Extraction of Range Information from Stereo Video Images

by Yogendra C. Shah

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

The City University
School of Electrical Engineering and Applied Physics
Department of Electrical and Electronic Engineering

DECEMBER 1985

List of Contents

List of Tables

List of Illustrations

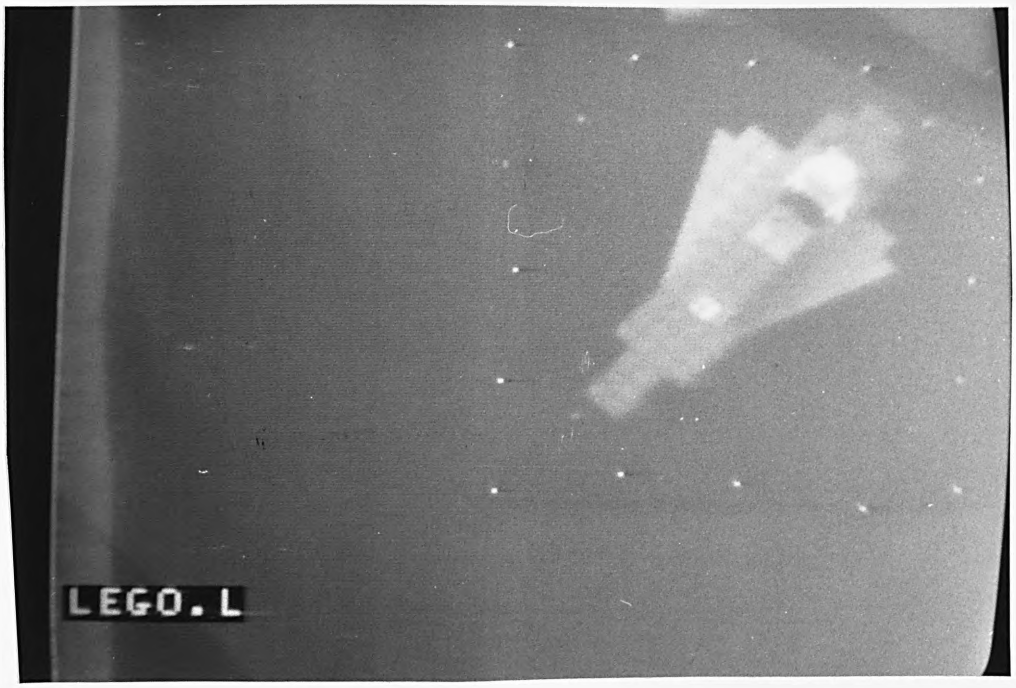


Plate 6.1a - Lego Left Image

Plate 6.1b - Lego Right Image

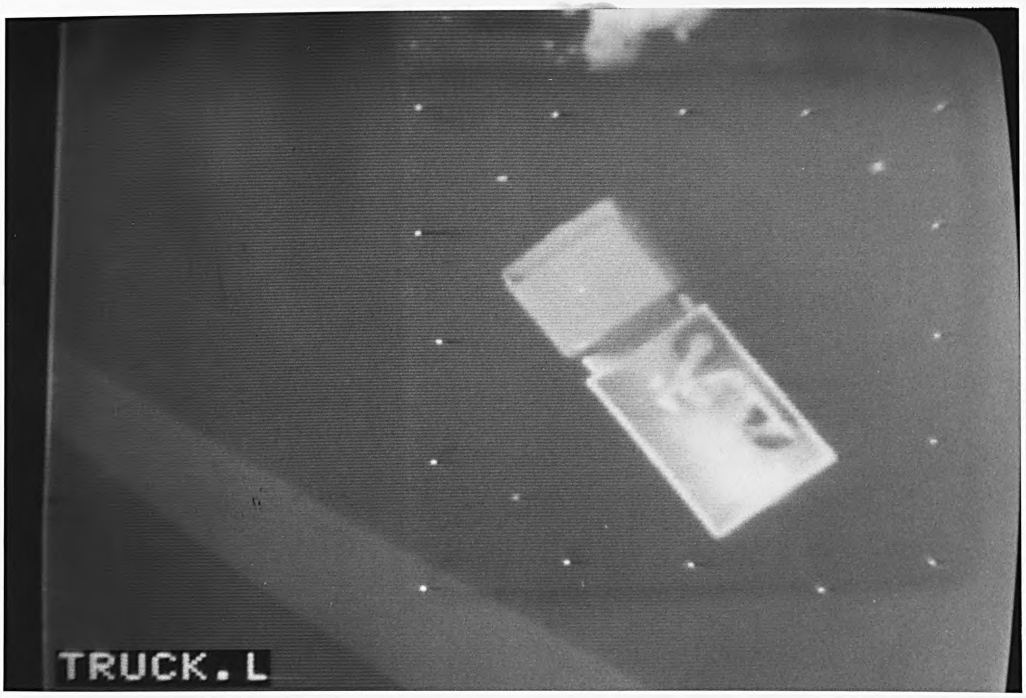


Plate 6.2a - Truck Left Image

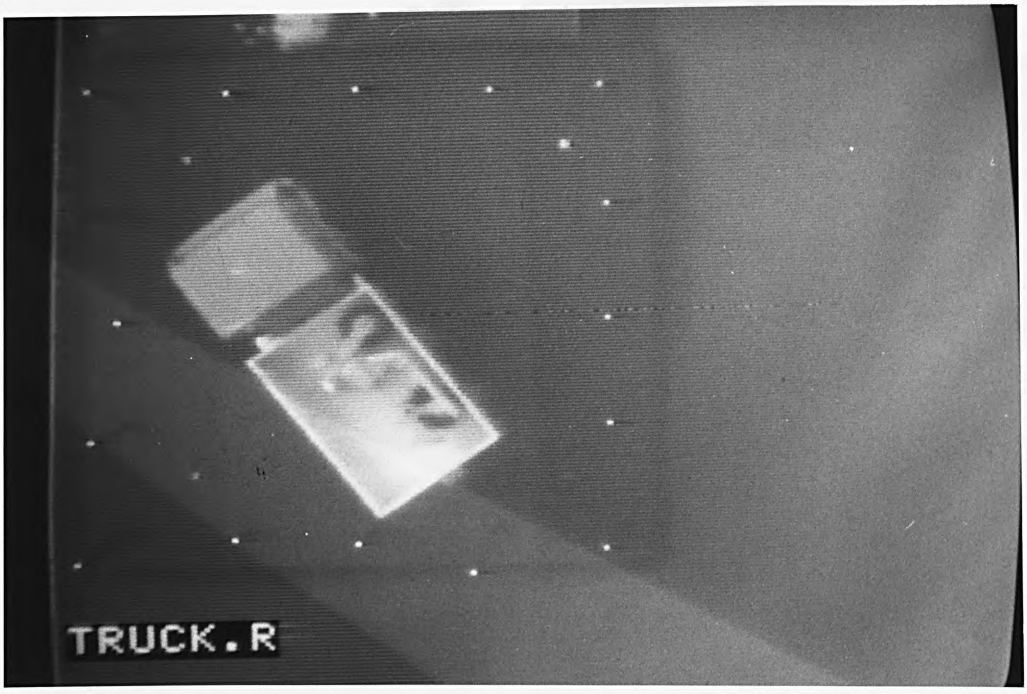


Plate 6.2b - Truck Right Image

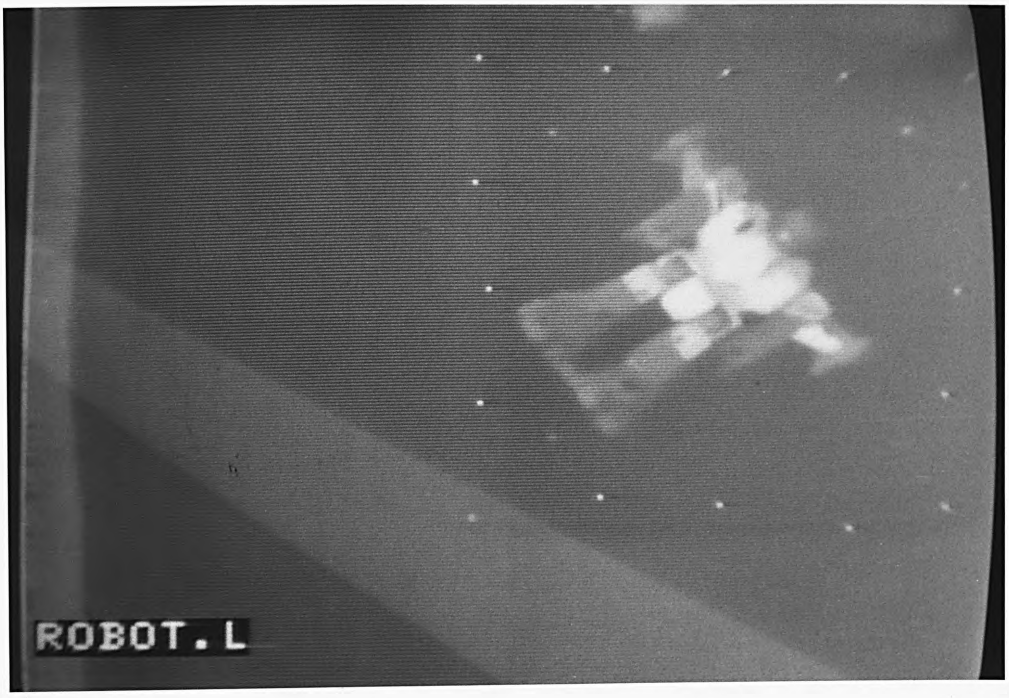


Plate 6.3a - Robot Left Image

Plate 6.3b - Robot Right Image

Acknowledgements

I would like to take this opportunity to thank my supervisor Mr. R. Chapman for his constant supervision and encouragement throughout this thesis without whom it would not have been possible.

Also many thanks to [redacted] for their expert advice in the photogrammetric aspects of this thesis.

I would like to thank GEC Avionics Ltd. for their financial support and use of their facilities; especially [redacted] for giving his time generously.

Finally I would like to extend my gratitude to my colleagues, [redacted] for their invaluable assistance and friendship throughout the period of this project.

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

The aim of this thesis is to outline a versatile stereo video measurement system capable of extracting three-dimensional information from a pair of stereo video images.

When stereo images are used to obtain measurements there are two points which have to be noted and these are that the distortions due to each camera system need to be modelled and that the relative orientation between the cameras needs to be calculated from known control points located in both images.

The City University stereo video system is fully described together with an outline of how measurements may be carried out with the aid of an operator who provides a means of obtaining corresponding stereo points.

An automated technique of acquiring these measurements is then described. The approach used to obtain automatic range information from the stereo images is that of estimating the range image by means of an area correlation process and then iteratively improving upon the range image by successive application of the mapping transformation of coordinates from one stereo image to another.

List of Symbols

x,y	image coordinates in general
x',y'	comparator coordinates in general
x_1,y_1	image one coordinates
x_2,y_2	image two coordinates
x_1',y_1'	image one comparator coordinates
x_2',y_2'	image two comparator coordinates
i,j,k	camera station one orthogonal unit vectors
I,J,K	camera station two orthogonal unit vectors
$V = v_1,v_2,v_3$	translation vector from camera two to camera one perspective centres
n_1,n_2,n_3	normal vector of reference image plane
V_1,V_2	translation vectors from object point to image one and object point to image two camera stations respectively
f	equivalent focal length for a camera, also known as the perspective distance
f_1,f_2	equivalent focal length for camera one and camera two respectively
r	radial distance from the principal point to an image point
r_{ij}	parameter in the relative rotation, between two cameras, matrix
m_{ij}	parameter in the absolute rotation matrix for a single camera
a_{ij}	parameter in matrix for translation of coordinates from one stereo image to the other
X,Y,Z	object space coordinates
X_c,Y_c,Z_c	camera station coordinates
$f(x,y)$	reference image
$f^*(x,y)$	reference image coordinates mapped onto sensed image coordinates

$h(x,y)$	sensed image
$g(x,y)$	gradient image associated with reference image
$r(x,y)$	range image associated with reference image
$r_t(x,y)$	true range image associated with reference image
$r_m(x,y)$	measured range image associated with reference image
$r_p(x,y)$	range image associated with photogrammetric reference image
$r_a(x,y)$	range image associated with digital reference image
$d\_r(x,y)$	incremental range image associated with reference range image
R	individual range value

Parameters used in the image-to-comparator coordinate transforn

$\Delta x_s, \Delta y_s$	symmetric radial lens distortion
a_1, a_2, a_3	parameters in mathematical model of symmetric radial lens distortion
$\Delta x_d, \Delta y_d$	asymmetric radial and tangential lens distortion
p_1, p_2	parameters in mathematical model of asymmetric radial and tangential lens distortion
$\Delta x_t, \Delta y_t$	scanning distortion at the camera tube
q_1, q_2, q_3	parameters in the mathematical model of scanning distortion at the camera tube
a, b, c, d, e	corrections for non-orthogonality and scale changes along image axes
$\Delta x, \Delta y$	total distortion corrections
x_m', y_m'	measured comparator coordinates

Parameters used in the comparator-to-image coordinate transforn

$\Delta x_s', \Delta y_s'$	symmetric radial lens distortion
a_1', a_2', a_3'	parameters in mathematical model of symmetric

	radial lens distortion
$\Delta x_d', \Delta y_d'$	asymmetric radial and tangential lens distortion
p_1', p_2'	parameters in the mathematical model of asymmetric radial and tangential lens distortion
b',c',d',e'	corrections for non-orthogonality and scale changes along image axes
$\Delta x', \Delta y'$	total distortion corrections

Parameters used in the photogrammetric-to-digital image transform

x_d, y_d	measured digital image coordinates
x_p, y_p	measured photogrammetric image coordinates
s,t,u,l,m,n	parameters in the affine transform from the photogrammetric coordinates to the digital coordinates

Introduction

Photogrammetry is an ideal form of measurement in situations where the difficulties, high cost of access and a hostile environment generally preclude direct measurement.

Up to the last few years a still camera system has been used for measurements together with the laborious and time consuming operations involved in extracting three-dimensional information. Any technique which can reduce the time and number of operations required is of great economic benefit. A prime example is that of the oil platform operators. In addition to saving money, reductions in diving time improve the chances of completing tasks on schedule in the limited good weather periods available at sea.

The use of still photographs enables very accurate measurements to be made but it may not always be possible to obtain still photographs in some hazardous situations such as underwater mensuration or inspection of nuclear fuel rods. In these situations a stereo video system with video cameras mounted onto a remotely operated vehicle would enable inspection without any danger to human life.

A stereo video system composed of a pair of video cameras, a stereo viewer, a pair of video monitors, an electronic floating mark (for measurement of conjugate stereo points) and micro-computer equipped with disc drives and plotter can provide fast acquisition of images for later analysis or even on the spot evaluation, up to now impossible with still camera photogrammetry.

In a still camera system it is absolutely necessary to have the use of a stereo comparator which can be bulky and cost anything up to £100,000. Inclusion of the cost of expensive metric cameras and a micro-computer with associated peripherals can push the cost up even further. Compare this with the cost of a flexible and portable stereo video system requiring readily available, inexpensive, hardware and the cost benefits and handling advantages can be easily seen.

Whereas a still camera system can take anything up to a few days to process a pair of stereo images, the stereo video system can provide results within hours or even on the spot results if only isolated points are required.

Any object, provided it is clearly visible in both views, can be measured and the results presented as distances, areas, volumes, contour plots or other graphical output as required. As the method relies upon software it is simple to modify the system to cater for a wide range of applications.

So far the cost and time involved in hiring an experienced operator has not been mentioned. An experienced operator can quickly and accurately obtain a series of conjugate stereo points which can then be used to reconstruct the area of interest in the scene. In order to provide measurements with the stereo video system, at the scene, the operator must be trained to extract measurements from the stereo scene. Consider the cost and time advantages if the requirement of a trained or experienced operator can be eliminated. This requirement can be satisfied if

the extraction of corresponding stereo points can be automated.

The research described in this thesis is aimed at developing a stereo video system capable of extracting three-dimensional information from a pair of stereo video images. The system developed eliminates the need for an experienced operator to laboriously extract conjugate stereo points except in the calibration of the camera system when a few points need to be measured. This task can be carried out by any reasonably competent operator.

The first chapter is a review of literature available regarding the automatic acquisition of three-dimensional information from a pair of stereo images.

The thesis then outlines a brief introduction to the theory of stereo imaging and how three-dimensional measurements may be obtained from a pair of stereo images taking into account the systematic distortions that occur in the imaging system such as symmetric lens distortion, decentreing lens distortion, scale changes along axes, shift of origin and non-orthogonality of the axes. This leads to a discussion of the City University stereo video system and construction details of a hardware unit built for obtaining manual photogrammetric measurements with the aid of the stereo viewer.

This is followed by the description of an algorithm for the automatic extraction of corresponding stereo points which uses the transformation from one image to another to exactly register areas from two stereo images. Hence the only errors which are not

accounted for are due to the reflective properties of the scene causing reflectance variations dependent upon the viewpoint (Shah et al., 1985).

The theory behind the stereo ranging algorithm is discussed with an outline of the final form of the algorithm, derived through application of the algorithm to two computer generated stereo models. Computer generated models offer investigation of the algorithm without having to deal with the problem of noise, correction of distortion in the imaging system and photometric properties of the scene. The development of the algorithm was carried out on an ICL (International Computers Limited) PERQ computer kindly supplied by the SERC (Science and Engineering Research Council), in the 'C` programming language.

Finally the results obtained by application of the algorithm to three models are discussed and a conclusion is drawn together with suggestions for further work to improve the performance of the algorithm.

1.0 Review of Literature

Three-dimensional information about a scene may be extracted using several methods varying from simple contact methods through to complex non-contact techniques requiring some form of energy measurement. The latter is of primary concern here and may be categorised as

1) Echometric,
2) Stereometric and
3) Reflectometric.

1.1 Echometric Techniques

Echometric systems essentially implement time of flight measurements to obtain a measure of the distance travelled by an electro-magnetic wave. A typical example is the laser spot ranger which can use two techniques. The first is to emit a sharp pulse and measure the time of its return. The second approach is to amplitude modulate a light signal and measure the phase difference between the forward going and return signals.

Alternatively ultra-sonic waves may be utilised since sound travels much slower than light, making the problem of timing the return signal much easier than in the laser range finder.

1.2 Stereometric Techniques

Stereometric systems are based on the projection of light patterns onto a scene and measuring the distortion on a camera.

Essentially geometric information in the illumination is used to help extract geometric information about the scene. One such example is light striping where a single plane of light is projected onto a scene, which causes a stripe of light to appear on the scene. The only part of the scene sensed by a viewing camera is the illuminated section and so the problem of correspondence is limited to a single dimension. The plane itself has a known position in object space and any point in the image lies along a straight line, therefore the intersection of the line and plane determines the world coordinates of the image point.

Other examples of the use of structured lighting are those of Posdamer and Altschuler (1982) who projected space-encoded light planes to minimise the number of images required to classify a scene. The other example described by Hall et al. (1982) is where a grid projection technique was used in an active-passive camera system. These techniques are particularly useful for industrial inspection where the lighting can be controlled and the object to camera distances are short.

1.3 Reflectometric Techniques

The reflectometric techniques process one or more images of the scene to obtain surface orientation and thus shape. The use of a dual camera system is known as stereo ranging and is based upon stereopsis, the human process of three-dimensional shape perception. Two images obtained from different positions can give absolute three-dimensional location of a point from the

corresponding image projections. This is achieved through triangulation of the two lines containing the image points and their respective perspective centres. The main problem with this approach has been the determination of corresponding stereo points. Various procedures have been described to tackle the problem of automatically acquiring conjugate stereo points. The methods can be separated into two groups. One method based upon some form of area correlation technique and the other based upon a two step procedure - feature extraction and then matching of these features.

1.3.1 Area Correlation Techniques

Of the area correlation techniques, various attempts have been made to refine the quality of the match such as the use of CCD (charge coupled device) delay lines to correlate the stereo images line by line (Nagata, 1982). Alternatively Dowman and Haggag (1977) described a method of correlation along epipolar lines and discussed the use of alternative, less time consuming methods of correlation such as sum of difference squared. A brief outline of the advantages and disadvantages of these methods is also given. Other methods have been based upon exhaustive 2-D correlation (Real and Fujimoto, 1984) and also 2-D correlation with a coarse to fine match approach (Levine et al., 1973). This type of technique enables the general location of a match point using a coarse match and then a fine match is used to analyse finer details of the images. The template size is selected from an analysis of the spread of intensity values (or variance). The

use of multiple resolution techniques can provide significant computational and functional advantages. The low resolution correlation operations ensure that gross image features are matched rather than details and with the application of high resolution correlation, the finer details may be matched over a limited area.

Model based matching (Burr and Chien, 1977) has also been attempted. Here several images were correlated to reveal edges which were subsequently compared with a wire framed model structure.

A different approach has been applied by Chong et al. (1982) who carried out correlation in the frequency domain rather than the spatial domain. Correlation in the frequency domain is an efficient procedure for suppressing information which is the same in both stereoscopic images, such as distant background, by the use of appropriate filters. Another technique of correlation in the frequency domain has been described by Bennett and Balasubramanian (1973) who implemented high precision, optical processing of near vertical transparancies. Two transparancies are projected by coherent light and by means of optical heterodyning techniques, the correlation coefficient of the transparancies may be measured by placing a detector area at the superposed image plane.

Area correlation techniques work satisfactorily for smoothly undulating relief but are unsuitable where surface slope or range are discontinuous, in relatively textureless regions and where

inter-image surface brightness disparities are extreme.

1.3.2 Feature Extraction Techniques

The feature extraction and matching processes are broadly speaking categorised into two groups. The first group is based upon extracting general image features and then matching. The second group, on the other hand, is based upon psycho-neurological studies and hence attempts to match features extracted, by human beings, in the early stages of visual perception.

There have been several mentions in the literature of feature based matching processes. Hardy and Zavodny (1981) described a method of matching features such as lines, vertex forming line sections ("X", "T" and "L") and region based spatial features (moments and texture). In a similar manner Barnard and Thompson (1980) detailed a stereo matcher which used features such as spots and corners and a consistency property which favoured points that have nearly the same disparities. Arnold and Binford (1980) carried out a statistical analysis of edges and based their matching constraints upon the probability distribution of edge angles and edge intervals.

The use of synthetic descriptions as a basis for the matching criterion is not at all unknown. Sidney (1981) used orthogonal trihedral vertices (interior and exterior corner elements of right parallelpipeds) as the basic matching tool. This process was found to be very versatile and applicable to wide angle stereo where there is gross perspective distortion. The vitterbi

algorithm, applied to speech processing, has also been used for stereo matching (Benard, 1983) where features such as edges (orientation and gradient), blobs (mean grey level) and so on were used as matching parameters.

A graph clustering technique which took into account rotation, translation, magnification and also perspective geometrical normalisation was proposed by Gilmore and Rowland (1983).

In a similar manner to the area correlation algorithms, a coarse to fine strategy may be applied to the matching of features such as edge angle, side intensities, contrast and so on (Baker and Binford, 1981).

A description of a task specific knowledge base which went directly to a higher level 3-D description rather than one based on surface has been described by Herman et al. (1984). This technique allows topology, shape, absolute size and absolute position of scene objects to be easily documented. Therefore information from several images can be grouped together to incrementally acquire the three-dimensional model.

Human stereopsis does not necessarily involve being aware of features to match in either view. Most human beings can fuse two images which by themselves show no apparent depth but when viewed stereoscopically, consists of a three-dimensional scene. Julesz's experiments with random dot stereograms (Julesz, 1971) support the theory that the human visual system is able to distinguish depth purely from disparity differences between images. Another observation was that disparity varies continuously across

unoccluded surfaces and discontinuously only at occluding edges.

A very simple cooperative algorithm, based on this property, was devised by Marr and Poggio (1976) which operated on many input elements and reached a global organisation by way of local, interactive constraints. Marr and Poggio (1977) later refined their algorithm to agree better with psychological data by use of early human visual processing cues such as edge matching at various levels of image smoothing. In a similar manner Mahew and Frisby (1980) used a Laplacian operator followed by a zero-crossing detector to extract edge features from both images before matching. The use of non-orientated spatial-frequency tuned channels enabled a coarse to fine resolution matching.

Attempts have been made to apply the Marr and Poggio stereo matcher in real time environments. Nishihara and Larson (1981) outlined a near real time implementation of the Marr and Poggio stereo matcher by the use of dedicated hardware. Recently Lloyd (1985) proposed a variation on the Marr and Poggio stereo matching algorithm by the use of a Laplacian of a Gaussian filter to detect edges and then classify them using information such as the edge angle, grey level to either side of the edge and edge strength. The edge strength was used as a guide for a coarse to fine match algorithm.

Feature extraction and matching algorithms, although immune to photometric properties of a scene, can provide accurate results. But at the same time it is difficult to match a scene pixel by pixel as in an area matching algorithm and so only small features

such as edges may be matched. Area based approaches offer the ability to use all the intensity information available whereas feature based matching has an inherent loss of information due to its approach.

2.0 Obtaining Measurements from Stereo Images

Three-dimensional information about a scene may be obtained, from a stereo pair of images, by the fact that the scene is being viewed from two different positions.

