, 2004 route des Lucioles, BP 93, F-06902 Sophia-Antipolis, Cedex, France.

Faugeras, O. and Robert, L., 1996. What can two images tell us about a third. International journal of computer vision, 18. Kluwer Academic Publishers. pp 5-19.

Foerstner, W., 1986. A feature based correspondence algorithm for image matching. Proc. ISPRS Commission III Symposium, Rovaniemi, Finland, August 19-22, 1986. In ISPRS, Vol.26 -3/3, pp. 150 - 166.

Fraser, C.S. 1997. Automation in digital close-range photogrammetry. Proc. First tasman surveyors conference, Newcastle, Australia.

Fua, P., 1997. From multiple stereo views to multiple 3-D surfaces. International journal of computer vision, 24(1). Kluwer Academic Publishers, Boston. pp 19-35.

Ghosh, S.K., 1972. Theory of stereophotogrammetry. Printed by Edwards Brothers Inc., U.S.A. pp. 18-120.

Gooch, R.M., Clarke, T.A., and Pushpakumara, D.D.A., and Ellis, T.J.,1996. Driving a DSP underground. Proc. DSP'96 Scandinavia. Copenhagen, Denmark, pp. 159-168.

Georgis, N., Petrou, M., and Kittler, J., 1998. Image and vision computing. Vol. 16, Elsevier Publishers. pp. 35-41.

Gonzalez, R. and Woods, R.E., 1993. Digital image processing. Addison-Wesley publishing company. 482 pages.

Gruen, A. and Baltsavias, E.P., 1988. Geometrically constrained multiphoto matching. Photogrammetric Engineering and Remote Sensing. Vol. 54(5), pp.633-641.

Haralick, R.M. and Shapiro, L.G., 1992. Computer and robot vision, Volume I. Addison-Wesley Publishing Company, Inc. U.S.A. 630 pages.

References and bibliography

Haralick, R.M. and Shapiro, L.G., 1993. Computer and robot vision, Volume II. Addison-Wesley Publishing Company, Inc. U.S.A. 366 pages.

Hegering, H.-G. and Lapple, A., 1993. Ethernet - Building a communications infrastructure. Addison-Wesley Publishing Company, London. 383 pages.

Hearn, D. and Baker, M.P., 1997. Computer graphics - C version. Second edition. Prentice-Hall, U.S.A. 624 pages.

Hitachi, 1995. IC memory data book. Hitachi document centre, Japan.

Horowitz, P. and Hill, W., 1985. The art of electronics. Cambridge university press. 704 pages.

Ito, M., and Ishii, A., 1986. Three-view stereo analysis. IEEE transactions on pattern analysis and machine intelligence. Vol. PAMI-8, No. 4. pp 524-532.

Kanatani, K., 1993. Geometric computation for machine vision. Oxford University Press, U.K. pp. 148-178.

Koch, R., 1994. Model based 3-D scene analysis from stereoscopic image sequences. ISPRS Journal of photogrammetry and remote sensing, 49(5), pp. 23-30.

Jack, K., 1996. Video demystified. Second edition. High text interactive Inc. U.S.A. 282 pages.

Kanade, T., Kano, H., and Kimura, S., 1999. Development of a video-rate stereo machine. Web site. CMU, Robotics institute, Carnegie Mellon University, U.S.A.

Kang, S.B. and Szeliski, R., 1996. 3-D scene recovery using ominidirectional multibaseline stereo. Proc. IEEE conference on computer vision and pattern recognition. pp. 364-370.

Kang, S.B., Szeliski, R., and Shum, H.-Y., 1997. A parallel feature tracker for extended image sequences. Note: Computer vision and image understanding. Vol.67, No.63, pp. 296-310.

References and bibliography

Kreyszig, E., 1988. Advanced engineering mathematics. Sixth edition. John Wiley & Sons Inc, U.S.A. 1294 pages.