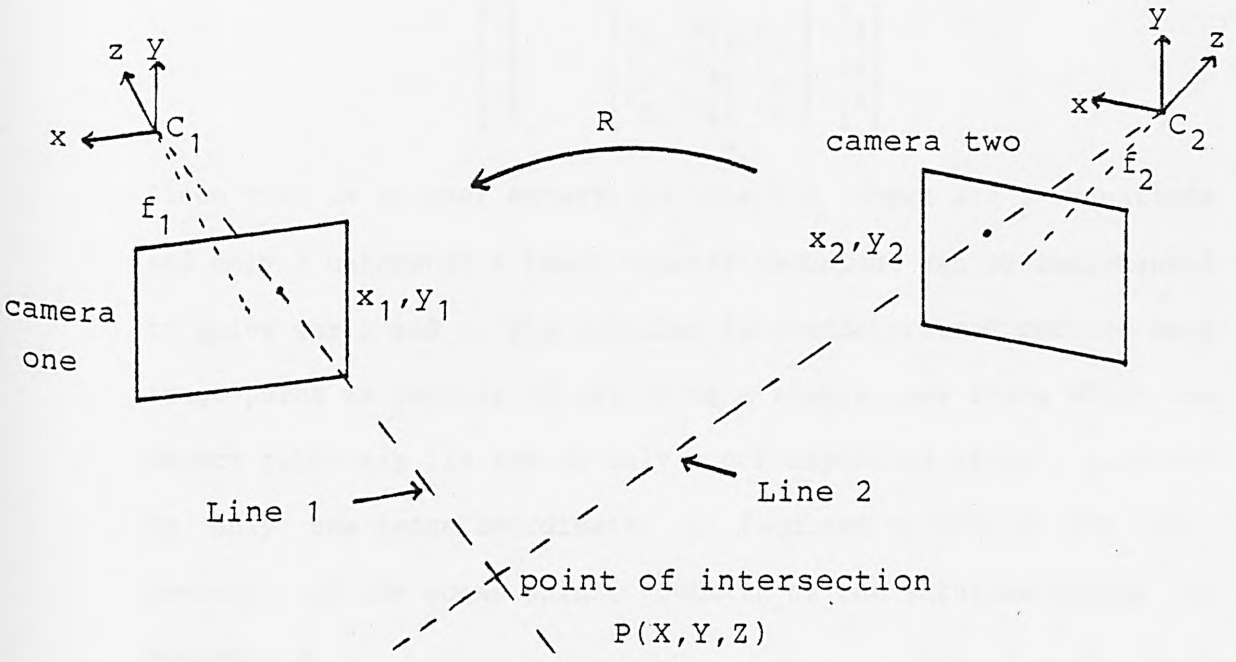


Figure 2.1 - Intersection of Conjugate Lines

Consider the case of a single point in the scene being viewed (figure 2.1). If the point can be located in both images then given that the relative orientation between the cameras is known, the three-dimensional world coordinates may be computed by simple triangulation. If the corresponding image coordinates are (x_1,y_1) and (x_2,y_2) then the world coordinates can be determined by the intersection of two lines - one being the line containing (x_1,y_1) and camera one perspective centre, the other being the line containing (x_2,y_2) and camera two perspective centre.

line 1: $\lambda.(x_1 i + y_1 j - f_1 k)$

line 2: $\mu.(x_2 I + y_2 J - f_2 K) + (v_1 I + v_2 J + v_3 K)$

where (i,j,k) and (I,J,K) are related by the rotation matrix (derived in Appendix A)

$$\begin{bmatrix} i \\ j \\ k \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} I \\ J \\ K \end{bmatrix}$$

Since this is an over determined solution (there are 3 equations and only 2 unknowns) a least squares technique can be implemented to solve for λ and μ. The solution is overdetermined because each image point is capable of providing a single line along which the object point may lie and so only a corresponding plane, provided by only one image coordinate, is required to obtain the true position of the scene point. Details of the solution appear in Appendix B.

The reconstruction of a scene consists of selecting an area of interest in one image and then selecting points within this area, sufficiently spaced to provide an adequate description of the scene being viewed. For example regularly spaced points forming a grid. This is followed by location of corresponding points in the second image and a computation of the world coordinates. Finally the three-dimensional coordinates are used to obtain an interpolated surface of the scene with the aid of a suitable interpolating routine.

2.1 Distortions in the Imaging System

Before any form of measurement can be made, it is necessary to calibrate the camera system geometrically, taking into account the systematic errors that may occur in the imaging system.

Image distortions are the result of deviation from a perfect central projection which is caused by both random and systematic errors. Their combined effect deteriorates the statistical significance since it is only unbiased random errors that are considered for the majority of photogrammetric adjustment procedures. The object of calibration is to single out systematic criteria of distortion. Systematic image distortions can be taken into account by the use of analytical calibration. It is possible to predict the mathematical functions that can model the effect of systematic distortion.

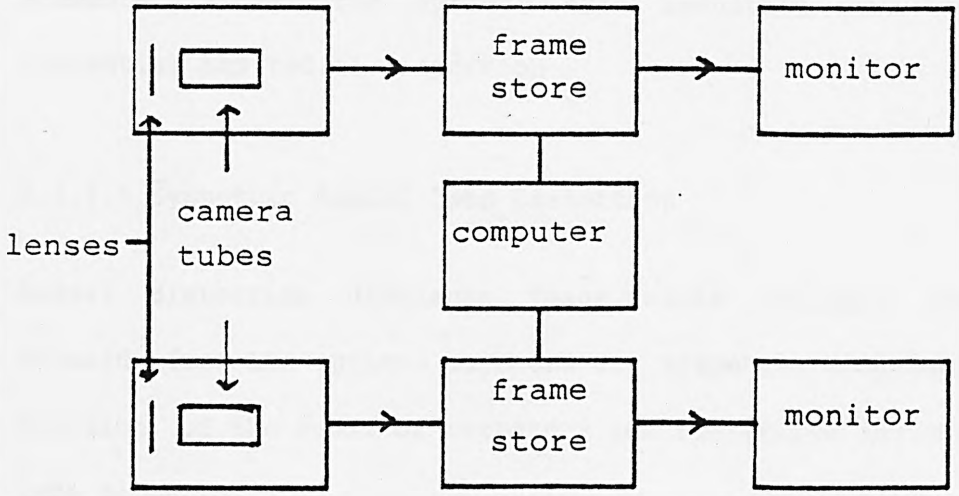


Figure 2.2 - Block Diagram of Stereo Imaging System

An analysis of the various types of systematic errors that may

occur in a stereo video imaging system was investigated by Wong (1968) and is briefly described below. Figure 2.2 illustrates the complete stereo video imaging system.

2.1.1 Distortions Due to the Lens System

There are three different kinds of lenses in a video system. These are optical lens, electric lens and magnetic lens. An optical lens is used to focus an image onto the face of the camera tube. Electric and magnetic lenses are used to focus the electron beam used for scanning the target of the camera tube.

Fortunately, the aberration and geometric distortion effects of these three kinds of lenses are completely analogous to each other. Geometric distortions produced by an optical lens system are results of two sources - imperfect lens design resulting in symmetric radial distortion and imperfect centreing of the lens elements along the optical axis resulting in asymmetric tangential and radial distortion.

2.1.1.1 Symmetric Radial Lens Distortion

Radial distortion displaces image points radially inward or outward from the optical axis and its symmetric component is a function of the radii of curvature and refractive indices of the lens elements.

The symmetric radial lens distortion, a function of the radial distance of the image point from the principal point, is best described by a polynomial. The case for an odd or even polynomial

is debatable but it has been found by various authors (Abdel-Aziz and Karara, 1974; Brown, 1972; Franke, 1966) that for non-metric cameras the distortions are best described by an odd polynomial of the form

$$\Delta r_s = k_0.r + k_1.r^3 + k_2.r^5 + k_3.r^7 + ...$$

The inclusion of the linear term $k_0.r$ generates an ill-condition because of its direct correlation to the principal distance. This is analytically explained by the fact that the parameter k_0 has the same effect as a change in the principal distance. In addition, it has been found from experimental study (Brown, 1972) that the significant parameters are k_1, k_2 and k_3 so that the equation is reduced to

$$\Delta r_s = k_1.r^3 + k_2.r^5 + k_3.r^7$$

Decomposed into its x and y axis components we have

$$\Delta x_s = x.(k_1.r^2 + k_2.r^4 + k_3.r^6)$$
$$\Delta y_s = y.(k_1.r^2 + k_2.r^4 + k_3.r^6)$$

2.1.1.2 Asymmetric Radial and Tangential Lens Distortion

The principle source of asymmetric radial and tangential lens distortion is the decentreing of the lens elements from the optical axis. This causes image displacement in various directions, commonly broken down into components along and perpendicular to the radial lines in the image plane. The systematic effect produced by a decentred lens (Conrady, 1919) is described by an equation of the form

$$\Delta x_d = (1 + p_3.r^2 + p_4.r^4 + \ldots).(p_1.(r^2 + 2.x^2) + 2.p_2.x.y)$$
$$\Delta y_d = (1 + p_3.r^2 + p_4.r^4 + \ldots).(p_2.(r^2 + 2.y^2) + 2.p_1.x.y)$$

The only significant terms in this expression (Brown, 1971) are p_1 and p_2 which reduces the equations to

$$\Delta x_d = p_1.(r^2 + 2.x^2) + 2.p_2.x.y$$
$$\Delta y_d = p_2.(r^2 + 2.y^2) + 2.p_1.x.y$$

2.1.2 Scanning Distortions at the Camera Tube

The use of magnetic scanning, in the present vidicon camera tubes, causes problems such as non-uniform field, fringe field and oscillations in the windings. With the presence of the axial focus field their effects, on the scanning pattern, are much more difficult to analyse.

One source of geometrical distortion inherent in the vidicon tubes is the non-uniform electric field set up between the target and the wall anode. If this field were perpendicular to the target, it would not have any effect on the electron beam. But this condition is satisfied only in the region close to the tube axis. At the outer edges of the target, the field makes a small angle with the tube axis and thus has a component radiating from the centre of the target. As a result of the interactions between the focusing and deflection fields and by the electric deceleration field at the camera tube, symmetric tangential distortions are caused. This form of distortion is zero at the principal point and increases with the radial distance as expressed by an even order polynomial of the form

$$\Delta r_t = q_1.r^2 + q_2.r^4 + q_3.r^6 + ...$$

decomposed into its x and y components

$$\Delta x_t = -y.(q_1.r + q_2.r^3 + q_3.r^5 + ...)$$
$$\Delta y_t = x.(q_1.r + q_2.r^3 + q_3.r^5 + ...)$$

2.1.3 Distortions Due to the Scanning Process

The most serious source of geometric distortion is due to the scanning process. In order to attain perfect scanning geometry the movement of the electron beam must follow the following pattern

a) The beam moves along each scan line with constant velocity.
b) The lines are straight.
c) The lines are equally spaced.
d) The beginnings and endings of the lines fall on lines forming the borders of the image field.
e) The lines all have the same inclination with the horizontal.

These problems are mainly present in vidicon cameras and the use of solid state cameras, where the image is pre-defined by an array of detectors and scanned digitally, is highly recommended.

2.1.4 Synchronisation Distortion

To produce a distortionless pattern the horizontal and vertical components of deflection at the camera head must be synchronised

in both frequency and phase with the video frame store.

Rigid synchronisation of the horizontal and vertical components of the electron beam motion is needed to ensure that the positions of the scan lines are identical from one frame to the next and that the electron beam always begins and ends the scanning of each frame at the same point on the tube.

Once again the use of solid state cameras enables these effects to be easily eliminated.

2.1.5 Distortions at the Transmission Channel

Geometric distortions are also caused by non-linear phase shift at the amplifiers and at the transmission channel. Due to non-linear phase shift, signals of high frequency arrive at the the receiver ahead of low frequency signals resulting in a shifting of the finer picture details with respect to the coarser elements. This kind of distortion can be minimised by the use of equalisers to provide as near a constant frequency response, within the transmission channel, as possible. With significant non-linear phase distortion the image quality can be rendered entirely useless for mensuration purposes.

2.1.6 Photometric Distortion

Photometric distortion is caused by non-linear transfer characteristics of the input face plate illumination to video voltage. If a camera views a 'test chart' which is a series of 'chips' of reflectance values such that the light reflected from

each chip is a linear increase on that from the adjacent chip, then the video signal so produced should be a series of linearly equal steps in the waveform from black level to white level.

Vidicon tubes have a non-linear transfer function such that the higher light levels do not produce the same change in voltage as do lower light levels and this is referred to as 'white crushing'. This non-linearity can only be overcome by calibration of the camera tube. An important factor for the algorithm to be described in chapter 5 is to have minimal photometric distortion and therefore a solid state camera is highly recommended since it has a linear transfer function.

2.2 Calibration of Cameras

Calibration of a camera system consists of determining the individual parameters that define the interior and exterior geometry of a camera. For this purpose a suitable reference system has to be established in a camera within which the properties of the projective geometry can be described.

A universally accepted reference system is based on both the location of the 'principal point' and the 'principal distance' (or equivalent focal length) of the camera, frequently referred to as basic interior orientation parameters (see figure 2.3). The principal point is defined as the point at the base of the perpendicular from the perspective centre. The principal distance is the distance from the perspective centre to the principal point.

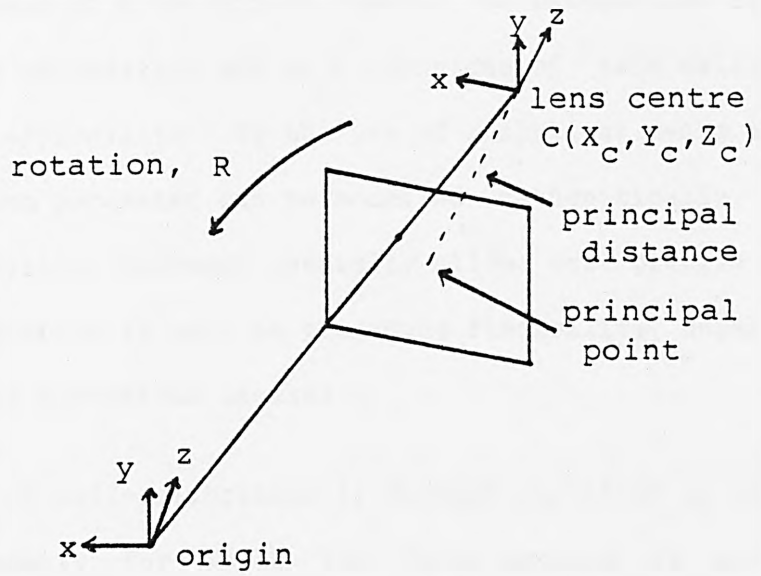


Figure 2.3 Collinearity Condition

For the majority of photogrammetric applications the interior projective geometry of a camera is mathematically represented as a central projection, any deviation is referred to as distortion. Causes of distortion consist of the distortions previously described as well as affinity errors such as non-perpendicularity and scale changes along axes. Therefore the calibration procedure involves the determination of basic interior and exterior orientation as well as distortion parameters for each camera in the stereo video system.

Cameras used for photogrammetric measurements can be divided into two categories - non-metric cameras and metric cameras. The difference between the two types being the interior orientation stability of the metric cameras. While for non-metric cameras partial or no information is available regarding calibration and its stability, metric cameras have well defined calibration data which can be assumed to be stable over a period of time.

In the case of a television camera, no information is available regarding calibration and so a technique of 'self calibration' is the most appropriate. By the use of analytical means any unknown calibration parameter can be modelled mathematically. Therefore the analytical approach generally allows more precise evaluation for calibration as well as providing flexibility, especially when applied to non-metric cameras.

The use of self-calibration is further justified in close range photogrammetry for which the focus setting is not a fixed parameter (as compared to aerial photogrammetry). Short object distances usually require finite and variable settings from inspection of one scene to another and so considering the fact that large variations, in both lens distortion and interior orientation, occur when focussing at different, finite distances, analytical self-calibration is best applied to close range photography.

2.2.1 Self-Calibration Technique Applied to Video Cameras

The basic mapping from the scene coordinates to the image coordinates is provided by the 'collinearity condition' (Wong, 1980). The collinearity condition states that any point in a scene, the camera perspective centre and the image point corresponding to the scene point all lie along the same line. The collinearity condition can be expressed as

$$\begin{bmatrix} x \\ y \\ -f \end{bmatrix} = \beta . \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} . \begin{bmatrix} X - X_c \\ Y - Y_c \\ Z - Z_c \end{bmatrix}$$

This may be reduced to provide the following collinearity equations

$$x = -f.(M_1 / M_3)$$

and

$$y = -f.(M_2 / M_3)$$

where $M_i = m_{i1}.(X - X_c) + m_{i2}.(Y - Y_c) + m_{i3}.(Z - Z_c) \quad i = 1,2,3$

This assumes that the comparator coordinate system has its origin at the principal point with orthogonal axes and no scale changes along the axes.

2.2.1.1 Image-to-Comparator Coordinate Mapping

In the case of a vidicon camera, due to the process of scanning the target, the coordinate axes have both a shift of origin and a lack of orthogonality (figure 2.4) in addition to any differential scale deformation.

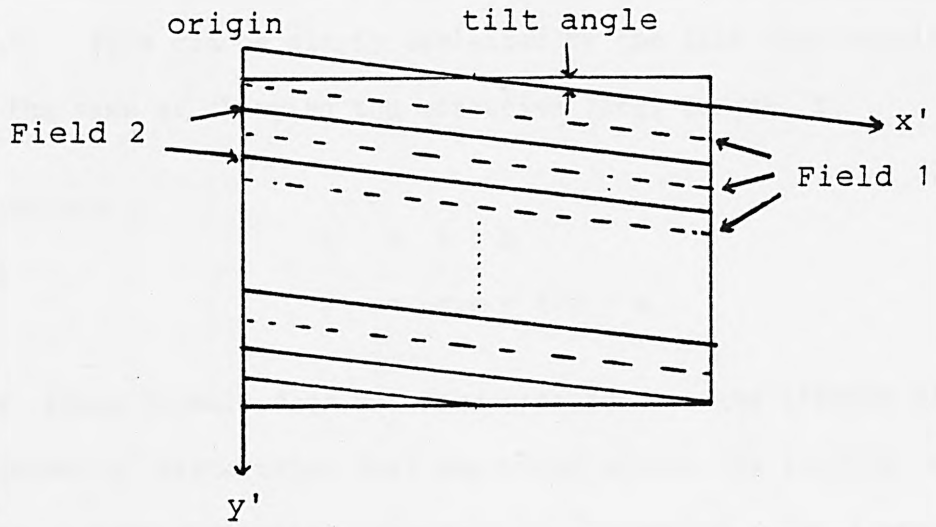


Figure 2.4 - Trace of Television Scan Lines

The transformation from image coordinates to comparator coordinates is described as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b \\ e \end{bmatrix}$$

where the parameter 'c' accounts for the non-orthogonality of the axes due to the scanning system, parameters 'a' and 'd' account for differential scaling and parameters 'b' and 'e' account for the shift of origin.

This can be simplified into

$$x' = a.x + b$$

and

$$y' = c.x + d.y + e$$

In the case of a video camera the focal length is an unknown quantity, so in order to avoid an ill-condition occurring in the solution of the equations and reduce the number of parameters, without losing generality, the parameter 'a' can be reduced to unity. This can be simply explained by the fact that varying 'a' is the same as changing the effective focal length, f.

therefore

$$x' = x + b$$

and

$$y' = c.x + d.y + e$$

The above formula does not take into account the effects of the systematic distortions that may occur within the imaging system such as lens distortion and scanning distortion. The formulation of the distortion equations is

$$\Delta x = \Delta x_s + \Delta x_d + \Delta x_t$$

and $$\Delta y = \Delta y_s + \Delta y_d + \Delta y_t$$

For the purposes of evaluating the exact area registration algorithm, a CCD (charge coupled device) camera was used and so the distortion correction due to the scanning at the camera tube (section 2.1.2) was not used in the derived solution.

Hence $$\Delta x = x.(k_1.r^2 + k_2.r^4 + k_3.r^6) + p_1.(r^2 + 2.x^2) + 2.p_2.x.y$$

and $$\Delta y = y.(k_1.r^2 + k_2.r^4 + k_3.r^6) + p_2.(r^2 + 2.y^2) + 2.p_1.x.y$$

where $$r^2 = x^2 + y^2$$

This gives the final form of the transformation equations, after including the distortion equations, as

$$x' = (x + \Delta x) + b$$

and

$$y' = c.(x + \Delta x) + d.(y + \Delta y) + e$$

Expansion of these equations yields two functions, F and G which represent the difference between the computed comparator coordinates, (x',y') and the measured comparator coordinates, (x_m',y_m') and these equations may then be solved using a least squares technique

$$F = x' - x_m' = 0$$

and

$$G = y' - y_m' = 0$$

The self-calibration method consists of using control points whose world coordinates and comparator coordinates are known. It is essential to provide a set of well distributed points in order

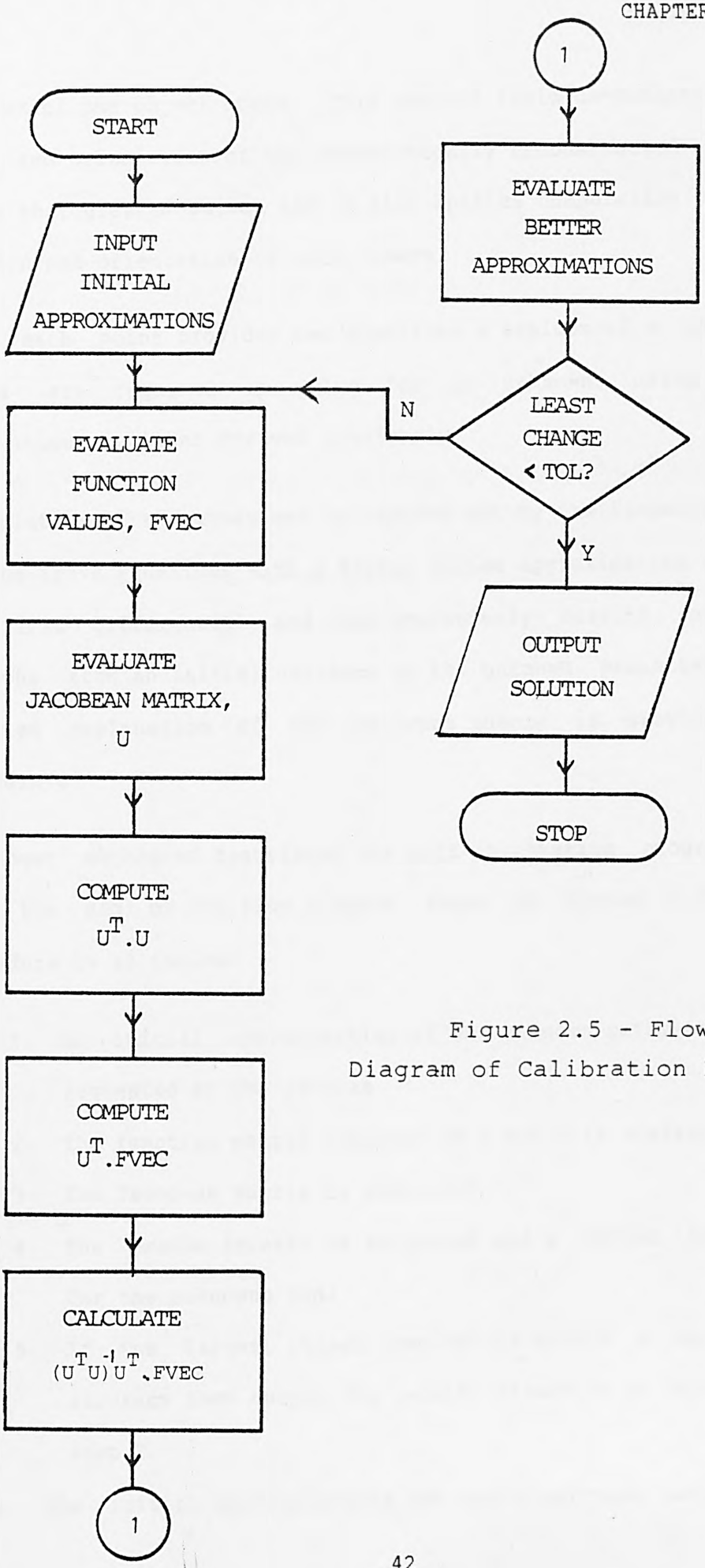


Figure 2.5 - Flow Diagram of Calibration Program

to control the object space. This control field determines the scale and orientation of the mathematically reconstructed model of the photographed object and it also enables computation of the location and orientation of each camera.

Since each point provides two equations a minimum of n control points are required to solve for 2n unknowns using the observation equations derived previously.

The solution of the equations is carried out by the linearisation of the above equations with a Taylor series approximation up to the first differential and then iteratively solving for the unknowns from an initial estimate of the unknown parameters. A detailed explanation of the relevant theory is provided in Appendix C.

The best method of describing the self-calibration program is with the aid of the flow diagram shown in figure 2.5. The procedure is as follows

1: An initial approximation of the unknown parameters is requested by the program.

2: The function matrix composed of F and G is evaluated.

3: The Jacobean matrix is evaluated.

4: The pseudo inverse is evaluated and a better estimate for the unknowns made.

5: If the largest change computed is within a specified accuracy then output the results otherwise go back to step 2.

Note: the initial approximations for the distortion parameters

k_1, k_2, k_3, p_1, p_2 and c are all set to zero.

A full derivation of the differential parameters in the Jacobean matrix can be found in Appendix D and the program listing, in the 'C' programming language, can be found in Appendix E.

2.2.1.2 Comparator-to-Image Coordinate Mapping

For purposes that will become apparent in chapter 5 it was also necessary to derive an equation which mapped comparator coordinates onto image coordinates. This can be solved after computation of the image-to-comparator coordinate transform has been determined from the above solution. From the previous solution we know that the comparator coordinates are given by

$$x' = (x + \Delta x) + b$$

and

$$y' = c.(x + \Delta x) + d.(y + \Delta y) + e$$

where Δx and Δy are functions of x and y.

Without affecting the accuracy of the model we can just as easily write the image coordinates, in terms of the comparator coordinates, as $(x + \Delta x', y + \Delta y')$ where $\Delta x'$ and $\Delta y'$ are functions of x and y which are in turn represented by

$$x = x' + b'$$

and

$$y = c'.x' + d'.y' + e'$$

where b',c',d' and e' allow for scale changes along axes, non-orthogonality of the axes and shift of axes.

The solution of these two equations is performed in a similar manner to that described for the absolute orientation, except in

this case, the control point image coordinates are used to compute the control point comparator coordinates and then the iterative method of solution is carried out for the reverse transform, from comparator-to-image coordinates.

The initial approximations for this, reverse transform are calculated from the previous results as

$$b' = -b, \quad d' = 1/d \quad \text{and} \quad e' = -e/d$$

The values for the distortion parameters $k_1', k_2', k_3', p_1', p_2'$ and c' are set to zero and the Jacobean evaluations can be found in Appendix D.

2.3 Alternative Approach to Solution

In order to reduce the number of control points, for which the world coordinates are required, an alternative method of solution consists of computing the relative orientation between the cameras and then solving for the orientation from the camera coordinate system to the world coordinate system separately as a two stage process.

The diagram in figure 2.6 illustrates the orientation of a pair of camera stations. Now to determine the relationship between the two cameras we need to link the world coordinates with the image coordinates. The coplanar model (Wong, 1980) states that any world point (X,Y,Z) and its corresponding image coordinates (x_1,y_1) and (x_2,y_2) are coplanar. Therefore, the direction vectors V_1, V_2 and V must satisfy the condition

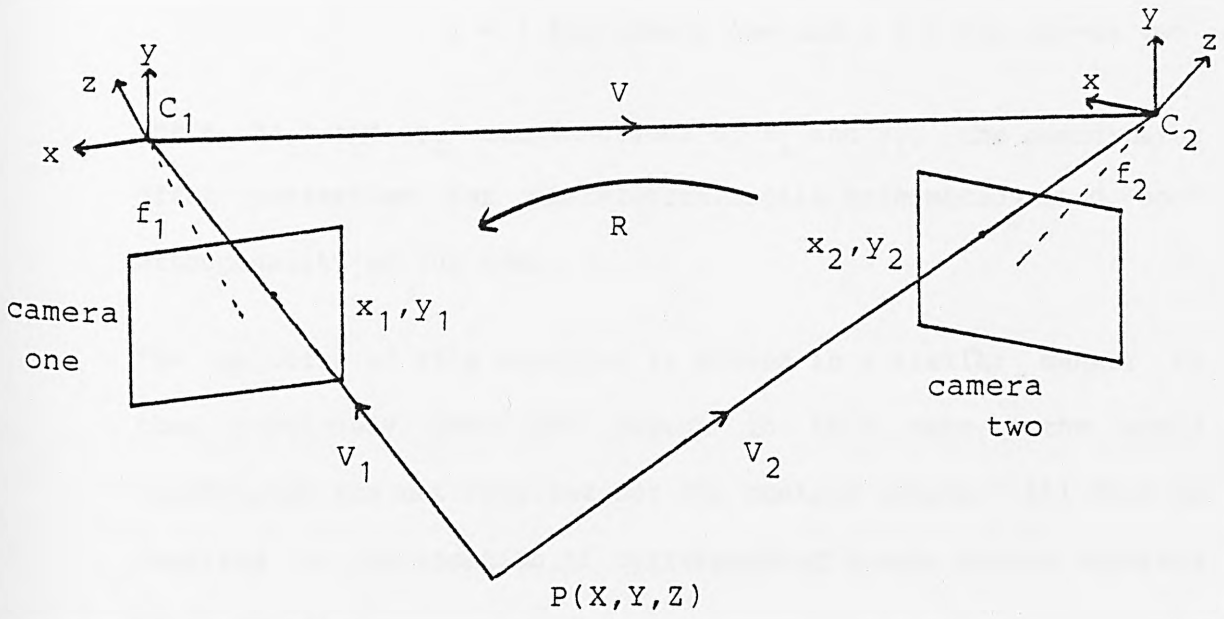


Figure 2.6 - Coplanarity Condition

$$(V_1 \times V_2).V = 0$$

that is

$$\det. \begin{bmatrix} x_1.r_{11} + y_1.r_{21} - f_1.r_{31} & x_2 & v_1 \\ x_1.r_{12} + y_1.r_{22} - f_1.r_{32} & y_2 & v_2 \\ x_1.r_{13} + y_1.r_{23} - f_1.r_{33} & -f_2 & v_3 \end{bmatrix} = 0$$

Where the first column represents direction vector V_1 measured relative to the camera two coordinate system.

Expansion of this determinant yields an equation linking the image coordinates of each camera and the relative orientation between the cameras. To avoid an ill-condition occurring in the solution a scale must be introduced such as

$$v_1^2 + v_2^2 + v_3^2 = 1$$

The transformation from the comparator-to-image coordinates is

$$(x_i + \Delta x_i', y_i + \Delta y_i')$$

i = 1 for camera one and i = 2 for camera two.

where $\Delta x_i'$ and $\Delta y_i'$ are functions of x_i and y_i, the coordinates after correction for differential scale deformation and non-orthogonality of the axes.

The solution of this equation is solved in a similar manner to that previously described except in this case, the world coordinates are not required for the control points. All that is required is the location of corresponding stereo points obtained by an operator.

For the above derived equations the minimum number of points required is 22 but for purposes of accuracy, a least squares method with more points should be used.

The solution for the absolute orientation is carried out as described in section 2.2.1 but in this case there are considerably fewer unknowns. For camera one the rotation and translation relative to the world coordinate system are unknown and for camera two the translation distance between the cameras is unknown although the direction has been determined. This is a total of 7 unknown parameters and so for an absolute orientation solution only two control points are required (each camera provides a pair of observation equations for each control point).

3.0 Stereo Video Measurement System

Stereopsis is the term applied to man's ability to fuse two slightly different monocular views of a scene into a single three-dimensional image. A stereoscopic viewer is a display system which presents two dissimilar pictures to the eyes so that stereopsis produces the impression of a three-dimensional scene.

Stereoscopic viewers for displaying photographs, in three-dimensions, are fairly simple in operation. Two photographs of an object are taken from slightly different viewpoints and by placing them in the focal plane of a binocular viewer each eye can see only one of the two pictures. Provided that the separation of the photographs has been set correctly, the brain can fuse them into one, apparently three-dimensional, scene.

3.1 The City University Stereo Video System

The City University stereoscopic system (plate 3.1) uses two monitors to display the pictures derived from two television cameras. Since the width of the television screens is larger than the eye separation distance, the monitors cannot be physically placed directly in front of each eye. Optical techniques have been used to solve this design problem. In addition the magnification and rotation of the image perceived by each eye may be independently adjusted for comfortable viewing. The separation distance between the two eyepieces can also be varied to suit the eye separation distance of the observer.

Some of the following information has been extracted from a paper

by Lindsey et al. (1985) with further details included for the design and construction of the marker generator/tracker.

Plate 3.1 - Photograph of the Stereo Video System

The stereo video system (figure 3.1) is designed such that two 7in. monitors are observed simultaneously using an optical stereo viewer. Measurement is made possible by two methods. One uses encoders mounted on the axes of the instrument, used to physically measure the movement of light marks introduced into the viewing optics. The other is by an electronic marker generator designed to project two electronic light spots independently onto each monitor image. The fusion of the individual light marks, by an operator, into one three-dimensional point provides the conventional floating mark.

The method of measurement from the stereo images is similar to normal photogrammetric mensuration, where two independent optical marks are introduced into the viewing system and the observer

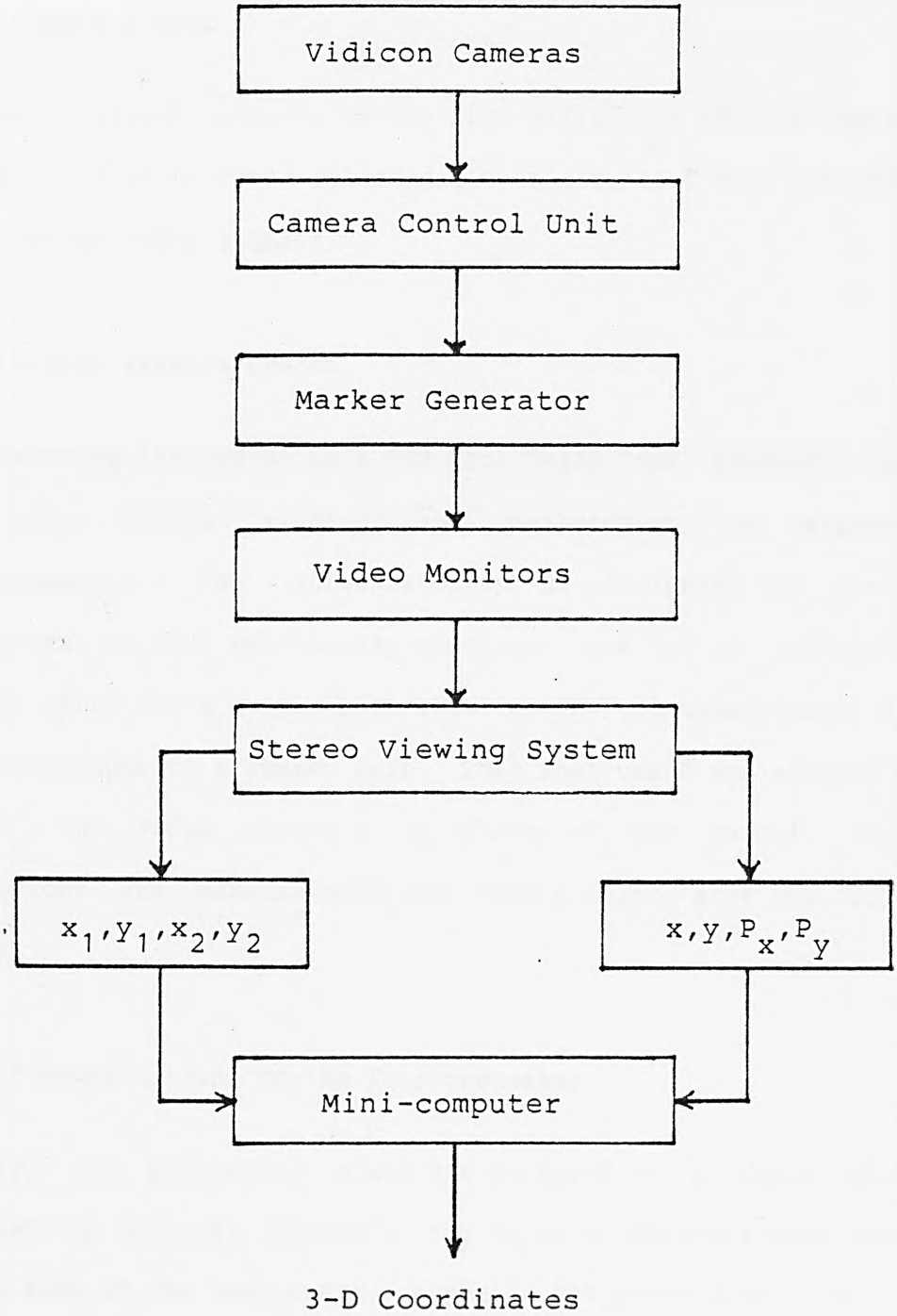


Figure 3.1 - Block Diagram of Complete Stereo Video System

fuses the marks to form a floating mark. This mark is then placed on the desired point, in the stereo model, to obtain the image coordinates of that point and hence its world coordinates.

3.1.1 Camera System

The camera system consists of two high definition vidicon cameras linked via a multi-cored camera cable to a control unit providing synchronised video signals.

3.1.2 Stereo Viewing System

The measuring instrument is a VEB Carl Zeiss Jena Interpretoskop used for stereo viewing and measurement in standard photogrammetry. The Interpretoskop is designed for photo-Interpretation with two viewing stations - one for an instructor and the other for a pupil, both able to view, simultaneously, the two photographs of a stereo pair. This instrument was adopted to accept two video monitors in place of the normal still photographs and measurements were then possible from the video images.

3.1.2.1 Modifications to the Interpretoskop

Firstly, the photostage plate was replaced by a frame which accepted the monitors. Secondly, two types of encoders were fixed to the body of the instrument, opposite the travelling rails, as axes. The encoders are two linear and one rotary such that one linear encoder is parallel to the x-axis and the other parallel

to the y-axis of the instrument. The x-axis encoder is used for both the left and right hand objective system. The rotary encoder was attached to the handle of the right hand objective. These encoders, when linked to a MDR-1 (Microprocessor Display and Recording system) allow four coordinate measurements to be continuously displayed on 6-digit counters. The MDR-1 is in turn linked to a mini-computer for data recording and processing.

3.1.3 The Floating Light Mark

In the optical system of the Interpretoskop, two bright dots of 40um diameter are incorporated to enable measurements to be made from the two video monitor images.

In order to achieve accurate measurement, the distortions caused by the complete imaging system must be modelled and corrected. To remove the instrumental error and the distortion in the monitor tubes it was decided to introduce an electronically generated mark in place of the optical mark. The electronic marks are bright spots superimposed onto the video signals before they arrive at the monitors.

3.2 Electronic Marker Generator/Tracker

The electronic marker generator is designed to be such that a 100ns pulse could be placed onto and along any line of the video signal. In addition the pulse, seen as a bright spot, can be manoeuvrable within the visible area of the image. Also, at the press of a button the current coordinates of the spot can be

transmitted, via an RS232 link, to a mini-computer.

An optional feature of the marker generator is to be able to track a bright target in real time and transmit its coordinates as it moves from frame to frame. This feature enables measurements to be made of the flight of a ball for example.

It was decided that although the marker generator/tracker could be built completely using hardware logic, it would be better to build a system based around a micro-processor. This would allow flexibility into the system and if required, at a later stage it would be possible to include the calibration computations into the system thereby providing a stand alone system capable of making simple measurements.

3.2.1 Basic Structure of Hardware

The basic structure of the marker generator/tracker is composed of an 8-bit micro-computer, hardware counter, shift register store, video sync separator and associated control circuits as illustrated in figure 3.2. Full circuit diagrams may be found in Appendix F.

3.2.1.1 Video Sync Separator

In order to process the video signal and superimpose a light spot onto the video image, it was necessary to extract the line and field sync signals from the composite video signal. The composite video signal is a video signal which has synchronising pulses added, the ratio of amplitude of video to sync can vary but in

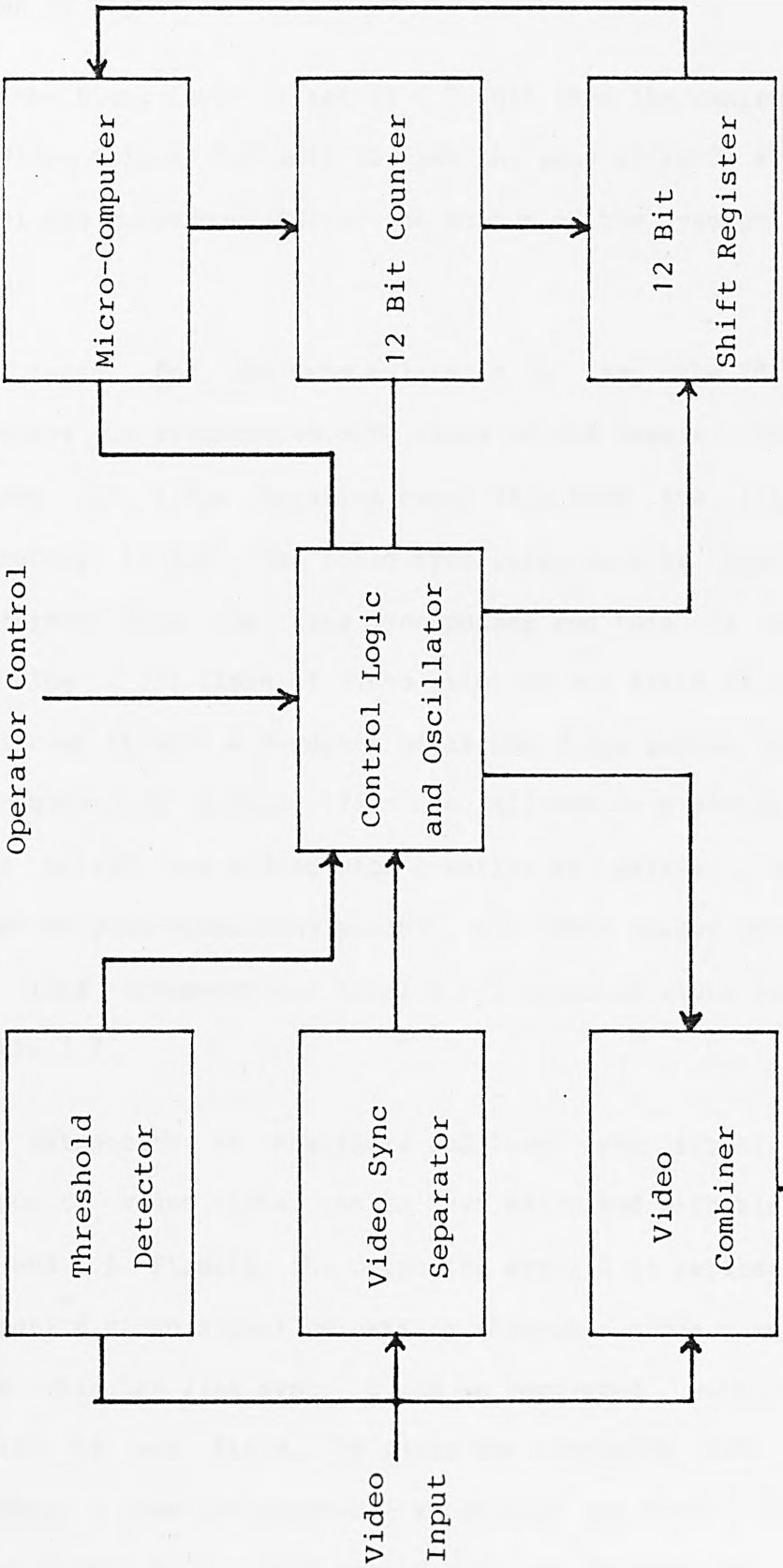


Figure 3.2 - Block Diagram of Marker Generator/Tracker

the standard 1 volt peak-to-peak composite video, it is 70:30 as shown in figure 3.3.

If the black level is set at 0.3 volt then the maximum amplitude of video will be 0.7 volt so that the peak white is at the 1 volt level and correspondingly, the bottom of the sync pulses is at 0 volt.

The reason for the sync pulses is to keep the display unit timebase in synchronism with those of the camera. The line sync pulses, of 4.7μs duration every 64μs keep the line timebase correctly locked. The field sync pulse must be distinguishably different from the line sync pulses and this is achieved by removing 2 1/2 lines of video prior to the start of a field and replacing it with a sequence of narrow 2.3μs pulses, known as pre-equalising pulses. This is followed by a series of broad, 27μs pulses and ending with a series of narrow 2.3μs pulses, known as post-equalising pulses. All these pulses occur at twice the line frequency and total 7 1/2 lines of video as shown in figure 3.3.

The extraction of the field and line sync signals from the composite video signal can be best described with aid of figures 3.4 and 3.5. Firstly, the composite sync, A is separated from the composite video signal by passing through a diode clamp circuit. From this the line sync, E can be recovered, within a maximum period of one field, by using the composite sync signal to trigger a non-retriggerable monostable set with a pulse width greater than half a line period but less than one line period. If