Lawrence, P.D. and Mauch, K., 1988. Real-time computer systems design: An introduction. McGraw-Hill Publishers. 556 pages.

Lattice Semiconductor, 1996. Data book, Lattice Semiconductor corporation, U.S.A.

Lipschutz, S., 1965. Theory and problems of probability. SI (Metric edition). McGraw-Hill Book Company, U.K. pp. 37-125.

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. Nature Vol.293. pp 133-135.

Luong, Q.-T., Deriche, R., Faugeras, O., and Papadopoulo, T., 1993. On determining the fundamental matrix: Analysis of different methods and experimental results. Research report no 1894. Programme 4: Robotics, Image and Vision. INRIA Sophia Antipolis, BP 93-Sophia-Antipolis, Cedex, France.

Luong, Q.-T., and Faugeras, O., 1996. The fundamental matrix: Theory, algorithms, and stability analysis. International Journal of computer vision, 17. Kluwer Academic Publishers, Boston, pp. 43-75.

Luong, Q.-T. and Vieville, T., 1996. Canonical representations for the geometries of multiple projective views. Computer vision and image understanding, Vol. 64, No 2, September. Academic Press Inc, pp. 193-229.

Mason, S.O., 1994. Expert system-based design of photogrammetric networks. PhD Thesis, Institute for geodesy and photogrammetry, Swiss Federal Institute of Technology (ETH), CH-8093 Zurich. 173 pages.

Mass, H- G., 1992. Complexity analysis for the establishment of image correspondences of dense spatial target fields. International archives of photogrammetry and remote sensing. Commission V, Vol. XXIX(B5). pp 102-107.

Mikhail, E.M., 1976. Observations and least squares. Dun-Donnelley publisher, New York. 484 pages.

National Semiconductor, 1996. Ethernet data book. National semiconductor corporation, U.S.A. 1-278 pages.

Nilsen, J.H. and Hadem, I., 1994. Photogrammetric tracking of traver particles in modelled ocean flows. ISPRS Journal of photogrammetry & remote sensing. Vol. 49. Number 6. pp. 9-20.

Oda, K., Tanaka, M., Yoshida, A., Kano, H., and Kanade, T., 1996. A video-rate stereo machine and its application to virtual reality. International archives of photogrammetry and remote sensing. Vol. XXXI(B5), Vienna. pp. 411-415.

Okutomi, M. and Kanade, T., 1993. A multiple-baseline stereo. IEEE Transactions on pattern analysis and machine intelligence. Vol. 15, No. 4, pp. 353-363.

Pavlidis, T., 1982. Algorithms for graphics and image processing. Springer-Verlag, Berlin. pp. 167-193.

Pla, F. and Marchant, J.A., 1997. Matching feature points in image sequences through a region-based method. Computer vision and image understanding. Vol. 66, No. 3, pp. 271-285.

Pushpakumara, D.D.A. 1995. The use of a Digital Signal Processor in a real-time image processing application. MSc Thesis, City University, U.K. 120 pages.

Pushpakumara, D.D.A., Gooch, R.M., Clarke, T.A., 1996. Real-time image processing and data communications for a real-time 3-D measurement system. Proc. 15th National Information Technology Conference, The Computer Society of Sri Lanka, pp. 1-9.

Pulnix. Miniature CCD camera-TM-6CN, operations and maintenance manual. Pulnix corporation. 216 pages.

Rosenfeld, A. 1998. Survey - Image analysis and computer vision:1997. Computer vision and image understanding. Vol. 70, No. 2, pp. 239-284.

Rothwell, C., Csurka, G., and Faugeras, O., 1995. A comparison of projective reconstruction methods for pairs of views. INRIA research report no. 2538. INRIA, 2004 route des Lucioles, BP 93, F-06902 Sophia-Antipolis, Cedex, France.