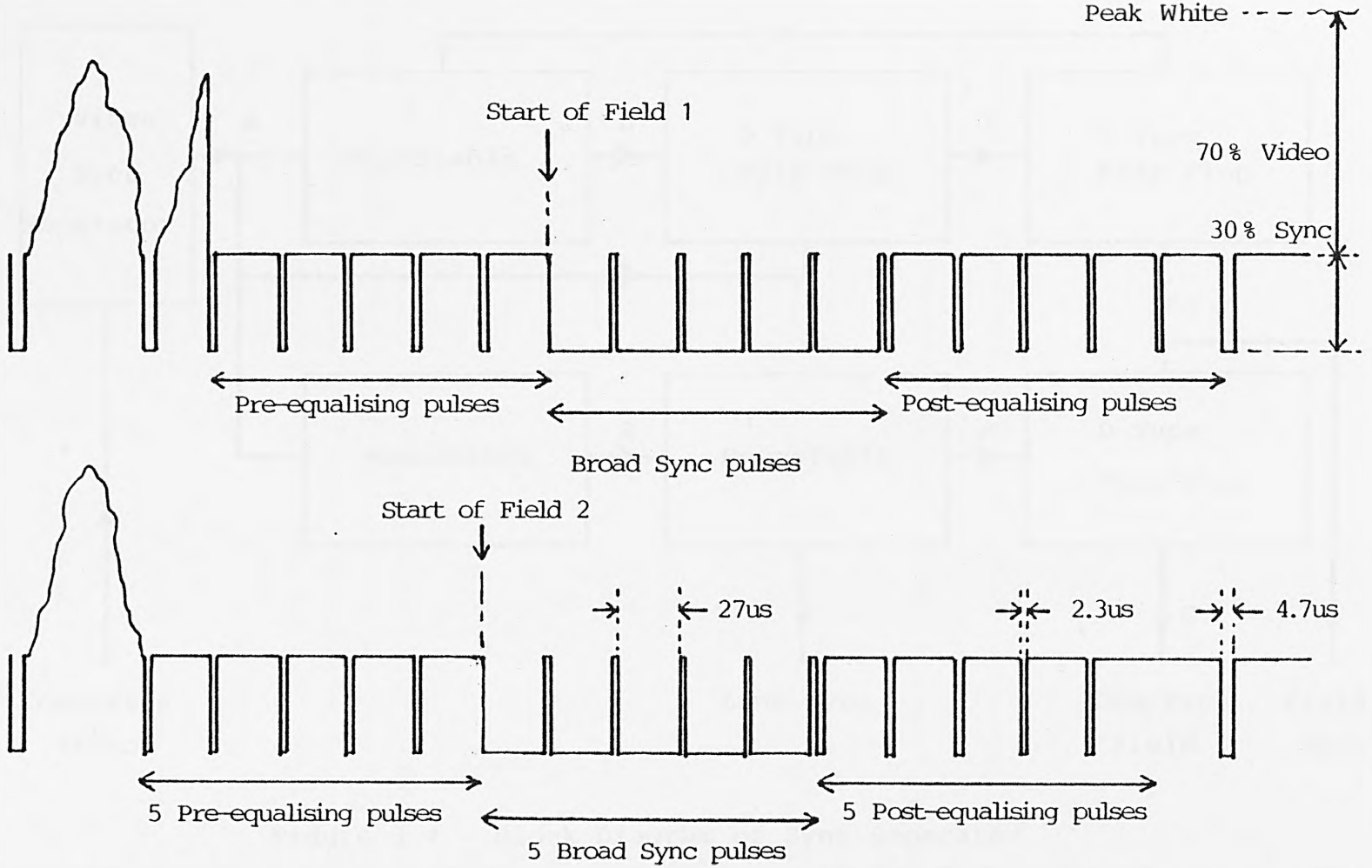


Figure 3.3 - Typical Representation of a CCIR Video Signal

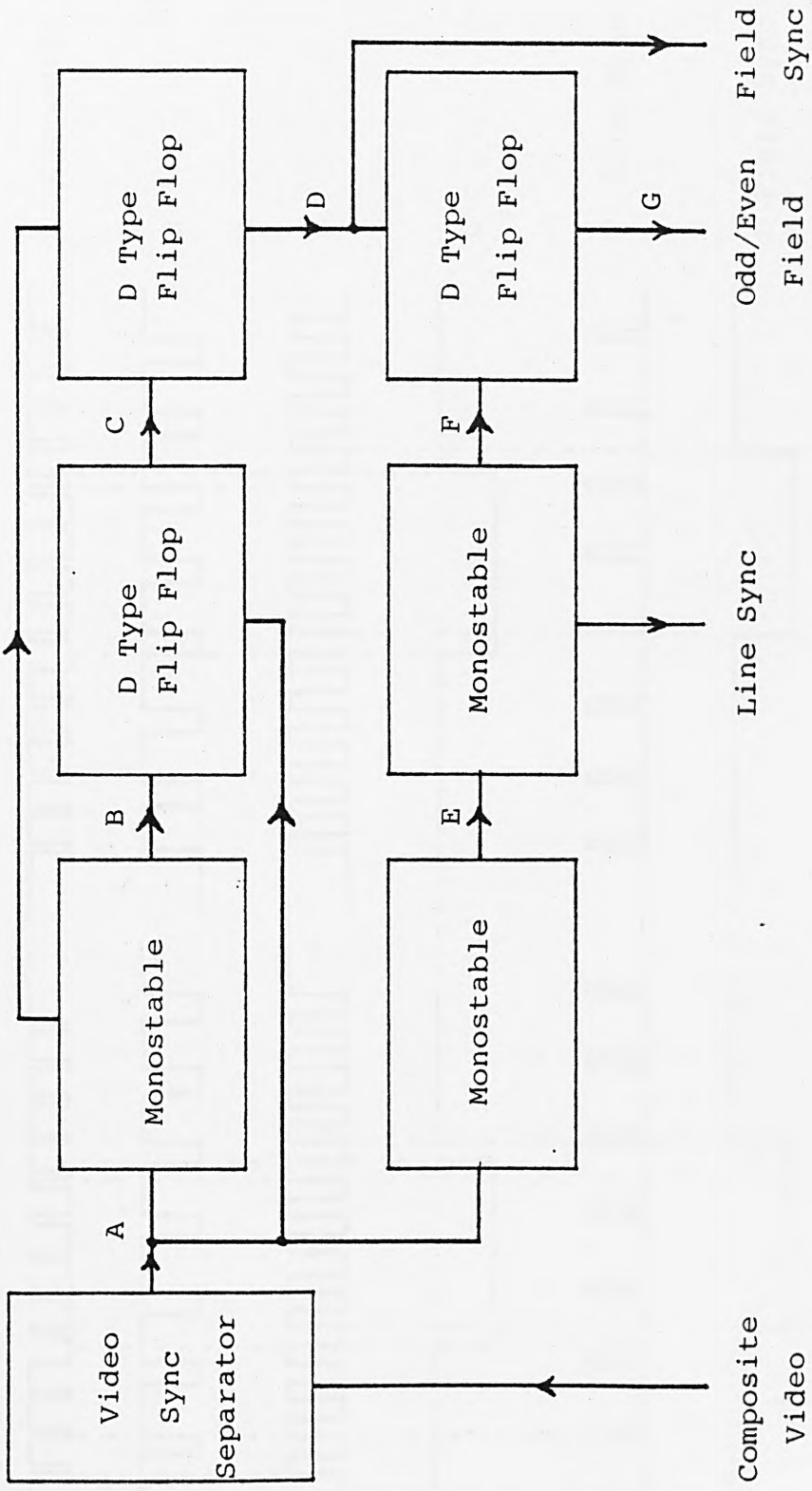


Figure 3.4 - Block Diagram of Sync Separator

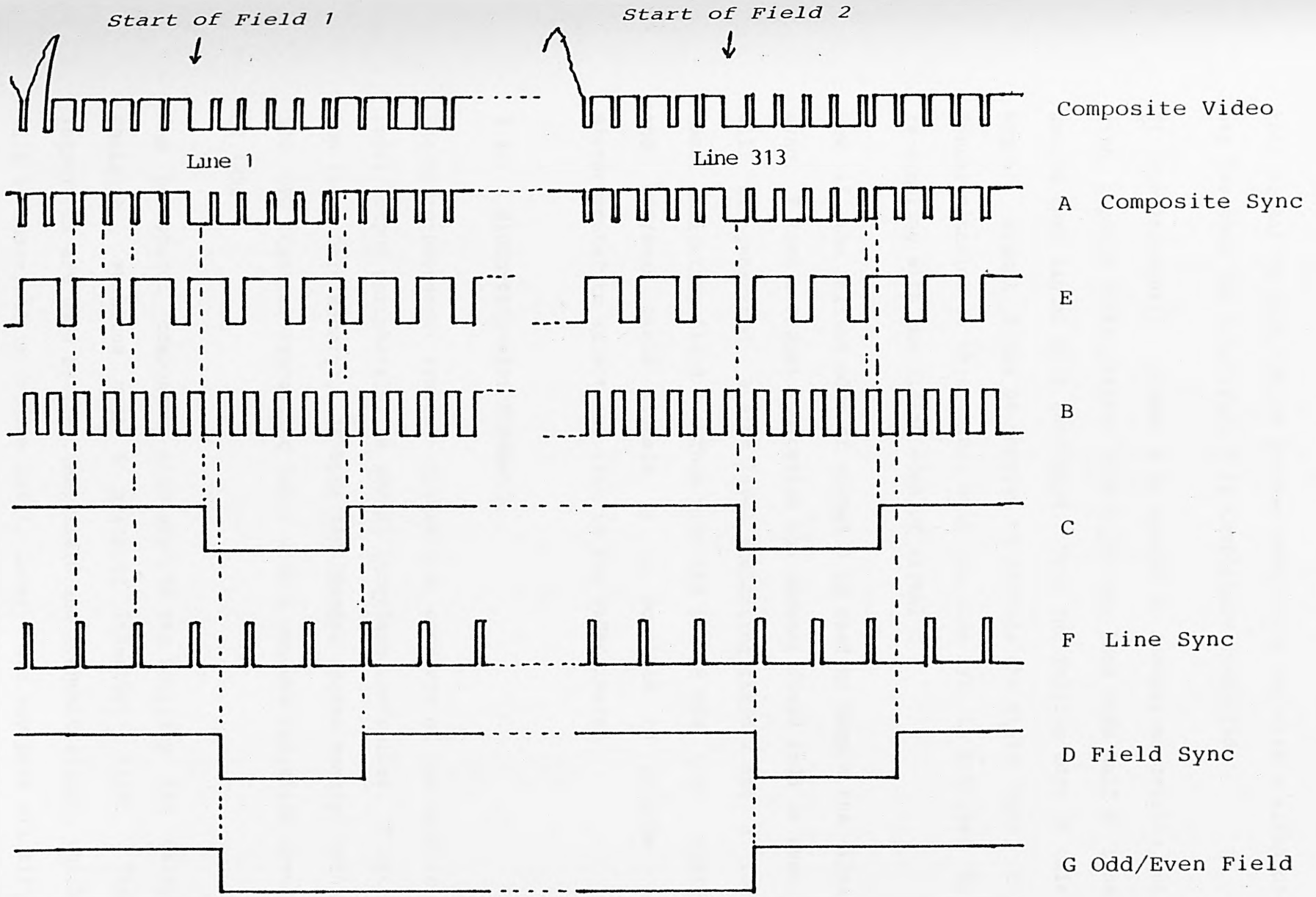


Figure 3.5 - Waveforms Associated with Sync Separator Diagram

this signal is then fed to another monostable set with a width of say 5μs then the line sync, F is completely recovered.

If simultaneously, signal A is passed to another monostable set with a pulse width greater than 2.3μs but less than half a line period then signal B is generated. Upon the falling edge of this signal, signal A can be sampled to provide the field sync, C. Synchronisation of this signal with the line sync is achieved by re-sampling with the rising edge of signal B.

Now if the falling edge of signal D is used to sample the line sync, F then a signal indicating the current field (odd or even) will be generated. A high level indicating field 2 and a low level indicating field 1. Thus with the aid of the line, field and odd/even field signals it is possible to program a microcomputer to select any line in the video image.

3.2.1.2 Micro-processor Schematic

The micro-processor system, figure 3.6, consists of two parallel input/output peripherals, a serial interface controller, 2K byte ROM (Read Only Memory), 1K byte RAM (Random Access Memory) and a Z80 CPU (Central Processing Unit) with a maximum operating speed of 4MHz.

The peripheral components also provide the facility for daisy chaining, enabling upto 6 priority interrupt lines. The interrupts are used by the line, field and transmit signals which will be described in further detail later. The top most priority level is that for the line sync followed by field sync and

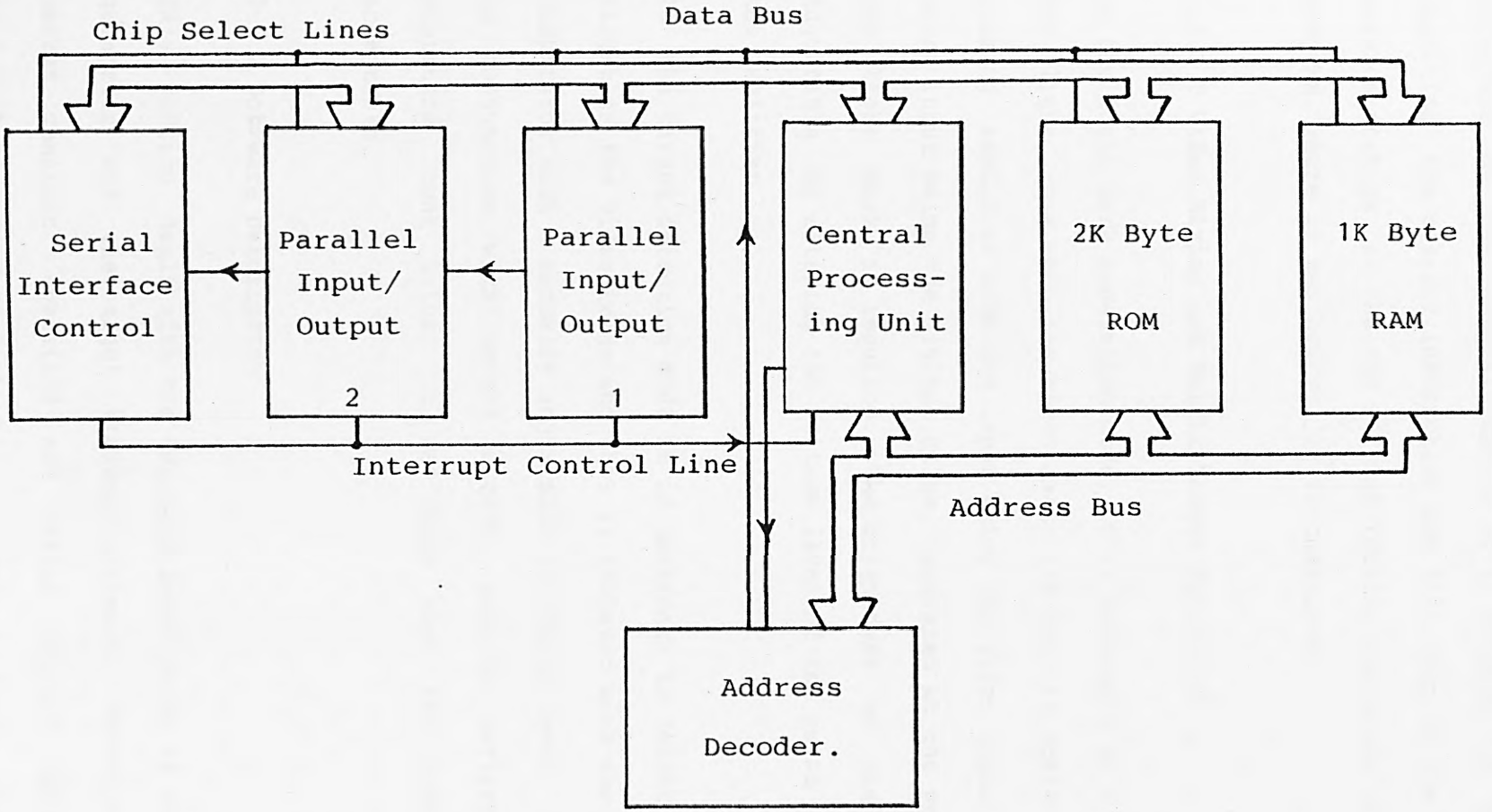


Figure 3.6 - Block Diagram of Micro-processor Layout

finally by the transmit interrupt line. The effect of an interrupt on the general running of a program is to finish executing the current instruction and then jump to the interrupt service routine. At the end of this routine execution of the main program, where it was halted, is re-commenced.

3.2.1.3 Video Mixing and Bright Target Detection

In the light mark generation mode, it is necessary to superimpose the light mark onto the video image and this is achieved by a summing amplifier with one input being the video signal and the other input being the digital pulse, generated at the point where the light mark is required. The brightness of the mark is adjustable by altering the voltage level of the pulse applied to the amplifier.

In the target detection mode it is necessary to detect a bright target on the video image and this is achieved with the aid of a comparator with a manually adjustable threshold level. When used in conjunction with bright targets, such as battery powered miniature light bulbs, it was found that the results were acceptable.

3.2.2 Software Description

This section deals with the detailed description of the marker generator and the target tracker software. Function between marker generator operation and target tracker operation is selectable with a switch which controls a logic signal, the

signal being low for marker operation and high for tracker operation. Full program listings, in assembler, are given in Appendix G.

3.2.2.1 Operation of Marker Generator

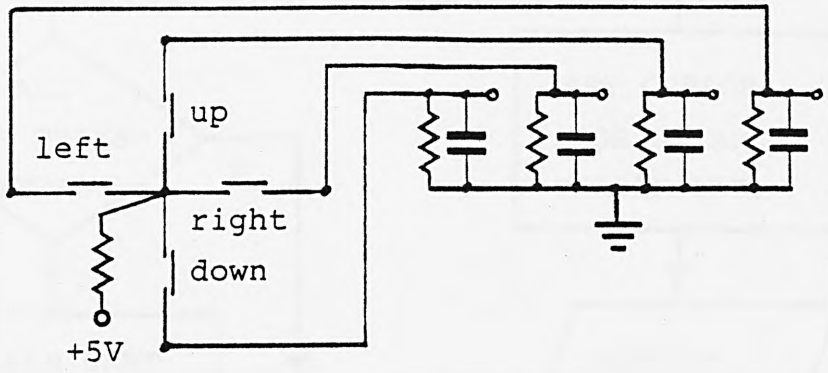


Figure 3.7 - Joystick Configuration

The marker operation is facilitated by the use of a switched joystick and a transmit button. The joystick is configured to provide four digital signals, as shown in figure 3.7, one for each of four directions - left, right, up or down. The movement of the joystick in the appropriate direction would cause that line to become high as compared to its normal low output.

When the micro-processor is reset the computer automatically begins execution of the marker generator program. The explanation of the individual routines is best explained with the aid of flow diagrams shown in figures 3.8 through 3.15. The overall program structure is built around four interrupt routines - line interrupt, field interrupt, transmit interrupt and serial

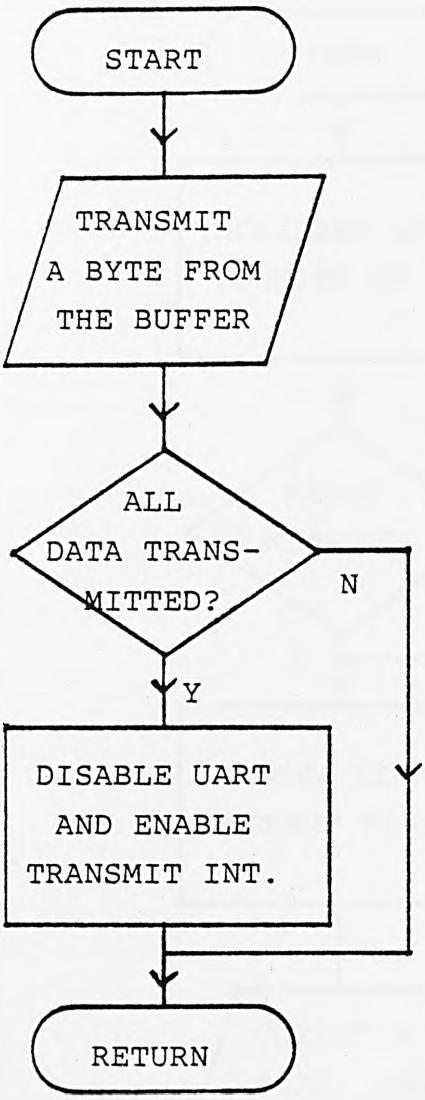


Figure 3.15 - Marker Generator /Tracker Serial Interface Interrupt

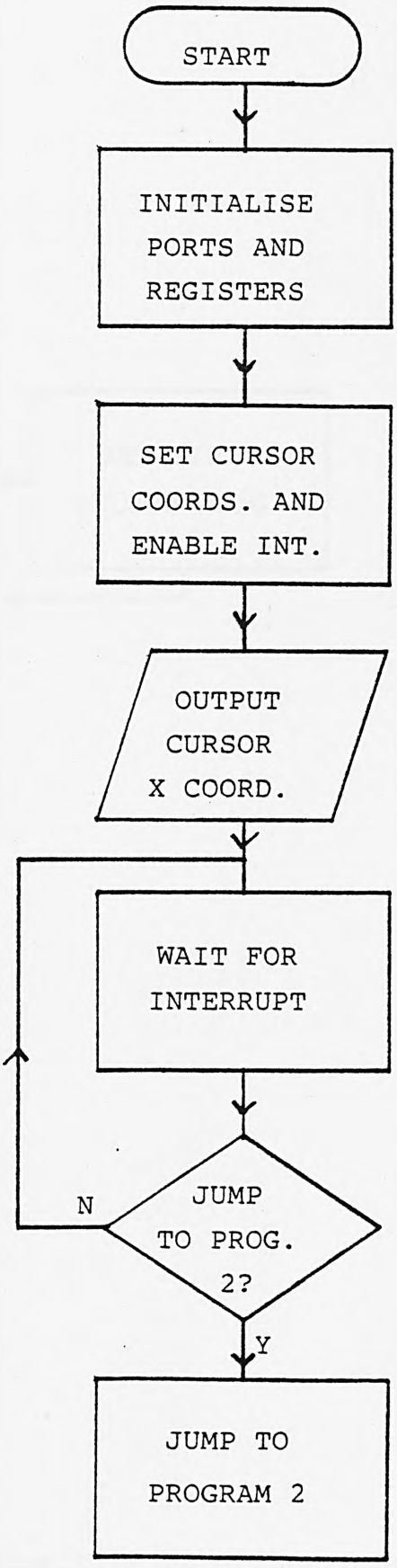


Figure 3.8 - Marker Generator Control Program

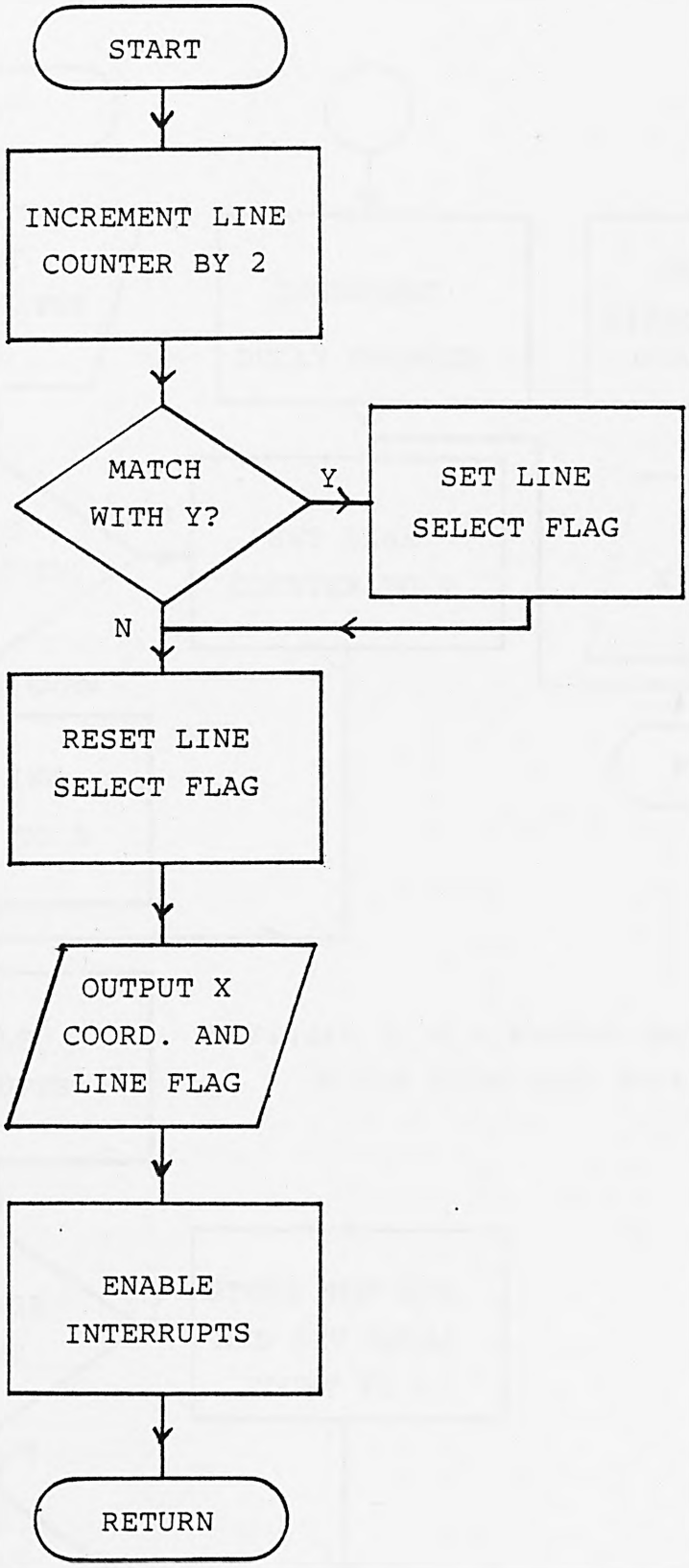


Figure 3.9 - Marker Generator Line Interrupt Routine

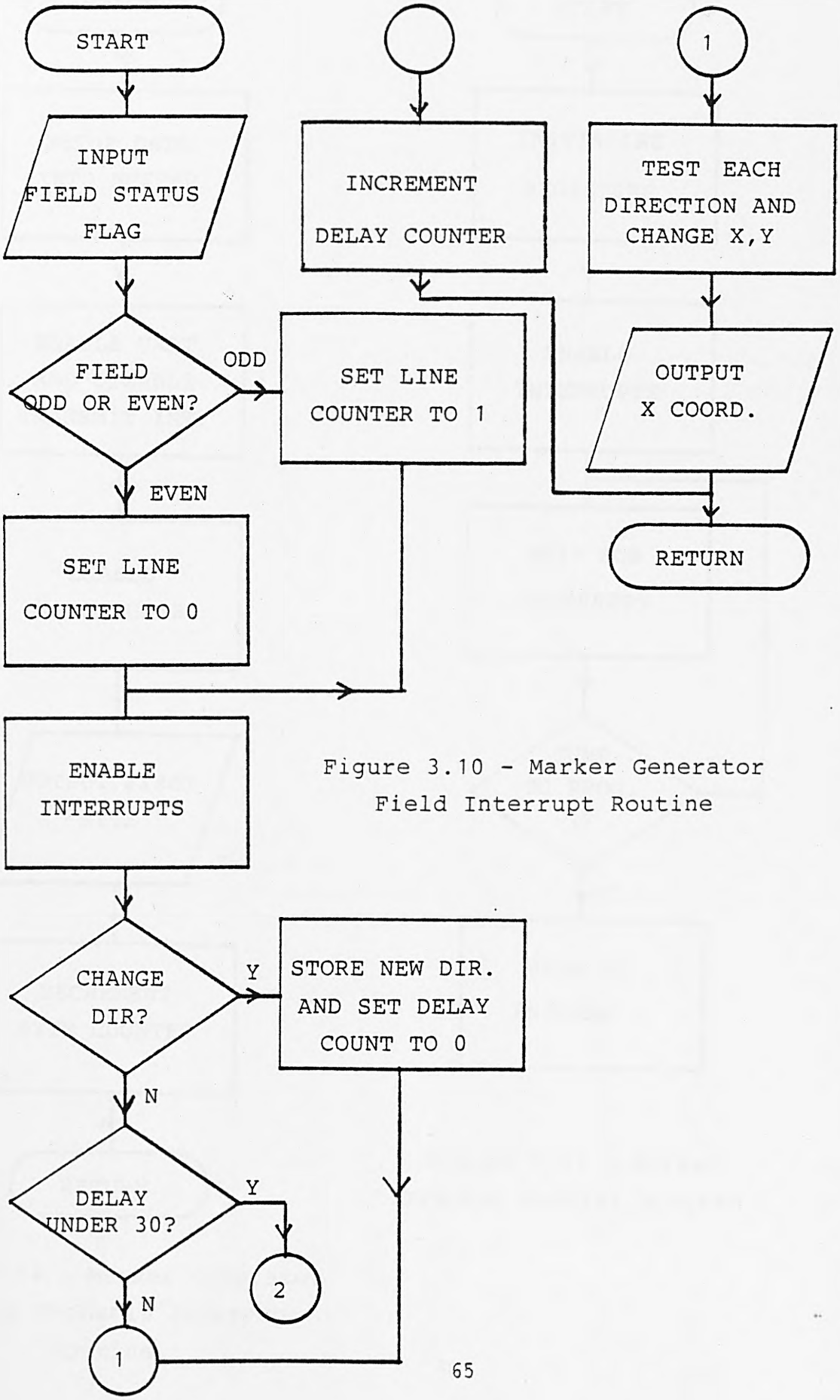


Figure 3.10 - Marker Generator Field Interrupt Routine

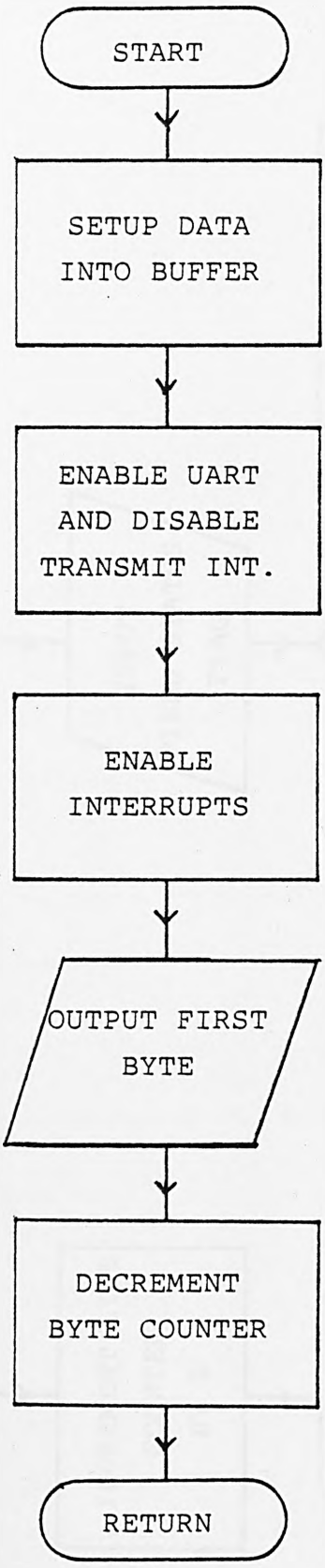


Figure 3.14 - Marker Generator /Tracker Transmit Interrupt Routine

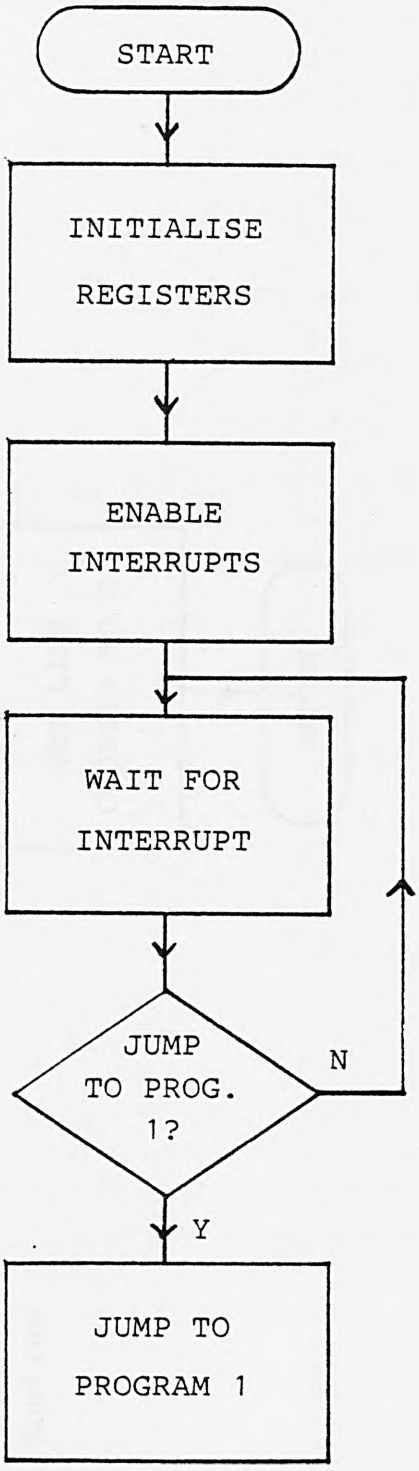


Figure 3.11 - Marker Tracker Control Program

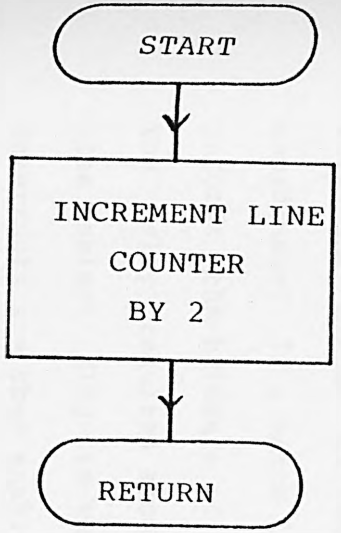


Figure 3.12 - Marker Tracker Line Interrupt Routine

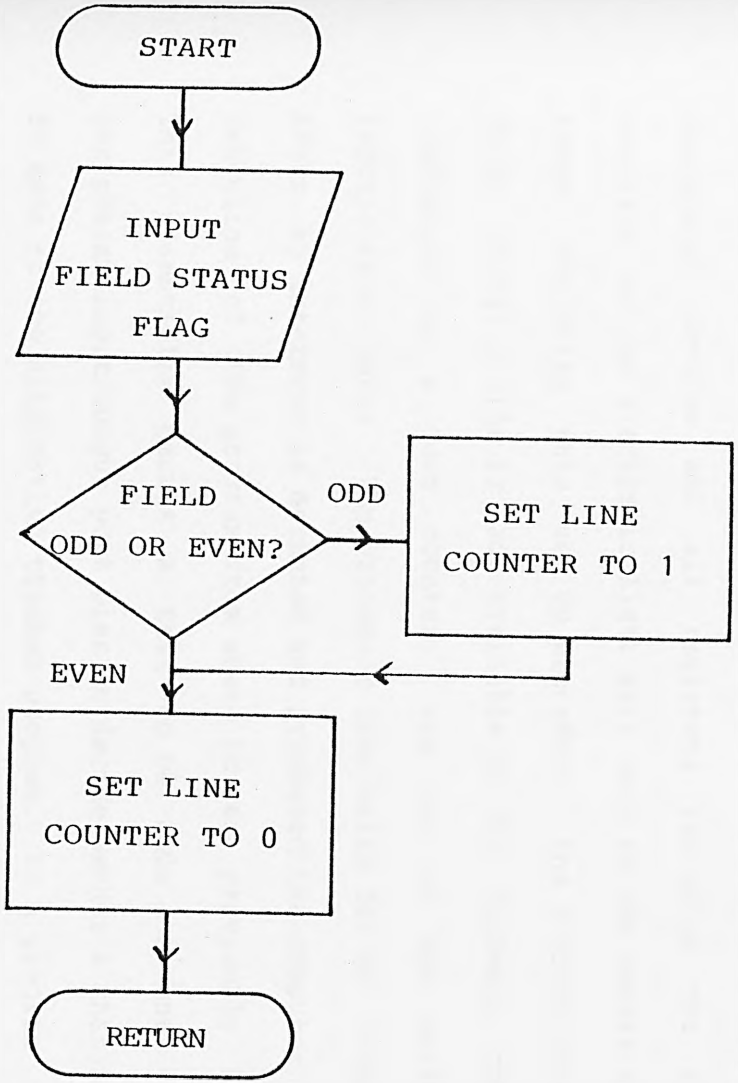


Figure 3.13 - Marker Tracker Field Interrupt Routine

communications controller interrupt.

The **main program** begins by initialising all ports of the peripheral devices and all registers including the initial position of the electronic light mark such as the centre of the image. Following this set up procedure, the x-coordinate (or delay along) a line is made available to the hardware counter, configured as a down counter, via one of the peripheral input/output ports. The processor then waits for an interrupt. After an interrupt is detected and processed the computer begins execution of the program from where it was previously halted. This essentially causes a test to be made on one of the peripheral input/output port pins to decide wether a change is to be made to the alternative tracker program. If a program change is required then a jump is made to the appropriate program otherwise a jump is made back up to the "wait for interrupt" instruction.

The **line interrupt** routine starts by incrementing the line counter, which holds the current line number, by two (since the video image is interlaced) and then comparing this updated value with the number of the line where the marker is required (y-coordinate). If a match is found then the line select flag, which informs the hardware of a match, is set otherwise it is reset. The delay required along a line (or the x-coordinate) and the line select flag is then output to the input/output ports. Interrupts are then enabled and program execution is re-commenced at the point where the interrupt occurred.

The processor speed of operation (4MHz) now becomes apparent since the time to execute this routine alone, at a 2.5MHz clock rate, is 60μs and hence if a Z80 CPU running at this speed was used then this would cause loss of some line interrupts and hence loss of synchronisation with the video signal.

The **field interrupt** consists of determining from one of the input/output port pins the current field. If the field is odd the the line counter is set to 1 otherwise it is reset to 0 thereby allowing alternate lines to be measured. Interrupts are then enabled and a test made on the joystick inputs to determine if the joystick signals have changed within the last 20ms (since the last field interrupt). If the joystick has moved, the new direction is stored, the marker moved one step in the appropriate direction and a delay counter is initiated. On the other hand if the joystick has not moved then a test is made of how long it has been since the last update of marker position coordinates. If the time is within 600ms (30 field interrupts) no change is made to the marker position otherwise the marker position is updated by one step in the specified direction. Finally the marker x-coordinate is output to the input/output ports and the interrupt routine is finished.

This procedure enables operation of the joystick on a single step movement by a small flick in the appropriate direction or a rapid sequence of steps in the specified direction by holding the joystick in that position, causing a single step followed by a rapid movement after 600ms.

The **transmit interrupt** routine starts by loading the x and y coordinates of the marker into a buffer. The serial communications interface is then enabled and the transmit interrupt disabled (so that a complete set of coordinates are transmitted before the next set can be transmitted). All interrupts are then enabled and the first byte of data from the buffer is loaded into the buffer of the serial interface controller. The byte counter (number of bytes to be transmitted) is decremented and the program execution returned to the point where the interrupt occurred.

The **serial communications interrupt** routine is the program executed when the serial controller is activated and the data byte, previously loaded into its buffer, has been transmitted. Another byte of data is loaded into the serial controller buffer and the byte counter decremented. If the byte counter is zero then all data has been transmitted - the serial controller is disabled, the transmit interrupt enabled and the interrupt routine is finished. On the other hand if the byte counter is not zero the program execution is returned to the point at which the interrupt occurred.

3.2.2.2 Operation of Tracker

For operation of the device as a bright target tracker the alternative program may be selected. In this case the program begins by initialising the registers, enabling interrupts and then executing a "wait for interrupt" instruction. Following the service of an interrupt routine the computer determines whether a

jump is to be made to the alternative program by testing one of the peripheral input/output port pins. If a jump is not made to the marker generator program then the processor makes a jump to the previous "wait for interrupt" instruction.

The **line interrupt** simply consists of incrementing the line counter by two and returning to the program executing before the interrupt.

The **field interrupt** begins by determining the current field from a test of the odd/even field signal. The line counter is set to 1 if the field is odd otherwise it is reset to 0. Control is then passed back to the point where execution was halted.

The **transmit interrupt** is exactly the same as in the marker generator program except that the x-coordinate is read from the appropriate pins of the input/output ports and the coordinates stored in the transmit buffer.

The **serial communications interrupt** routine is the same as the routine used in the alternative program and requires no further explanation.

3.2.3 Hardware/Software Interaction

Other than the description of the hardware already detailed in the previous sections it is necessary to explain the operation of the hardware in conjunction with the micro-processor system.

3.2.3.1 Marker Generator Configuration

In this mode the up/down counter is operated in the 'down' mode, selected via the 'program select' logic signal and its function is as follows. At the beginning of each line period the x-coordinate provided by the input/output ports of the micro-computer is loaded into the counter and the counter is fed with a 10MHz clock. Upon reaching zero a pulse of 100ns duration is generated and passed on to the control circuits. The control electronics combines this pulse with the line select flag, also provided by the computer, and mixes the resultant pulse onto the video signal thereby introducing a bright spot onto the image at the desired position.

3.2.3.2 Tracker Configuration

In the case of tracking the counter is implemented as an 'up' counter, once again determined via the 'program select' logic signal. At the beginning of each line period the counter, running at 10MHz, is reset. Whenever a bright target is detected the contents of the counter are automatically loaded into a shift register for temporary storage. At the same time an interrupt is transmitted to the micro-processor. In due time, when the computer has had time to process the interrupt routine, the data is read from the shift register via the input/output ports.

4.0 Dynamic Programming Algorithm for Stereo Ranging

In the reconstruction of a three-dimensional scene from a digital stereo pair of images, two problems must be solved. The first is the geometrical calculation of the three-dimensional position of a scene point from its stereo projections. The second is a rather more complex problem and concerns the problem of correspondence. When considering a point on one image how can the conjugate point be located in the other image, provided it is not occluded, in order to determine the 3-D position of the scene point ?

In classical photogrammetry this problem is solved by a human operator. The human scene-matcher can select the true match position based upon the detailed differences remaining after mentally registering the two images. A potential match point is examined by mentally analysing the two images at the trial match position. If the observed residual match errors, after removing all geometrical distortions, are comparable to the type of match errors known a priori then the trial match is judged to be the true match position.

This approach has not been possible in automated scene-matching since there has been no way to exactly register the two images. Unless this is carried out, the true match errors will be saturated by spurious match errors. Even slight residual mis-registration due to rotation, scale change and so on can offset the true match point.

The algorithm to be described in this section shows a new process that automatically registers the two images. An exact area

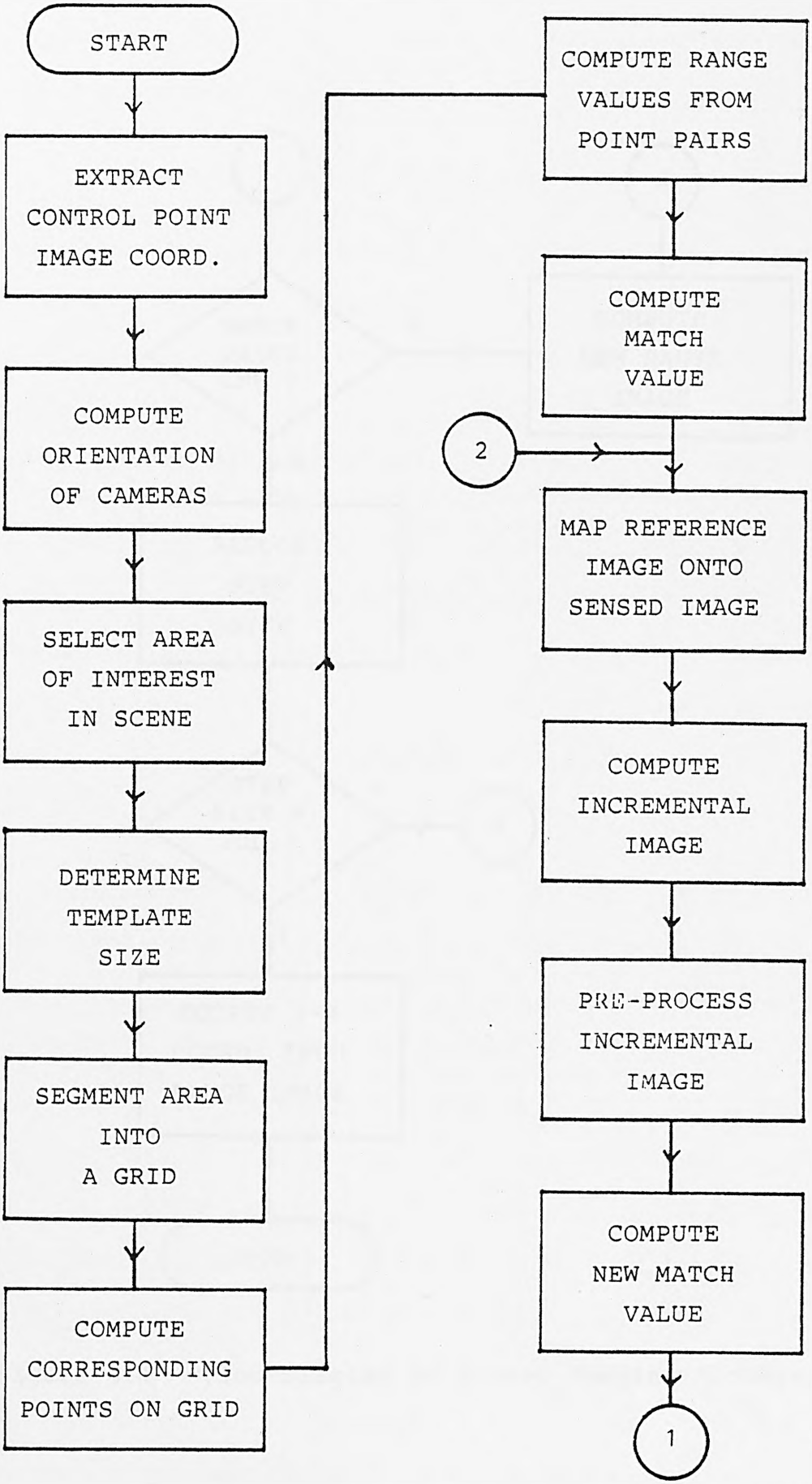
registration process which automatically removes all residual distortion due to the geometry of the stereo imaging system will be demonstrated. This process allows the match judgement to be based upon an analysis of the residual differences of structure and intensity existing after all differences due to geometrical distortion have been corrected.

A dynamic programming technique is implemented to enable an approximate registration of the two images to be improved upon by an iterative procedure which applies the exact transform from one stereo image to the corresponding stereo image. This process causes convergence of the approximate 3-D structure, represented by the scene, to the the true 3-D structure based upon an optimisation of the residual intensity differences existing after correction of all geometrical distortion.

The algorithm described enables range measurements to be obtained from the two stereo images directly without operator intervention, except in the initial stages when the system is being calibrated. Even this could be automated with appropriate hardware for measurement of a few scene coordinates.

4.1 Algorithm Description

The basic structure of the algorithm is best described with the aid of the flowchart illustrated in figure 4.1. The flowchart details the final form of the algorithm after an analysis of variations on the algorithm, described in chapter 5. The primary criterion of the algorithm is to minimise the sum of residuals


START
EXTRACT CONTROL POINT IMAGE COORD.
COMPUTE ORIENTATION OF CAMERAS
SELECT AREA OF INTEREST IN SCENE
DETERMINE TEMPLATE SIZE
SEGMENT AREA INTO A GRID
COMPUTE CORRESPONDING POINTS ON GRID
COMPUTE RANGE VALUES FROM POINT PAIRS
COMPUTE MATCH VALUE
2
MAP REFERENCE IMAGE ONTO SENSED IMAGE
COMPUTE INCREMENTAL IMAGE
PRE-PROCESS INCREMENTAL IMAGE
COMPUTE NEW MATCH VALUE
1

Cont.

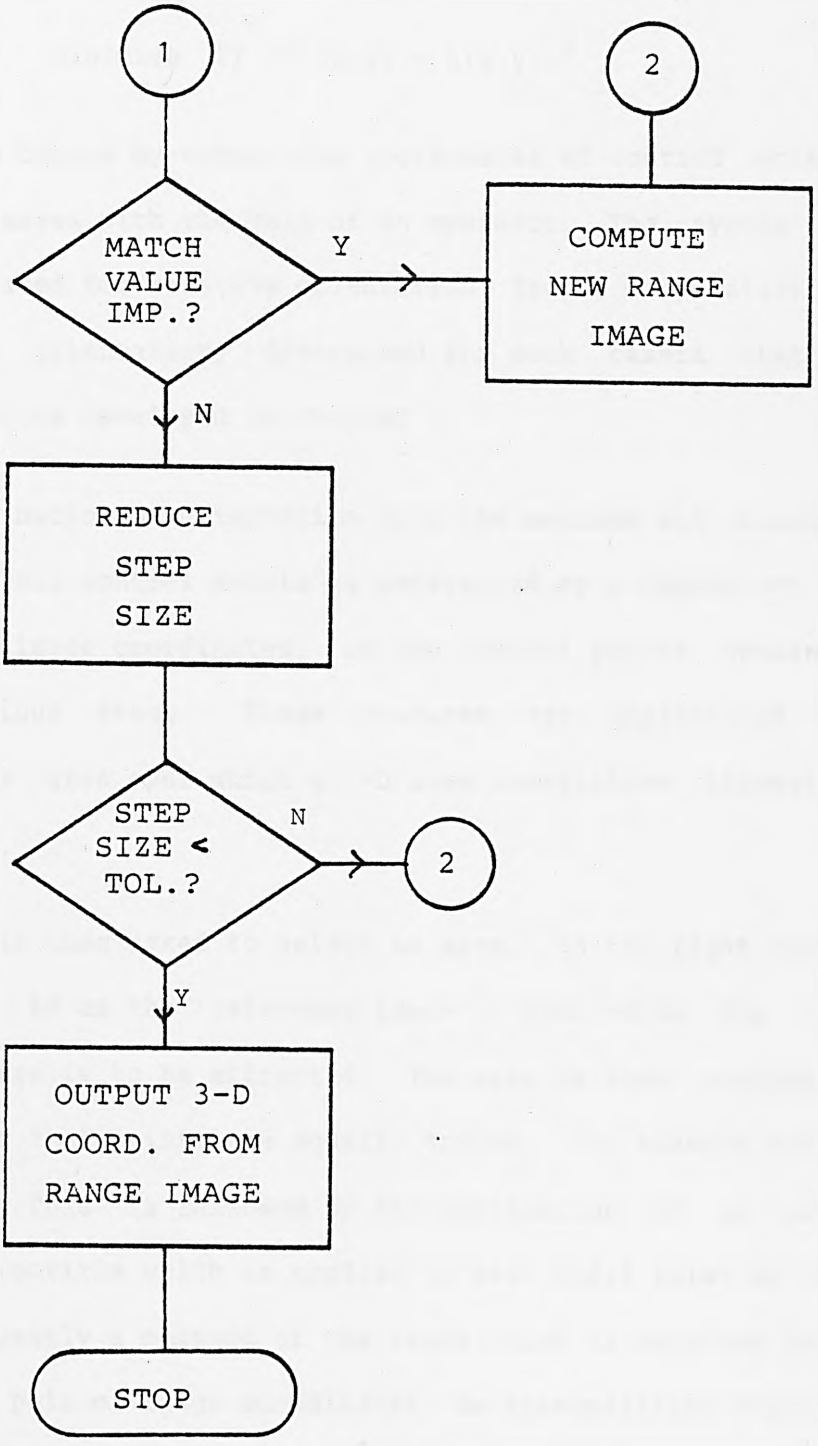


Figure 4.1 - Flow Diagram of Stereo Ranging Algorithm

squared when correction for all geometrical distortions have been catered for, as formulated below

$$\text{minimise} \quad \Sigma\Sigma \, (f^{*}(x,y) - h(x,y))^{2}$$

The algorithm begins by extracting coordinates of control points from both images with the help of an operator. The system is easily calibrated for relative orientation, from a computation of the absolute orientation, determined for each camera station using the program developed in chapter 2.

After determination of orientation data the maximum and minimum parallax for all control points is determined by a comparison of the extracted image coordinates, of the control points obtained in the previous stage. These measures are implemented to determine the area over which a 2-D area correlation algorithm will operate.

The operator is then asked to select an area, in the right image now referred to as the 'reference image', over which the 3-D scene structure is to be extracted. The area is then segmented into a grid so that points are equally spaced, for example every four pixels. This is followed by the application of an area correlation algorithm which is applied to each nodal point on the grid. Subsequently a measure of the range value is obtained from the conjugate pair of image coordinates. An interpolation routine is then applied to the grid to provide approximate range values for non-computed points within the grid. This technique enables efficient computation of approximate range values, within the selected area, without consuming useful computer time.

The main part of the algorithm may now be applied which attempts to improve upon the approximate range values computed. From an estimate of the range values a mapping from the reference image to the expected left image, referred to as the 'sensed image', is carried out and a comparison made with the actual sensed image. This mapping enables a match value to be calculated from a comparison of the transformed and the actual sensed image.

In addition a more accurate prediction of the range values may be obtained by an analysis of small perturbations in the range image. The change in a range value is either an increment or a decrement and the choice is based upon the change which provides an improvement in the 'individual match value'. The individual match value is the absolute value of the residual intensity remaining after mapping a reference image point onto the sensed image. The image composed of changes in range values will be referred to as the 'incremental image'.

Various techniques were implemented to pre-process this incremental image in order to provide a better overall improvement in the range image before applying the changes to the range image. It is necessary to pre-process the incremental image since it is possible that for a few points a 'false match' may occur due to for example occlusion. This false match would in turn provide an incorrect change in the range value associated with these points.

To minimise this type of mis-match error, the fact that there is a close correlation between surrounding range values is

considered to be a very useful criteria. For example a local averaging may be carried out to the incremental image. The different types of approach studied on a set of computer generated stereo images is fully detailed in chapter 5.

Next, the match value is computed for the case when the range image is the summation of the range image and the incremental image. If the match value is equal to or better than the previous match value, the incremental image is added to the range image and hence a new range image has been calculated. On the other hand if the match value is not improved then the incremental image is discarded and the step size used to inspect changes in range values is halved.

If the step size is reduced to a previously determined accuracy then the final range image is implemented to extract 3-D data about the scene and the program is halted. If the step size is not less than the specified accuracy then a jump is made to carry out another iteration of the procedure just described.

4.2 Detailed Description of Subroutines

Having carried out an analysis of the algorithm structure the individual subroutines will now be fully described. The exact area registration process (figure 4.2) operates on two intensity images, $f(x,y)$ and $h(x,y)$, a gradient image $g(x,y)$ associated with the reference image and two images associated with the range values of points within the reference image. These are the range image, $r(x,y)$ and the incremental image, $d\_r(x,y)$. The range

image rather than storing the intensity distribution of the projected scene, stores the distance of the corresponding point in the intensity image resolved along a constant reference direction. The reference direction can be the normal vector of a base plane upon which a model is situated.