Ruther, H. 1996. Multi-image correlation for digital photogrammetric measurement systems. International archives of photogrammetry and remote sensing. Vol. XXXI(B5). Vienna. pp. 499-505.

Sabel, J.C., van Veenendaal, H.L.J., and Furnee, E.H., 1993 . PRIMAS, a real-time *3-D* motion analysis system. Optical 3-D measurement techniques II - applications in inspection, quality control and robotics. Published by Wichmann publishers. pp. 530-537.

Shah, Y.C., 1985. Extraction of range information from stereo video images. PhD thesis. City University, London. 215 pages.

Shapiro, L. and Rosenfeld, A., (Editors) 1992. Computer vision and image processing. Academic Press, Inc. pp. 593-607.

Shih, T.Y., 1990. An investigation into the computational procedures of space resection by collinearity equations. Photogrammetric record. Vol.13, No.75, pp. 433-440.

Shortis, M.R., Clarke, T.A., Short, T., 1994. A comparison of some techniques for the subpixel location of discrete target images. Videometrics III. SPIE Vol. 2350. Boston, U.S.A., pp. 239-250.

Slama, C.C., (Editor), 1980. Manual of Photogrammetry, 4th Edition, American Society of Photogrammetry. pp. 1-400.

Smit, J.L., 1997. Three dimensional measurement of textured surfaces using digital photogrammetric techniques. PhD Thesis. University of Cape Town. 134 pages.

Sommerville, D.M.Y. 1934. Analytical geometry of three dimensions. Cambridge University Press, U.K. pp. 1-73.

Sonka, M., Hlavac, V., and Boyle, R., 1993. Image processing, analysis and machine vision. Chapman and Hall Publishers. 542 pages.

Sony, 1992. Semiconductor IC data book, CCD cameras & peripherals. Distributed by Anzac, Burnham, U.K. pp. 235-266.

Stockman, G., Chen, J.-L., Cui, Y., and Reynolds, H., 1997. Measuring body points on automobile drivers using multiple cameras. Image and vision computing. Vol. 15. Elsevier Publishers, pp. 317-329.

Stroud, K.A., 1986. Further engineering mathematics, programmes and problems. Macmillan Education. 775 pages.

Texas Instruments, 1989. TTL data book, Vol.1. Texas Instruments Inc. U.S.A.

Texas Instruments, 1984. TTL data book, Vol.2. Texas Instruments Inc. U.S.A.

Texas Instruments, 1995. Higher performance FIFO memories, Designers hand book, Texas Instruments Inc. U.S.A.

Triggs, B., 1997. Linear projective reconstruction from matching tensors. Image and vision computing, Vol. 15. Elsevier Publishers, pp. 617-625.

Trucco, E. and Verri, A., 1998. Introductory techniques for *3-D* computer vision. Published by Prentice-Hall Inc., NJ, U.S.A. pp. 139-170.

Venkateswar V. and Chellappa, R., 1995. Hierarchical stereo and motion correspondence using feature groupings. International journal of computer vision, Vol. 15. Kluwer Academic Publishers, Boston, pp. 245-269.

Vujovic, N. and Brzakovic, D., 1997. Establishing the correspondence between control points in pairs of mammographic images. IEEE Transactions on image processing, Vol. 6, No. 10. pp. 1388-1399.

Wang, X., 1997. Separate adjustment of close range photogrammetric measurements. PhD Thesis. City University, London. 200 pages.

Wang, W. and Duncan, J.H., 1996. Recovering the three-dimensional motion and structure of multiple moving objects from binocular image flows. Computer vision and image understanding. Vol.63, No.3, pp. 430-446.

Webb, J.A., Warfel, T., and Kang, S.B., 1997. A scalable video rate camera interface. Internal report no. CMU-CS-94-192 (15213-3891). School of computer science, Carnegie Mellon University, U.S.A.

West, G.A.W., 1982. Automatic visual inspection of printed circuit boards. PhD thesis, City University, London. pp. 106-246.