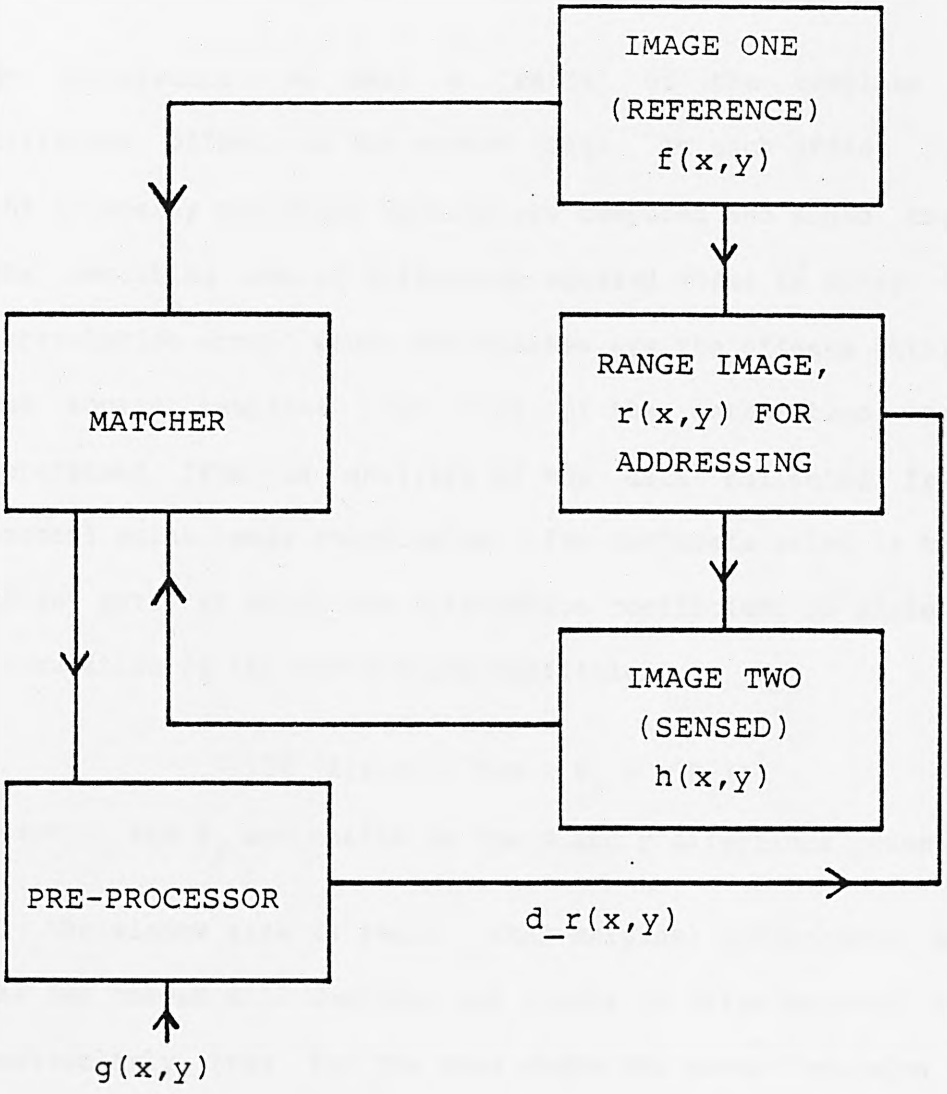


Figure 4.2 - Block Diagram of Algorithm Structure

Note: a full program listing in the 'C' programming language may be found in Appendix H.

4.2.1 Area Correlation Algorithm

The area correlation algorithm essentially selects nodal points on the grid of the selected area in the reference image. For each point a small template is extracted and correlated, over a previously determined area, with the corresponding sensed image to determine an approximation for the corresponding stereo point.

By correlation we mean a 'shift' of the template across different offsets in the sensed image. At each offset position the intensity residuals squared are computed and added together. The resulting sum of difference squared forms an entry in the 'correlation array' whose coordinates are the offsets attained by the source template. The size of the correlation array is determined from an analysis of the data collected from the control point image coordinates. The conjugate point is then the offset entry at which the correlation coefficient is minimum. The formulation of the correlation coefficient is

$$\Sigma\Sigma \, (f(x,y) - h(x + S_x, y + S_y))^2$$

where S_x and S_y are shifts in the x and y directions respectively

If the window size is small, then marginal differences between the two images will dominate and result in false matches. This is particularly true for the case where the scene contains large areas of constant texture. On the other hand, large windows will result in poorly defined edges of the objects in the picture but will reduce false matches. Therefore in the algorithm, the size of template used was set to 39x39 pixels.

It should be pointed out that the camera geometry plays a vital

part in the arrangements of the correlation algorithm in addition to the number of correlations required for each point. These points are briefly outlined below.

4.2.1.1 Epipolar Lines

A family of epipolar lines is traced out by the intersection of the plane containing both camera perspective centres and the world point with the image planes of each camera as illustrated in figure 4.3. Any point located in one image must have its corresponding point in its conjugate epipolar line. Therefore to reduce correlation time the previous 2-D correlation scan area is contracted into a 1-D strip.

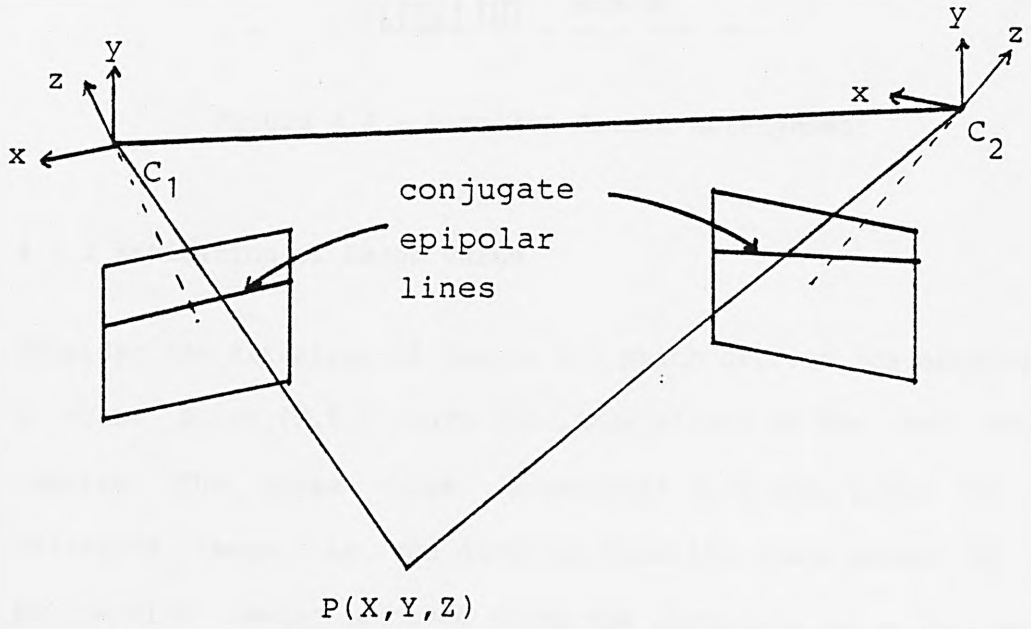


Figure 4.3 - Representation of Epipolar Lines

A special case of the use of epipolar lines is when a pair of

video cameras are placed parallel to the same plane in 3-D space (figure 4.4), with a rotation about the y-axis only. In this situation the epipolar lines lie along consecutive scan lines and so correlation can be carried out along a single line only. This is the simplest method of addressing within the two images as compared with any other geometrical camera arrangement.

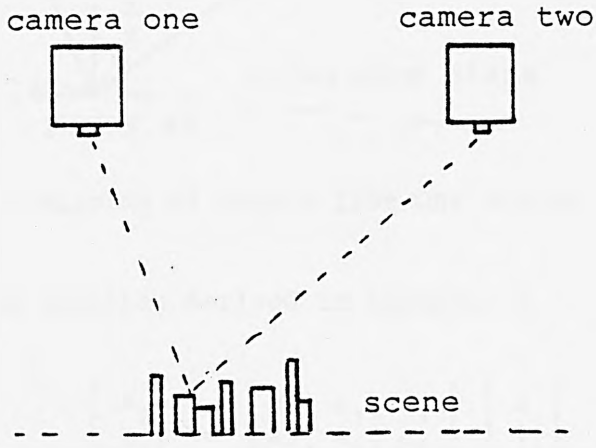


Figure 4.4 - Parallel Camera Arrangement

4.2.2 Estimation of Range Value

Consider the situation of figure 4.5 which depicts the mapping of a scene point (X,Y,Z) onto the image planes of the two stereo cameras. The range value, associated with the point in the reference image, is the distance from the scene point to the perspective centre resolved along the direction (n_1,n_2,n_3) which may be the normal vector of a reference plane in object space.

The image point (x_1,y_1) in the reference image and its corresponding point (x_2,y_2) in the sensed image are associated by

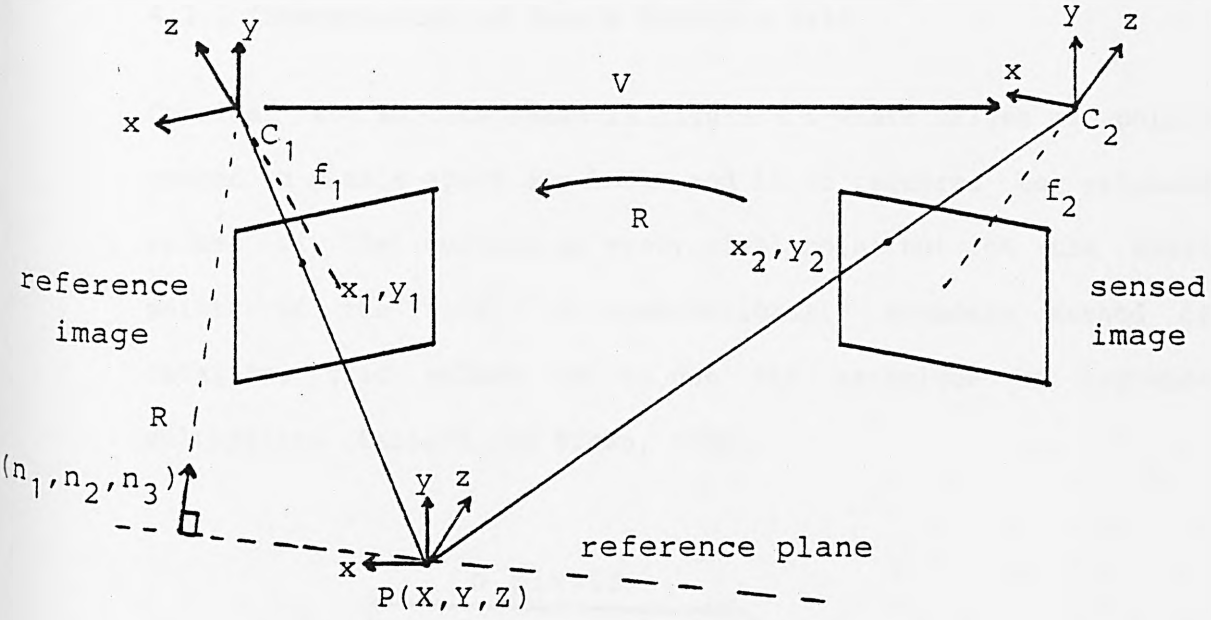


Figure 4.5 - Mapping of Points from One Stereo Image to Another

the following equation derived in Appendix I.

$$\begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

where

$$a_{ij} = r_{ij} - (v_i n_j / R)$$

The range value can be found by a least squares approximation as (full derivation in Appendix J)

$$R = (a.c + b.d) / (a^2 + b^2)$$

where

$$a = x_2.R_3 + f_2.R_1 \qquad b = y_2.R_3 + f_2.R_2$$

$$c = x_2.c_3 + f_2.c_1 \qquad d = y_2.c_3 + f_2.c_2$$

$$R_i = r_{i1}.x_1 + r_{i2}.y_1 - r_{i3}.f_1$$

and

$$c_i = v_i.(n_1.x_1 + n_2.y_1 - n_3.f_1)$$

In order to include the distortions in both images the points should be rectified with the equations derived in chapter 2, before applying the above formula.

4.2.3 Interpolation of Points Within a Grid

Consider the surface shown in figure 4.6 where values of points spaced n pixels apart are known and it is required to estimate values of the surface at every pixel point not on the nodal points of the grid. A computationally economic method of obtaining grid values is to use the technique of Lagrange multipliers (Ballard and Brown, 1982).

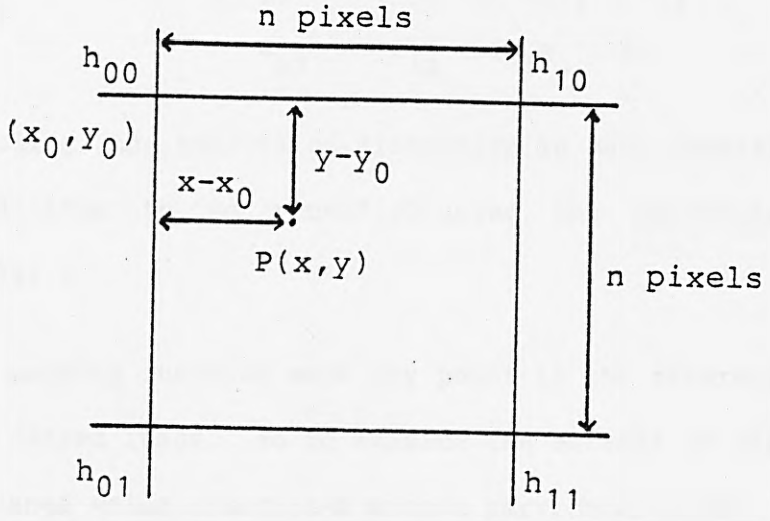


Figure 4.6 - Interpolation of a Regular Surface

The Lagrange multiplier enables intermediate points within the grid to be computed without affecting the nodal points. The value of the height, h(x,y) at point P(x,y) is given by

$$h(x,y) = (1 - p).(1 - q).h_{00} + (1 - p).q.h_{01} + p.(1 - q).h_{10} + p.q.h_{11}$$

where

$$p = (x - x_0)/n \quad \text{and} \quad q = (y - y_0)/n$$

This is the approach used for the determination of range values

for all points within the area under investigation in the the reference image.

4.2.4 Computation of Incremental Range Values

As previously mentioned the mapping from the reference image to the sensed image is represented by

$$\begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

where

$$a_{ij} = r_{ij} - (v_i n_j / R)$$

Including the effects of distortion in both cameras enables the coordinates to be rectified using the equations derived in chapter 2.

The mapping function maps any point in the reference image onto the sensed image, so to examine the effects of small changes in the range value associated with a particular pixel, (x_1, y_1) the individual match value defined as

$$\text{individual match value} = |(f^*(x,y) - h(x,y))|$$

may be computed on a pixel by pixel basis. The match value is merely the absolute value of the residual intensity remaining after the point (x_1, y_1) has been mapped onto the sensed image. From the examination it is possible to determine wether a small change should be added or subtracted from the associated range value to improve the match value of that pixel. This is essentially a hill-climbing technique.

The step value or increment is selected to be inversely proportional to the differential of the distance in the image plane, of the sensed image, with respect to the range. The derivative is computed from the equations derived in Appendix K, shown below

$$\partial x/\partial R = -f_2.(c_1.R_3 - c_3.R_1) \;/\; (R.R_3 - c_3)^2$$

$$\partial y/\partial R = -f_2.(c_2.R_3 - c_3.R_2) \;/\; (R.R_3 - c_3)^2$$

where $$c_i = v_i.(n_1.x_1 + n_2.y_1 - n_3.f_1)$$

and $$R_i = r_{i1}.x_1 + r_{i2}.y_1 - r_{i3}.f_1 \qquad i = 1,2,3$$

So for a small change in the range value, the corresponding change in the image plane is represented approximately by

$$\partial s/\partial R = |\partial x/\partial R| + |\partial y/\partial R|$$

The incremental step is then a scale factor multiplied by $\partial R/\partial s$. The scale factor, at the beginning of the algorithm, is 1.0 and this is the step size referred to in the description of the algorithm in the earlier section.

In order to minimise errors due to mis-matched points a further condition imposed on the incremental range computation function was that the intensity gradients in both images prior to calculation of the incremental step should be within 5 degrees. The selection of this value will be explained in chapter 5.

The gradient values were computed using the central difference equations (Lanczos, 1967) which state that for a point, P(x) on a line of equally spaced points the gradient is given by

$$\text{gradient} = (\Sigma\, f(x + n).n)/(\Sigma\, n^2)$$

for values of n from -t to +t where t is the number of pairs of points used.

Applying this same formula to a two dimensional grid and extending the formula only upto n = 2 gives

$$\text{gradient along x-axis, } G_x = (\Sigma\ f(x + n,y).n)/(\Sigma\ n^2)$$

$$\text{gradient along y-axis, } G_y = (\Sigma\ f(x,y + n).n)/(\Sigma\ n^2)$$

therefore the gradient angle at point P(x,y) is

$$\text{arc tan } (G_y\ /\ G_x)$$

If the comparison of gradient angles reveals an angle greater than 5 degrees, the incremental step is set to an improbable value such as 100 so that the interpolation routine can select this point when scanning the incremental image for mis-matched points.

4.2.4.1 Interpolation of Sensed Image

During computation of the individual match values in the above procedure, if the change in range value is liable to produce less than a single pixel step in the image plane then the match value is computed from the interpolated intensity value, at the mapped position in the sensed image. The interpolation formula used is the Lagrange multiplier as used in section 4.2.3

4.2.5 Pre-processing of Incremental Image

In order to reduce errors due to mis-matched points the

incremental image is pre-processed by passing a variable averaging template over the incremental image. The template size is obtained from corresponding points in the gradient image.

The gradient image is obtained at the beginning of the algorithm by computing gradient values in the reference image and scaling so that the maximum template size is 15x15. The gradient value calculated from the intensity image in the following manner. The gradients, G_x and G_y along the x and y axes respectively, at a point P(x,y) are calculated as

$$G_x = f(x + 1,y) - f(x - 1,y)$$
$$G_y = f(x,y + 1) - f(x,y - 1)$$

Hence an estimate of the gradient at P is

$$\text{Gradient} = |G_x| + |G_y|$$

4.2.6 Calculation of Match Value

The match value is defined as the sum of residuals squared after the reference image has been mapped onto the sensed image, that is

$$\text{match value} = \Sigma\Sigma \, (f^*(x,y) - h(x,y))^2$$

where the difference is calculated after all errors due to geometric distortions have been removed. This means that the coordinates on the reference image are rectified, then mapped onto the sensed image and finally converted to comparator coordinates before any comparison can take place.

4.2.7 Interpolation for Mis-matched Points

In the situation of section 4.2.4 where the gradients between the approximated corresponding points have a gradient angle difference greater than 5 degrees, the pre-set tolerance, it is necessary to compute an approximate incremental range value from the surrounding points in the incremental image. This is carried out by a modification of the interpolation technique described by Falconer (1971). The interpolation algorithm states that given a surface with height values randomly placed, the estimation of the height at a point, P(x,y) may be approximated by the formula

$$\text{height}(x,y) = \Sigma\text{weight}(i).\text{height}(X(i),Y(i)) \;/\; \Sigma\text{weight}(i)$$

where the weighting function is

$$\text{weight}(i) = (S - D)^2 \;/\; D^2$$

and D is the distance of the data point (X(i),Y(i)) from the point P and S is the radius of a circle centred on P. Only data points within the circle are used and in order to obtain a reasonable estimation of the height it is necessary to select a radius which provides on average, at least six data points.

For the purposes of simplicity and efficiency in computation, the algorithm is modified as follows. Instead of data points within a circle a square is centred on the unknown point. Starting with the closest contour, consisting of 8 pixels, an inspection is made for data points in the incremental image and partial weights computed (to be detailed below) for each data point located. If at the end of the contour a minimum of six data points have not been found then the next contour, consisting of 16 pixels is

traced and so on until weights for at least six data points have been computed.

To calculate the data point weight factor the radius, S of the effective circle centred on the unknown point is required. But since the largest square contour will not be known until a minimum of six data points have been located, the value for S will not be known. A solution to this problem is to expand the above equation and extract partial summations which do not require the use of S.

The weight is given by

$$\text{weight}(i) = (S^2 - 2.S.D + D^2) / D^2$$
$$= (S^2/D^2) - (2.S/D) + 1$$

therefore the height equation reduces to

$$\text{height}(x,y) = (a.S^2 - 2.b.S + c) / (d.S^2 - 2.e.S + f)$$

where $a = \Sigma\ (\text{height}(X(i),Y(i)) / D^2)$

$b = \Sigma\ (\text{height}(X(i),Y(i)) / D)$

$c = \Sigma\ (\text{height}(X(i),Y(i)))$

$d = \Sigma\ (1 / D^2)$

$e = \Sigma\ (1 / D)$

and f = count for number of data points encountered

This has been simplified to computing six partial summations which do not require the radius S and so whenever a data point is encountered, whilst tracing a contour, the partial summations, 'a - f' can be carried out. After at least six data points have been

found the value of S is estimated as

$$S = (\text{square side length} + 0.5) \; / \; \sqrt{2}$$

where the square side considered is the outermost square in the contour trace and the height is then estimated. The 0.5 is added so that if there is a data point at the corner of the square, its weighting is not eliminated.

4.2.8 Determination of World Coordinates from Range Image

Once the range image has been found it is a simple matter to compute the three-dimensional coordinates from the range image, $r(x,y)$ and the image coordinates, (x_1,y_1). The coordinates of the world point are given by

$$X = x_1.R \; / \; (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

$$Y = y_1.R \; / \; (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

$$Z = -f_1.R \; / \; (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

The full derivation of the above formulae can be found in Appendix L.

5.0 Algorithm Design Using Computer Generated Images

In the previous chapter an incremental image was introduced in the explanation of the stereo ranging algorithm and this chapter is concerned with various methods in which pre-processing may be applied to the incremental image. Pre-processing of the incremental image allows modifications to the range image to be controlled so that range values of adjoining pixels are always correlated, ensuring a sensible solution from the algorithm.

In the evaluation of the algorithm it is essential to minimise the noise in the stereo images used for testing of the algorithm, so that only known factors influence the performance of the algorithm. The method chosen was to generate a pair of computer generated models and project stereo images of these models. These images have no photometric distortion or noise to deteriorate the algorithm performance. The models chosen were - a smoothly varying structure, such as a hilly terrain, and a structure with sharp edges.

The undulating structure model enabled the algorithm performance to be optimised, whereas the model with sharp edges enabled investigation of the algorithm when presented with a model, consisting of large amounts of occluded information between the camera stations.

The ideal situation for photogrammetric mensuration is to place the cameras such that the ratio between the camera separation and the average distance to the scene is between 1 in 4 and 1 in 6. It was therefore decided to place the cameras with the camera

perspective centres at (120,140,550) and (240,140,550) in object space. An image size of 128x128 was deemed suitable for evaluation of the algorithm and so an equivalent focal length of 300 for the cameras was chosen and a rectangular model placed parallel to the z-axis with corners at (20,20,0) and (340,260,0). This is sufficient to cover every image point with at least one point mapped from the interpolated surface, without any form of coordinate scaling required in the image plane.

Since the cameras were displaced along the x-axis to the left and right of the centre of the model, it was necessary to tilt the cameras towards the centre of the model to provide optimum model coverage. With reference to figure 5.1, which shows a side view of the left camera, the tilt about the y-axis may be calculated from simple geometry.

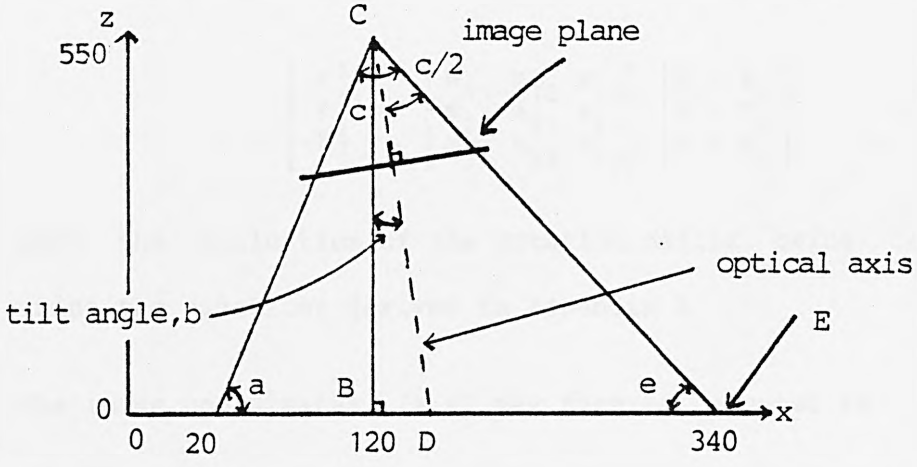


Figure 5.1 - Side View of Left Camera

C is the perspective centre and line CB is the perpendicular onto

the base of the model. Line CD passes through the perspective centre and is perpendicular to the centre of the image. By geometry

$$a + c + e = 180^\circ$$

hence $$c = 180^\circ - a - e$$

therefore the tilt angle is given by

$$\begin{aligned} \text{angle BCE - angle DCE} &= (180^\circ - 90^\circ - e) - (90^\circ - a/2 - e/2) \\ &= (a - e) / 2 \\ &= (\tan^{-1}(550/100) - \tan^{-1}(550/220)) / 2 \\ &= 5.7^\circ \end{aligned}$$

The collinearity equations were used to project the model coordinates onto the image coordinates with a rotation of 5.7 degrees about the y-axis for the left image and -5.7 degrees for the right image. The transformation equations for each camera are therefore

$$\begin{bmatrix} x \\ y \\ -f \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} X - X_c \\ Y - Y_c \\ Z - Z_c \end{bmatrix}$$

with the evaluation of the rotation matrix being carried out using the equations derived in Appendix A.

The image coordinates, (x,y) may then be computed as

$$x = \text{integer part of } [-f.(R_1 / R_3) + 64.5]$$

and $$y = \text{integer part of } [-f.(R_2 / R_3) + 64.5]$$

where $R_i = m_{i1}.(X - X_c) + m_{i2}.(Y - Y_c) + m_{i3}.(Z - Z_c)$ $i = 1,2,3$

The offset of 64.5 was added to provide a reference origin, on

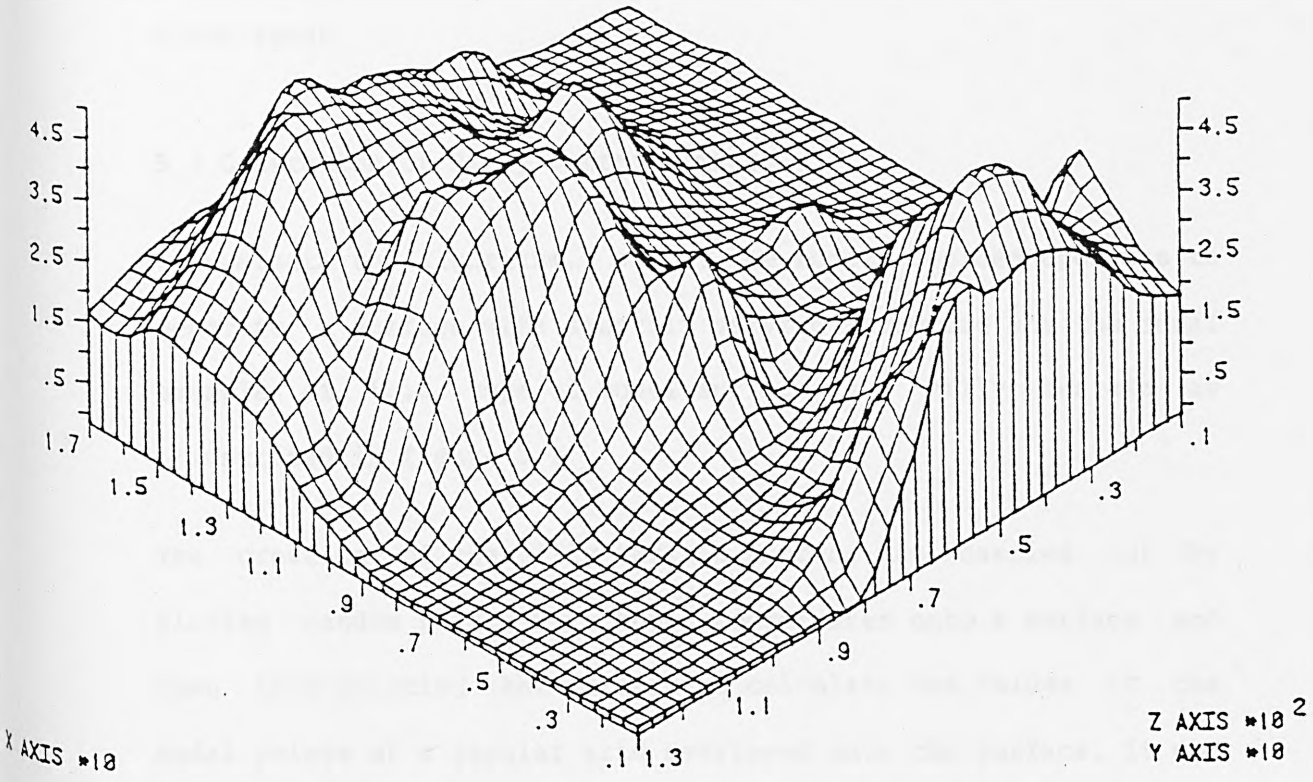


Figure 5.2a - Hilly Terrain Model

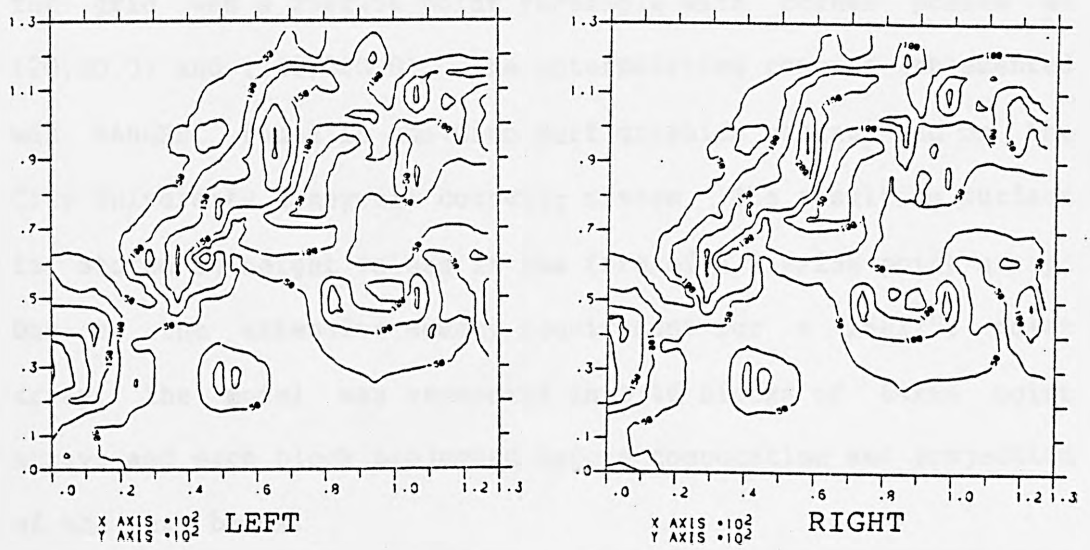


Figure 5.2b - Contour Plots of Hilly Terrain Stereo Images

the image plane, at the top left hand corner as required for a video image.

5.1 Generation of Hilly Terrain Images

To evaluate the algorithm, one of the models decided upon was to generate a surface with smoothly varying structure and the ideal example of this type of model surface is a hilly terrain as illustrated in figure 5.2.

The creation of the hilly terrain surface was carried out by placing random points with known coordinates onto a surface and then interpolating the surface to calculate the values at the nodal points of a regular grid overlayed onto the surface. It was necessary to interpolate the surface in order to provide a set of points on the surface with known coordinates so that the mapping equation above could be implemented.

The grid was a 256x256 point rectangle with corner points at (20,20,0) and (340,260,0). The interpolating routine implemented was RANGRD, found in the Gino Surf graphics package run on the City University Honeywell computer system. The resulting surface is stored as height values in the form of a 256x256 point array. Due to the extensive memory requirement for a 256x256 point array, the model was segmented into 16 blocks of 64x64 point arrays and each block projected before computation and projection of the next block.

The procedure involved in producing the stereo images was as follows (see figure 5.3).

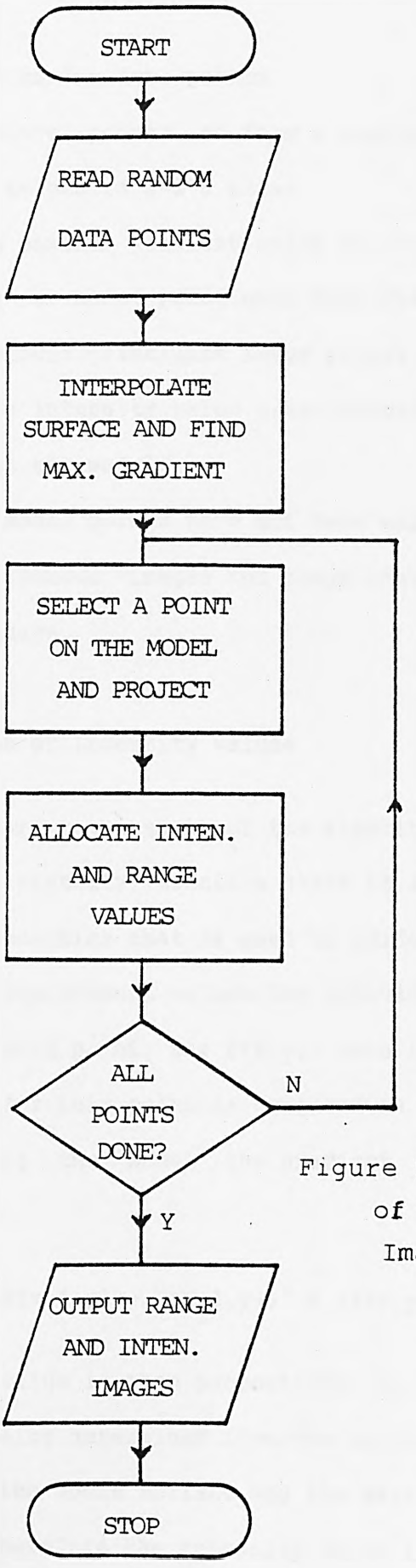


Figure 5.3 - Flow Diagram of Hilly Terrain Image Generator

1: Read in random data points.

2: Interpolate points to form a regular grid and store height values in a 2-D array.

3: Compute maximum gradient value of interpolated surface.

4: Project a model point onto both stereo images taking into account coincident image points.

5: Allocate intensity value proportional to gradient of the point in the model.

6: If all model points have not been mapped go back to 4.

7: Output stereo images and range image associated with right image.

5.1.1 Computation of Intensity Values

It is a necessary requirement of the algorithm to have images with sufficient intensity variation since it is the comparison of image point intensities that is used to guide the process which selects range improvement values for individual pixels. After projection of a grid point, say P(x,y), onto the image plane, the intensity value for this point is computed as follows. From the grid representing the model the gradient at that point is calculated from

$$\text{gradient} = \sqrt{[(F(x-1,y) - F(x+1,y))^2 + (F(x,y-1) - F(x,y+1))^2]}$$

The intensity value is then proportional to the gradient, the scale factor being determined from the maximum gradient value calculated over the whole surface and the maximum intensity value required, 255. Therefore the intensity value is

intensity = gradient * (255 / max. gradient)

The value 255 being chosen to provide a grey level resolution of 8-bits.

The computer generated images were intended to represent a real scene, so if a real image is considered then the simplest and most featureless case is one in which the surface being analysed is of one colour and of uniform reflectance. For uniform overhead illumination the brightness of the image would depend upon the gradient. The brightness would be greatest where the gradient is zero and would be low where the gradient is very large. This is the principle basis for the generation of the computer generated stereo images except that the images were inverted so that the intensity is low where the gradient is low and high where the gradient is high.

In practice the images would also vary in brightness due to observer position, surface reflectance characteristics, colour and surface defects. These factors have not been taken into consideration in the generation of the stereo model.

5.1.2 Multiple Point Coincidence

Whilst mapping points onto an image a note is made of the range values of the mapped points since the possibility may occur where a point may map onto an already occupied image point. This can happen when multiple model points lie along the same line as the perspective centre and the image point as shown in figure 5.4a. A situation like this is dealt with by associating a range image

with the intensity image, the range value of a point being the distance of that point in object space to the perspective centre resolved along a constant reference direction. The direction chosen in this case was the normal vector of the plane upon which the model was situated, that is the direction of the z-axis. The range value is therefore simply the z-component of the vector from the perspective centre to the model.

When a point is mapped onto an image point already occupied, the intensity value allocated to that image point corresponds to the intensity value of the point with the smallest range value. This rule deals with the situation of multiple point overlap and generates a true perspective view of the hilly terrain model.

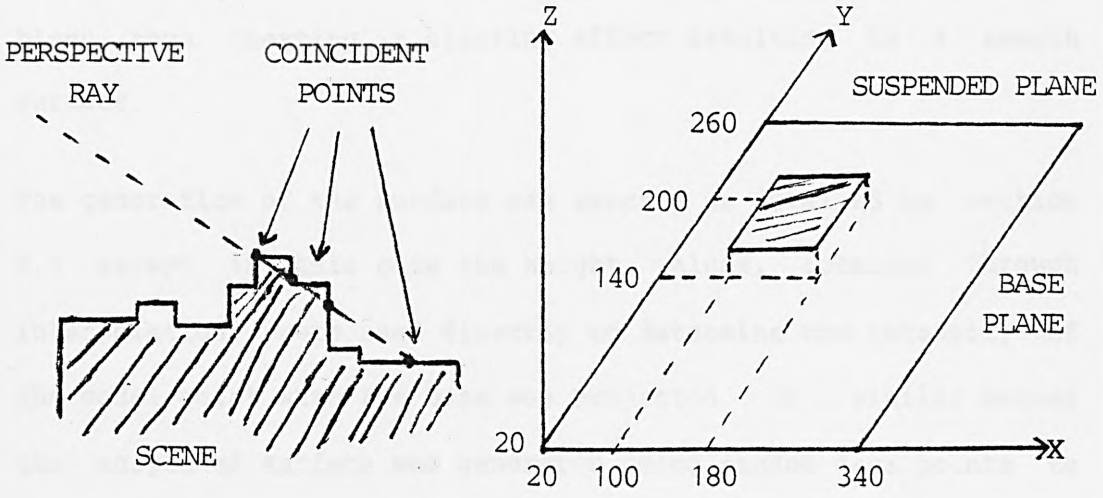


Figure 5.4
a - Multiple Point Coincidence, b - Planar Surface Model

5.2 Generation of Planar Surface Images

The creation of a planar surface model (figure 5.4b) is essentially the generation of a model base with a flat surface

suspended above it. Evaluation of the algorithm performance, when a model with sharp edges is presented to it is then possible. A model with sharp edges is required to analyse the effects of occluded regions, as an extreme example, on the algorithm.

Since the algorithm is based on intensity analysis it was necessary to place texture onto the flat surface as well as the base of the model. By texture we mean a variation in intensity accross the image so that there are image features which may be used by the stereo ranging algorithm. Texture is a visual primitive with certain invariant properties which occur repeatedly in different positions, deformations and orientations inside a given area. An example of such a primitive is an intensity pixel. As the area becomes larger, the texture tends to blend thus creating a blurring effect resulting in a smooth surface.

The generation of the surface was exactly as detailed in section 5.1 except in this case the height values, obtained through interpolation, were used directly to determine the intensity of the model point when the base was projected. In a similar manner the suspended surface was generated using random data points to predict the intensity of model points. The placement of the suspended surface was at a position 100 units above the base plane, with the corners at (100,140,100) and (180,200,100).

The flowchart for creation of the stereo images is shown in figure 5.5 and detailed below.

1: Read in random data points for base plane.

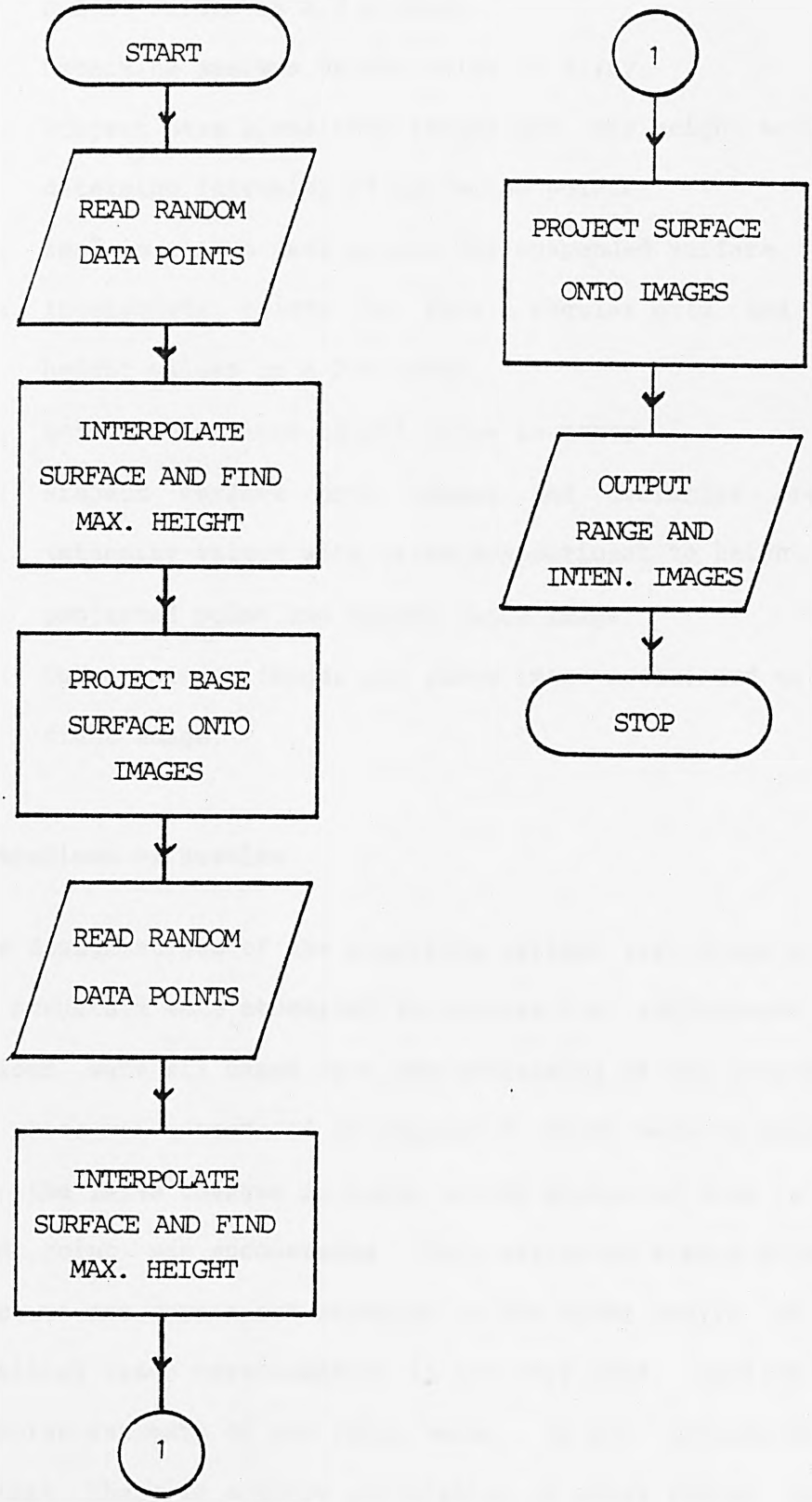


Figure 5.5 - Flow Diagram of Planar Surface Image Generator

2: Interpolate points to form a regular grid and store height values in a 2-D array.

3: Determine maximum height value in array.

4: Project base plane onto images and use height values to determine intensity of projected points.

5: Read in random data points for suspended surface.

6: Interpolate points to form a regular grid and store height values in a 2-D array.

7: Determine maximum height value in array.

8: Project surface onto images and overwrite previous intensity values with value proportional to height of projected point and append range image.

9: Output stereo images and range image associated with the right image.

5.3 Comparison of Results

In the design stages of the algorithm various variations on the basic structure were attempted to improve its performance. The variations were all based upon pre-processing of the incremental image, which was introduced in chapter 4. These were necessary to reduce the false changes in range values predicted when a mis-matched point was encountered. This may occur when a point is visible to one camera but obscured to the other camera or when the initial range approximation is not very good, leading to an even worse estimate of the range value. In all situations the fact that there is a close correlation in range values between adjacent pixel points was used as a guide to the pre-processing

techniques applied to the incremental image.

The selection of techniques implemented are described in detail in this section. A comparison of results is obtained by computing the maximum, mean and RMS (root mean square) error when the true range image is subtracted from the optimised range image, derived through application of the algorithm. The subtraction is carried out on a pixel by pixel basis, that is given the true range image, $r_t(x,y)$ and the measured range image, $r_m(x,y)$ the mean and RMS value is given by

$$\text{mean} = \Sigma|(r_t(x,y) - r_m(x,y))| \quad / \text{ number of pixels}$$

$$\text{RMS} = \sqrt{[\Sigma(r_t(x,y) - r_m(x,y))^2 \quad / \text{ number of pixels}]}$$

The algorithm was iteratively applied until the match value for the present iteration became worse than the match value for the previous iteration. When this occurred the step size was reduced by a factor of two and the incremental image re-evaluated. When a step size of 0.5 pixel had been reached and the algorithm could not proceed further without a reduction in step size, the algorithm was halted. The match value is the sum of intensity difference squared when the reference image has been mapped onto the sensed image, taking into account all geometric distortion. The results provided are for the case when the initial range increment was proportional to 1 pixel.

The significance of the match value is that it enables examination of the algorithm at every iteration so that although the range increments are computed on a local analysis of the images, the range image convergence is guided by the overall

effect of changes to the range image. So that for example, the algorithm converges to a global minimum rather than a local minimum. Hence in the results presented in this section, the match value can reach a very small value but only at the expense of decreasing the overall accuracy of the solution. It is better to reach a solution where all points are within a specified accuracy rather than a solution where most of the points are very accurate with a few points considerably off target.

5.3.1 Fixed Size Smoothing Template

In order to improve the estimated range improvement values in the incremental image, it was pre-processed by filtering with a fixed size smoothing template before adding to the range image. By smoothing with a nxn template we mean the substitution of every incremental image point with an average of the nxn area centred on the appropriate point. So the value at the point P(x,y) is represented by

$$d\_r(x,y) = (\Sigma_j \Sigma_k \; d\_r(x + j, y + k)) \; / \; n.n$$

where $j,k = -n/2$ to $+n/2$

This smoothing enabled the incremental range values for adjacent pixels to be closely correlated. The larger the template size, the higher the correlation and the smaller the template size, the lower the correlation.

The results obtained after using a template size from 1x1 through to 11x11 are shown in table 5.1 for both computer generated models.

	Template Size	Max. Error	Mean	RMS	Match Value
Hilly Terrain Images	1x1	70.724	9.723	13.724	48833
	3x3	91.438	24.186	30.633	205144
	5x5	88.564	23.872	30.419	223418
	7x7	87.513	23.681	30.270	155774
	9x9	88.668	23.808	30.366	143232
	11x11	87.795	23.769	30.394	188727
Planar Images	1x1	145.385	58.591	63.889	269334
	3x3	110.121	56.814	70.942	1139287
	5x5	107.394	56.487	70.662	1282095
	7x7	105.421	56.382	70.647	1357895
	9x9	104.595	56.692	71.531	1427195
	11x11	104.767	56.621	71.492	1413177

Table 5.1 - Fixed Size Smoothing Template

It can be seen that the best results for the hilly terrain model are obtained when there is no averaging (1x1 template) of the incremental image because this technique arrives at the least maximum, average and RMS error on completion of the algorithm. On the other hand the results of the planar model indicate at first glance that better results may be obtained by using a large averaging area but closer inspection of the results indicates that the RMS value is least for a template size of 1x1 by quite a wide margin, the mean value is only marginally larger and the maximum value is considerably larger indicating that there are a few points that have strayed from the true range value. Overall, the performance of the algorithm is better when using no averaging.

5.3.2 Variable Averaging Template

The application of a fixed averaging template caused an averaging of the incremental image in all types of terrain, irrespective of

the model structure and so at points where there was large structural variation in the model, the tendency was to over smooth the incremental image and at points of low structural variation, the tendency was not to filter the incremental image sufficiently.

As a result of this technique, the next path to be investigated was the application of a variable averaging template dependant upon local characteristics of the reference image. The local quality function chosen was the gradient value. The idea was that the gradient at an image point would provide an indication of the variation in range values around that image point.

At the beginning of the algorithm, the gradient image was calculated from the reference image and scaled so that the larger gradients correspond to small template sizes and the smaller gradients correspond to large template sizes. A linear scale was chosen so that the template size at a particular pixel was directly proportional to the gradient value at that pixel. The template size was

$$\text{template size} = \text{scale}.(1 - \text{gradient}/\text{max. gradient}) + 1$$

where the value for 'scale' is decided from the maximum template size required, 'gradient' is the intensity gradient and 'max. gradient' is the maximum gradient computed over the whole image. The gradient at the point P(x,y), is computed from the gradients G_x and G_y along the x and y axes respectively as

$$G_x = f(x + 1,y) - f(x - 1,y)$$

and $$G_y = f(x,y + 1) - f(x,y - 1)$$

Hence an estimate of the gradient at P is

$$\text{gradient} = |G_x| + |G_y|$$

This complies with the theory that when there is no texture or very little, the averaging area should be increased, due to the reliability of the computed incremental range values being low at these points and when there is a large amount of texture, the averaging area should be reduced.