Wolf, P.R., 1983. Elements of Photogrammetry with air photo interpretation and remote sensing, 2nd Edition McGraw-Hill. 614 pages.

Zheng, Z. and Wang, W., 1992. A general solution of a closed-form space resection. Photogrammetric Engineering and Remote sensing, Vol.58(3). pp 327-338.

Zhang, Z., 1996. On the epipolar geometry between two images with lens distortion. IEEE Proc. International conference on pattern recognition, Vol. I. Vienna, Austria. pp. 407-411.

Zhang, Z., 1996. Determining the epipolar geometry and its uncertainty: A review. INRIA research report no. 2927. INRIA, 2004 route des Lucioles, BP 93, F-06902 Sophia-Antipolis, Cedex, France.

Appendix

Partial differentials of collinearity equations with respect to camera exterior parameters and *3-D* locations:

Collinearity equations can be written as follows,

$$F_x = x = -f\left(\frac{M1}{M3}\right) \tag{A.1}$$

$$F_y = y = -f\left(\frac{M2}{M3}\right) \tag{A.2}$$

where,

$$M1 = \left(m_{11}(X - X0) + m_{12}(Y - Y0) + m_{13}(Z - Z0)\right) \tag{A.3a}$$

$$M2 = \left(m_{21}(X - X0) + m_{22}(Y - Y0) + m_{23}(Z - Z0)\right) \tag{A.3b}$$

$$M3 = \left(m_{31}(X - X0) + m_{32}(Y - Y0) + m_{33}(Z - Z0)\right) \tag{A.3c}$$

Partial differentials with respect to unknown *3-D* locations (X, Y, Z) are,

$$\frac{\partial F_x}{\partial X} = f\left((m_{11}M3 - m_{31}M1)/M3^2\right) \tag{A.4}$$

$$\frac{\partial F_y}{\partial X} = f\left((m_{31}M1 - m_{11}M3)/M3^2\right) \tag{A.5}$$

$$\frac{\partial F_x}{\partial Y} = f\left((m_{12}M3 - m_{32}M1)/M3^2\right) \tag{A.6}$$

$$\frac{\partial F_y}{\partial Y} = f\left((m_{32}M1 - m_{12}M3)/M3^2\right) \tag{A.7}$$

$$\frac{\partial F_x}{\partial Z} = f\left((m_{13}M3 - m_{33}M1)/M3^2\right) \tag{A.8}$$

$$\frac{\partial F_y}{\partial Z} = f\left((m_{33}M1 - m_{13}M3)/M3^2\right) \tag{A.9}$$

Appendix

Partial differentials with respect to unknown camera exterior parameters (*X0, Y0, Z0, ω, φ, κ*) are,

$$\frac{\partial F_x}{\partial X0} = f\left((m_{31} M1 - m_{11} M3)/M3^2\right) \tag{A.10}$$

$$\frac{\partial F_y}{\partial X0} = f\left((m_{31} M2 - m_{21} M3)/M3^2\right) \tag{A.11}$$

$$\frac{\partial F_x}{\partial Y0} = f\left((m_{32} M1 - m_{12} M3)/M3^2\right) \tag{A.12}$$

$$\frac{\partial F_y}{\partial Y0} = f\left((m_{32} M2 - m_{22} M3)/M3^2\right) \tag{A.13}$$

$$\frac{\partial F_x}{\partial Z0} = f\left((m_{33} M1 - m_{13} M3)/M3^2\right) \tag{A.14}$$

$$\frac{\partial F_y}{\partial Z0} = f\left((m_{33} M2 - m_{23} M3)/M3^2\right) \tag{A.15}$$

$$\frac{\partial F_x}{\partial \omega} = f\left(\left(M3\left(\frac{\partial M1}{\partial \omega}\right) - M1\left(\frac{\partial M3}{\partial \omega}\right)\right)/M3^2\right) \tag{A.16}$$

$$\frac{\partial F_y}{\partial \omega} = f\left(\left(M3\left(\frac{\partial M2}{\partial \omega}\right) - M2\left(\frac{\partial M3}{\partial \omega}\right)\right)/M3^2\right) \tag{A.17}$$

$$\frac{\partial F_x}{\partial \phi} = f\left(\left(M3\left(\frac{\partial M1}{\partial \phi}\right) - M1\left(\frac{\partial M3}{\partial \phi}\right)\right)/M3^2\right) \tag{A.18}$$

$$\frac{\partial F_y}{\partial \phi} = f\left(\left(M3\left(\frac{\partial M2}{\partial \phi}\right) - M2\left(\frac{\partial M3}{\partial \phi}\right)\right)/M3^2\right) \tag{A.19}$$

$$\frac{\partial F_x}{\partial \kappa} = f\left(\left(M3\left(\frac{\partial M1}{\partial \kappa}\right) - M1\left(\frac{\partial M3}{\partial \kappa}\right)\right)/M3^2\right) \tag{A.20}$$

$$\frac{\partial F_y}{\partial k} = f\left(\left(M3\left(\frac{\partial M2}{\partial \kappa}\right) - M2\left(\frac{\partial M3}{\partial \kappa}\right)\right)/M3^2\right) \tag{A.21}$$

Appendix

where,

$$\frac{\partial M1}{\partial \omega} = (\cos\omega \sin\phi \cos\kappa - \sin\omega \sin\kappa)(Y - Y_0)$$
$$+ (\sin\omega \sin\phi \cos\kappa + \cos\omega \sin\kappa)(Z - Z_0)$$

(A.22)

$$\frac{\partial M3}{\partial \omega} = (-(\cos\omega \cos\phi)(Y - Y_0) - (\sin\omega \cos\phi)(Z - Z_0))$$

(A.23)

$$\frac{\partial M1}{\partial \phi} = -(\sin\phi \cos\kappa)(X - X_0) + (\cos\phi \sin\omega \cos\kappa)(Y - Y_0)$$
$$- (\cos\phi \cos\omega \cos\kappa)(Z - Z_0)$$

(A.24)

$$\frac{\partial M3}{\partial \phi} = ((\cos\phi)(X - X_0) + (\sin\omega \sin\phi)(Y - Y_0) - (\cos\omega \sin\phi)(Z - Z_0))$$

(A.25)

$$\frac{\partial M1}{\partial \kappa} = -(\cos\phi \sin\kappa)(X - X_0) + (-\sin\omega \sin\phi \sin\kappa + \cos\omega \cos\kappa)(Y - Y_0)$$
$$+ (\cos\omega \sin\phi \sin\kappa + \sin\omega \cos\kappa)(Z - Z_0)$$

(A.26)

$$\frac{\partial M3}{\partial \kappa} = 0$$

(A.27)

$$\frac{\partial M2}{\partial \omega} = (-\cos\omega \sin\phi \sin\kappa - \sin\omega \cos\kappa)(Y - Y_0)$$
$$+ (-\sin\omega \sin\phi \sin\kappa + \cos\omega \cos\kappa)(Z - Z_0)$$

(A.28)

$$\frac{\partial M2}{\partial \phi} = (\sin\phi \cos\kappa)(X - X_0) + (-\sin\omega \cos\phi \sin\kappa)(Y - Y_0)$$
$$+ (\cos\omega \cos\phi \sin\kappa)(Z - Z_0)$$

(A.29)

$$\frac{\partial M2}{\partial \kappa} = (-\cos\phi \cos\kappa)(X - X_0) + (-\sin\omega \sin\phi \cos\kappa - \cos\omega \sin\kappa)(Y - Y_0)$$
$$+ (\cos\omega \sin\phi \cos\kappa - \sin\omega \sin\kappa)(Z - Z_0)$$

(A.30)