	Gradient Template Size	Max. Error	Mean	RMS	Match Value
	1x1	46.746	9.453	12.746	192041
Hilly	3x3	95.267	24.242	30.471	186930
Terrain	5x5	109.552	26.188	32.338	498869
Images	7x7	105.153	28.336	34.755	870124
	9x9	123.841	31.258	37.759	1046907
	11x11	139.599	33.700	41.108	1248537
	1x1	106.389	54.508	61.211	1362564
	3x3	121.687	57.926	70.620	723672
Planar	5x5	129.956	57.168	69.768	825564
Images	7x7	133.931	55.825	68.236	1024989
	9x9	157.945	55.114	67.658	1330710
	11x11	174.040	54.076	66.870	1572832

Table 5.2 - Variable Averaging Template

The gradient information was incorporated into the algorithm by firstly extracting the gradient image from the reference stereo image and then scaling the image for a suitable template size. The smoothing template size was therefore variable from pixel to pixel. In the results presented in table 5.2 a smoothing template was passed over the gradient image before scaling the gradient image for a maximum template size of 15x15. The filtering of the gradient image was carried out for a range of different template

sizes varing from a 1x1 template (no averaging of gradient image) through to a 11x11 template size.

The pre-smoothing of the gradient image enabled the effects of small changes in the averaging template to be investigated. From the results it can be seen that the best results are obtained when the gradient image is not smoothed at all. Nevertheless it is worth pointing out that when the gradient image is smoothed, the hilly terrain scene gives a considerably worse performance on the algorithm than the planar scene which gives only slightly worse results.

In both cases it can be seen that the application of a variable averaging template, based upon gradient values, reduces the maximum, average and RMS error and hence improves the overall algorithm performance.

5.3.3 Elimination of Mis-matched Points

Although the application of a variable smoothing template did improve the results, there were still errors occurring in the algorithm as indicated by the maximum errors obtained in the previous section.

So far, the fact that there may be a point that has been mis-matched has not been dealt with fully. By mis-matched we mean situations where the initial range value may lead to a local minimum which is not the global minimum sought or where a point is masked to one of the camera stations. The figure indicates two cases - one where there are multiple points in one image which

map onto a single point in the conjugate image, the other where the initial range value leads to a local minimum, point A rather than the global minimum, point B.

Theoretically, when two approximately corresponding points are in the correct region the intensity gradients in both images should be of the same order in angular direction. Utilisation of this point was explored with a comparison of the gradients and setting a tolerance for the 'differential angle' from 5 degrees to 30 degrees in steps of 5 degrees. The differential angle is the absolute value of the angle remaining after subtraction of the angle computed in one image from the angle computed in the other image at the potential match point.

When a test is carried out on a pair of potential corresponding points, if the residual angle remaining after subtraction of the gradient angles in both images was greater than the tolerance angle, then the incremental range value associated with the reference image for this particular point was set to zero. On the other hand if the residual angle was within the tolerance angle then the incremental range value was computed as described in chapter 4.

	Tolerance Angle	Max. Error	Mean	RMS	Match Value
Hilly Terrain	5	39.907	9.472	12.681	209086
	>10	88.993	23.968	30.398	209086
Planar Images	5	102.264	54.009	60.942	1447205
	>10	104.085	56.410	70.941	1447479

Table 5.3 - Elimination of Mis-Matched Points

The results obtained from a comparison of gradient angles are shown in table 5.3. It is clear that the afore mentioned theory is valid since the results for a tolerance angle of 10 degrees or more give a worse performance than a tolerance angle of 5 degrees. In addition the maximum and RMS errors are also an improvement upon the previous methods of pre-processing the incremental image where no comparison of angles was attempted. The mean values for both models are also comparable.

5.3.3.1 Computation of Gradient Angle

The gradient values were obtained as described in chapter 4, using the central difference equations. Consider the reference image, f(x,y) and the sensed image, h(x,y) then the angle difference is given by

$$\text{residual angle} = \tan^{-1} G_1 - \tan^{-1} G_2$$

where G_1 = gradient angle in image 1 and G_2 = gradient angle in image 2.

The gradient angles are computed from the gradient along the x-axis, G_x and the gradient along the y-axis, G_y as

$$G_1 = G_{y1} / G_{x1}$$

$$G_2 = G_{y2} / G_{x2}$$

where

$$G_{x1} = \Sigma\, n.f(x + n,y) / \Sigma\, n^2$$

$$G_{y1} = \Sigma\, n.f(x,y + n) / \Sigma\, n^2$$

$$G_{x2} = \Sigma\, n.h(x + n,y) / \Sigma\, n^2$$

and $$G_{y2} = \Sigma\, n.h(x,y + n) / \Sigma\, n^2$$

for n = -2 to +2

To compute the absolute angle is computationally expensive and so since it is a comparison of angles that is of interest an alternative method is as follows

$$\text{difference angle, } d = \tan^{-1} G_1 - \tan^{-1} G_2$$

$$\tan d = \tan (\tan^{-1} G_1 - \tan^{-1} G_2)$$

but $\tan(a - b) = (\tan a - \tan b) / (1 - \tan a.\tan b)$

therefore $\tan d = (G_1 - G_2) / (1 - G_1.G_2)$

Since d is a fixed parameter in the algorithm tan d can be easily computed at the beginning of the algorithm. This saves computing the absolute gradient angles during the main part of the algorithm.

5.3.4 Application of Interpolation to Mis-matched Points

In the previous case when a mis-matched point was encountered, the incremental range value was allocated as zero. This technique, although effective, caused regions of zero incremental range values where the region was occluded to one of the camera stations. As a consequence there were regions in the range image which remained unaltered. So instead of allocating a zero value, an interpolation from surrounding points was investigated and the results obtained are shown in table 5.4. The interpolation was carried out as described in chapter 4.

	Max. Error	Mean	RMS	Match Value
Hilly Terrain Images	37.926	8.684	12.092	171717
Planar Images	104.817	53.915	60.662	1372996

Table 5.4 - Interpolation of Mis-Matched Points

The results indicate that the overall algorithm performance is improved once again since the mean and RMS errors are better than the previous case. The maximum error for the hilly terrain model is an improvement and for the planar model it is only slightly worse.

5.3.5 No Averaging on Incremental Image

In this test, a direct application of the incremental image, after applying the 5 degree criterion for location of mis-matched points, was used to compute the new range image. This means that no filtering was carried out on the incremental image. The results obtained are shown in table 5.5 and it can be seen that there is an improvement in the maximum, mean and RMS error when compared with the results of section 5.3.1 where no mis-match points were sought. But when comparing these results with those of section 5.3.4, the best results obtained so far, there is a deterioration in algorithm performance.

	Max. Error	Mean	RMS	Match Value
Hilly Terrain Images	55.107	9.490	13.150	70371
Planar Images	136.917	57.899	63.323	424179

Table 5.5 - No Averaging of Incremental Image

The idea of this test was that the mis-match point criterion would be sufficient to improve the algorithm performance and hence the smoothing of the incremental image may not be required. These results indicate that there is an improvement due to the mis-match point criterion, as compared with the results of

section 5.3.1 with a 1x1 averaging template, but there is still a need for some form of averaging of the incremental image.

5.3.6 The Case of Equal Match Values

In all the previous examples when calculating the optimal change in range value for a particular point, an equal individual match value was considered an improvement and the incremental step value appropriately calculated. In this test an equal match value was not assumed an improvement and the incremental step value was set to zero. The point to be investigated was that when an approximate match point is located in an area of virtually no texture, the algorithm would follow a path which may not necessarily improve the range estimate if the case of equal individual match values was considered as an improvement.

	Max. Error	Mean	RMS	Match Value
Hilly Terrain Images	37.724	8.226	11.861	166260
Planar Images	102.567	54.816	61.632	1432308

Table 5.6 - The Case of Equal Match Values

The results in this case (table 5.6) are the best obtained so far for the hilly terrain model and are comparable for the planar model. This indicates that the theory of not considering equal individual match values as an improvement in the estimate of the range value, provides a better algorithm performance.

5.4 Discussion

The results of the preceeding study indicate that the design of the algorithm, as described in this thesis, is valid and that it is possible to reduce the errors due to, for example, the mis-matching of image points by inspection of the intensity gradients on both stereo images at an approximate match position. The application of an exact mapping from one stereo image to another has enabled some of the problems involved with traditional area correlation techniques to be eliminated and hence the only source of errors that remains to be accounted for is that due to photometric distortion and occlusion of image points. A start has been made to eliminate occlusion errors by inspection of image intensity gradients but further analysis of the images will enable additional parameters to be examined and matched, to reduce the mis-match error. The problem of photometric distortion has been partly compensated for by averaging the incremental image.

The application of the algorithm to the hilly terrain model has helped to examine the effects of comparing intensity gradient angles to select mis-matched points. The interpolation of mis-matched points has also been examined and it has been shown that interpolation of mis-matched points does indeed improve algorithm performance. The planar model has helped to examine algorithm performance when presented with sharp changes in the model structure, an extreme condition which causes large amounts of feature occlusion. The analysis of these results also indicate that the algorithm performs better when an attempt is made to

reduce the mis-match error.

To sum up, the final form of the exact area registration algorithm is as follows. From an analysis of the reference image, the gradient image is extracted for purposes of pre-processing the incremental image. An approximation of the range image is estimated from the cross-correlation of small areas from the reference image with the sensed image, over a specified area of interest. A mapping is carried out from the reference image onto the sensed image and from a comparison with the actual sensed image, range improvement values are calculated to improve the initial approximation of the range image.

The incremental image is then pre-processed, this involves the location of corresponding points and then examining their intensity angles. If the intensity angles are within a given tolerance angle, the incremental range value is calculated in the usual manner. On the other hand if the residual angle is outside the tolerance angle, the incremental range value is interpolated from surrounding pixels. This is followed by smoothing the incremental image with the aid of the gradient image to select a variable averaging template size. Finally, a computation is made of the new match value. If the new match value is an improvement upon the previous match value then the new range image is computed and if there is no improvement then the step size is reduced and the incremental image re-calculated.

This process is carried out iteratively until the solution converges to an optimum range image from which the three-

dimensional model coordinates may be extracted. Convergence is assumed when the step size has been reduced to a pre-specified value.

6.0 Application of Stereo Ranging Algorithm to Real Scenes

In order to verify the technique of automatically acquiring three-dimensional information, with the aid of the stereo ranging algorithm, it was necessary to test both the camera calibration program and the algorithm on real scenes where the recorded stereo images are not ideal and where problems such as noise and photometric distortion are present.

The idea behind this chapter was to obtain results from the stereo ranging algorithm and compare these results with those obtained through normal photogrammetric procedures. Two stereo video images of a model were obtained by placing a CCD camera in two separate positions, sufficient to provide an approximate camera separation to range ratio of between 1/4 and 1/6. The images were obtained through the MV200 image processing system designed by Vision Dynamics Ltd., U.K., the size of the images being 256x256 pixels. Once obtained, the image coordinates of control points were extracted from both stereo images for purposes of camera calibration.

The photogrammetric measurements were obtained by producing video hard copies of the stereo images using a video hard copier provided by Prostab Ltd., U.K. The Computrace 1500 video hard copier produces a hard copy of a video image by segmenting each video line into 1024 pixels and passing the appropriate pixel intensity to an array of optical fibres which are arranged to focus their beams onto moving photographic paper. Hence no distortion is introduced into the hard copy as compared with

conventional video hard copiers based on flat cathode ray tubes.

Hard copies enable sub-pixel identification of corresponding stereo points as opposed to the stereo video system, consisting of the marker generator as a measurement tool, which only provides a limited accuracy of the order of half a pixel. By the process of accommodation, the operator is able to view the stereo model very comfortably even though the individual images may be restricted in quality due to a 256x256 pixel image size. In addition the video hard copies enable the transformation between the photogrammetric comparator coordinates and the digital comparator coordinates (merely an affine transformation) to be simplified, an important point to consider when comparing the results of the two processes.

The best possible accuracy in the photogrammetric measurements was obtained by using hard copies. Since these results were to be used as a standard in evaluation of the stereo ranging algorithm, it was essential to minimise the measurement error in the photogrammetric data.

6.1 Design of Control Field

Before any form of measurements could be carried out it was necessary to calibrate the camera system as accurately as possible. The camera geometry was determined by the application of the camera calibration program developed in chapter 2. To define as close as possible the camera geometry, it is absolutely necessary to provide a set of well distributed control points

with known object space coordinates.

It is possible to solve the calibration problem accurately with a limited number of control points but due to the quantisation error inherent in digitising the stereo images it is wise to use as many control points as possible.

Any form of control field may be implemented. However in order to achieve accurate triangulation of image coordinates it is necessary to have a variation in range of the order of 10-30% of the camera to object distance. In close range photography this condition would be difficult to achieve due to the depth of field restrictions. In the control field implemented up to 10% variation in range of control points was provided.

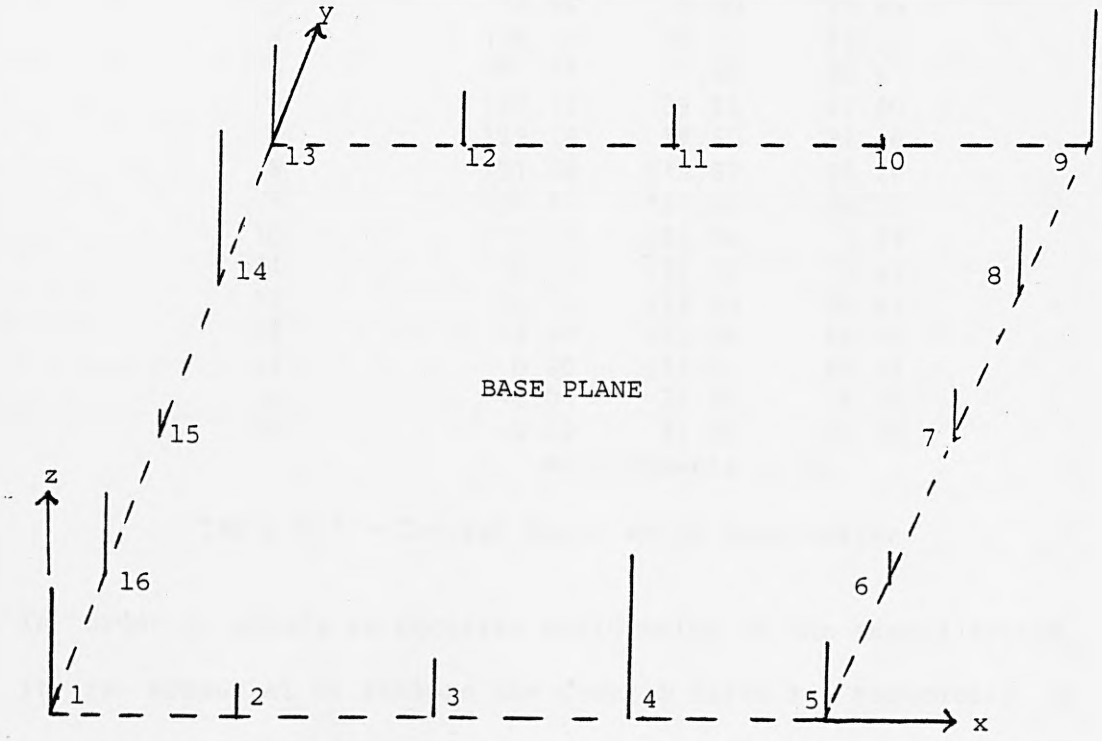


Figure 6.1 - Control Field for Camera Calibration

The control field was composed of a flat base with vertical pegs placed upon it such that the pegs consisted of a selection of differing heights as illustrated in figure 6.1. Sixteen pegs were placed randomly around the perimeter of the square 18x18cm base so that there was no situation where a series of control points (the tips of the pegs) formed a straight line in three-dimensional space. The control field was painted black and the tips of the pegs painted white to provide clearly visible control points on the stereo images.

6.1.1 Measurement of Control Point Coordinates

Control Point	X	Y	Z
1	0.00	0.00	50.00
2	38.26	-0.20	11.30
3	76.56	0.00	23.44
4	115.10	-0.24	64.23
5	151.74	1.60	30.61
6	152.12	39.98	11.40
7	152.18	77.60	21.16
8	151.88	115.62	26.86
9	150.47	151.59	54.37
10	113.69	153.06	3.84
11	72.12	152.75	15.46
12	36.73	152.04	20.68
13	-2.20	153.26	38.57
14	-0.60	114.50	60.14
15	-0.04	76.40	8.73
16	-0.02	37.48	34.44

measurements in mm

Table 6.1 - Control Point World Coordinates

In order to obtain an accurate calibration of the camera system, it is essential to measure the control field as accurately as possible so that there can be no doubt of the error in the solution due to errors in measurement of the control point object space coordinates. The control field was mounted on a milling

machine and a reference point was then manipulated to a position just above the tip of a peg and the three-dimensional coordinates noted. This procedure was repeated for all the pegs used as control points. The accuracy with which the milling tool could be positioned was 0.2mm, well within the accuracy required for all three orthogonal axes. The control point object space coordinates obtained are shown in table 6.1.

6.1.2 Calibration of Camera System

The camera used for the measurements was a CCD camera giving an image size of 256x256 pixels. So in the calibration equations derived in chapter 2, the systematic distortions due to lens aberration were not included since the random errors due to quantisation of the image coordinates were far greater and would swamp the systematic errors. This fact was confirmed by including the afore mentioned parameters in the calibration program, whence it was noted that the distortion parameters became dominant in the solution and made it very difficult for the Newton-Raphson method to converge when close to the solution.

Hence the only internal camera orientation parameters catered for were those due to differential scale changes along axes, non-orthogonality of the axes, shift of axes and the principal distance.

The procedure adopted for determination of the internal and external camera geometry was to locate image coordinates of all visible control points and then applying the camera calibration

program developed in chapter 2. In order to increase the accuracy of the solution the fact that the photogrammetric comparator coordinates were more accurately measured than the corresponding digital images was used to an advantage. The photogrammetric comparator coordinates were first calibrated and then the computed camera station coordinates and rotation about the x and y axes were used as a fixed parameter in the solution of the digital image solution. This is possible because the video hard copies were duplicates of the digital images and hence the only difference between the two sets of comparator coordinates lies in the internal camera orientation and the z-axis rotation. As a further improvement in precision the digital images were calibrated simultaneously by a simple modification of the camera calibration program because it was known that, apart from equivalent focal lengths, the internal camera orientation was exactly the same for both the left and right images (since they were taken with the same camera).

Finally to confirm the calibration parameters for both the digital images and the hard copies, the range values for all the control points, given the conjugate stereo points, were computed and compared with the true range values. The measured range values were obtained by triangulation of the image coordinates as previously described in chapter 5. The true range values were obtained by computing the z-coordinate of the vector from the scene point to the camera perspective centre, since the normal vector of the reference base plane was chosen to be along the z-axis. This corresponds to the reference plane being the base of

the control field.

6.2 Obtaining Measurements from a Stereo Pair of Images

To obtain measurements from a pair of stereo images and confirm results obtained through application of the stereo ranging algorithm, it was decided to use the photogrammetric results as a standard. The results obtained through application of the algorithm were then compared with this standard.

6.2.1 Obtaining Photogrammetric Measurements

It was possible to obtain measurements as in conventional photogrammetry by the use of a stereo comparator. A stereo comparator essentially enables comfortable viewing of a stereo model and within this equipment is incorporated a conventional floating mark for aid in the extraction of conjugate stereo points. The pairs of image coordinates were then used to obtain a corresponding range value, through application of the formulae derived in chapter 4.

To reconstruct a three-dimensional scene an area of interest is selected in the scene and the points of major change such as edges, corner points and so forth are selected and their conjugate stereo points extracted. Once the selected area has been sufficiently covered the corresponding points are triangulated to obtain their three-dimensional coordinates and then interpolated to provide a representation of the stereo model surface, using a suitable interpolating routine.

6.2.2 Algorithm Application

The stereo ranging algorithm was applied to a small area selected in the reference image and the configuration of the algorithm was exactly as that described in chapter 4. But as opposed to the computer generated images, the real images did have photometric distortion present and so to minimise this form of distortion the mean and variance of the stereo pair were equalised.

6.2.2.1 Photometric Equalisation of a Stereo Pair of Images

The best and simplest method of equalising the photometric distortion between two images is to equalise their mean and variance. The mean and variance of two images may be equalised by computing the mean, m_1 and variance, σ_1^2 of image one and the mean, m_2 and variance, σ_2^2 of image two. From the variances the image with the largest variance is selected and the other image is then scaled with respect to the former. For example, if the second image had the greater variance then the first image would have its pixel intensities modified using the following formula.

$$\text{new intensity} = (\text{old intensity} - m_1).(\sigma_2 / \sigma_1) + m_2$$

Application of this equation to all points on image one would equalise the mean and variance between the stereo images.

As a further advantage of equalisation it can be shown (Merchant, 1981) that the classical two dimensional correlation function, $R(u,v)$

$$R(u,v) = \frac{\iint (f(x,y) - g(x-u,y-v))dxdy}{\sqrt{\iint f^2(x,y)dxdy . \iint g^2(x-u,y-v)dxdy}}$$

is related to the corresponding normalised sum of difference squared function, S(u,v)

$$S(u,v) = \frac{\iint (f(x,y) - g(x-u,y-v))^2 dxdy}{\iint f^2(x,y)dxdy}$$

by the equation

$$S(u,v) = 2.(1 - R(u,v))$$

This means that the classical correlation can be carried out by simply equalising the mean and variance of the stereo images and then computing the sum of difference squared function which requires only a difference function and a table lookup for the square function (i.e. it does not require a multiply function).

6.2.3 Comparison of Results

The results of the stereo ranging algorithm and the results obtained through conventional photogrammetry can only be compared when the photogrammetric comparator coordinates, (x_p, y_p) have been translated into the corresponding digital image coordinates, (x_d, y_d). The transformation between the two sets of coordinates requires accounting for differential scale deformation, translation of principal point, non-orthogonality of axes and finally rotation of axes. These parameters can be accounted for by the use of an affine transformation of the form

$$x_d = s.x_p + t.y_p + u$$

and

$$y_d = l.x_p + m.y_p + n$$

The solution of the parameters s,t,u,l,m and n was carried out by application of the Newton-Raphson method as detailed in Appendix C. The previously determined control point comparator coordinates were used to solve for the unknowns in the above equations.

Upon computation of the above parameters it was possible to transform the photogrammetric reference image coordinates into the corresponding digital reference image coordinates. Therefore given the corresponding range values obtained from the photogrammetric measurements, it was possible to apply an interpolation routine to these results and hence obtain range values for all pixel points within the selected area on the digital reference image. This interpolated surface, $R_p(x,y)$ was then compared with the range image, $R_a(x,y)$ obtained through application of the stereo ranging algorithm and the maximum, mean and RMS errors were then extracted as follows.

$$\text{mean} = \Sigma|(r_a(x,y) - r_p(x,y))| \ / \text{ number of pixels}$$

$$\text{RMS} = \sqrt{[\Sigma(r_a(x,y) - r_p(x,y))^2 \ / \text{ number of pixels}]}$$

6.3 Evaluation of Algorithm Performance

This section deals with application of the algorithm to three stereo models. The models chosen were - a robot, a truck and a Lego model. The following sections give results of camera calibration and algorithm performance when applied to the stereo models and a discussion is left until the end. Plates of the stereo models may be found at the beginning of the thesis, the quality poor since the pictures are from TV monitor.

6.3.1 Lego Model

The Lego model consisted of Lego bricks assembled to resemble a model aircraft. A stereo pair of images for this model are shown in plate 6.1.

6.3.1.1 Calibration of Stereo Images

Upon calibration of the video hard copies the results obtained were as given in table 6.2. The calibration data is as shown below.

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	0.000	0.000	0.000	0.000	611.5	608.2
2	24.063	3.008	29.600	2.377	650.2	647.3
3	46.287	1.847	50.685	0.713	638.1	634.3
4	71.448	-2.023	71.062	-3.942	597.3	593.9
5	90.661	1.038	95.682	-1.299	630.9	630.4
6	89.678	19.634	97.031	17.267	650.1	649.5
7	91.576	36.448	97.456	33.755	640.4	636.4
8	92.540	53.532	97.843	51.128	634.7	631.2
9	94.500	70.889	96.564	68.577	607.2	605.8
10	69.402	70.388	76.526	68.302	657.7	657.8
11	46.798	70.536	52.135	69.313	646.0	647.5
12	24.541	70.625	28.899	69.860	640.8	637.5
13	0.470	71.807	71.492	-2.200	623.0	620.0
14	-0.089	53.579	-0.830	53.698	601.4	599.3
15	2.637	36.560	8.631	36.688	652.8	651.9
16	0.555	18.082	3.903	18.073	627.1	630.6

RMS error in range = 3.1mm

Table 6.2 - Lego Model - Calibration Confirmation, Photogrammetric Images

Left Camera

Perspective centre (52.777, 114.631, 661.527)

Rotation (-0.07506, 0.02917, 0.01982) in radians

Perspective distance = 376.113

Internal camera orientation parameters

b = 22.012 c = 0.00678 d = 0.76883 e = 31.802

b' = -22.012 c' = -0.00882 d' = 1.30068 e' = -41.171

Right Camera

Perspective centre (187.452, 117.013, 648.653)

Rotation (-0.07344, 0.03862, 0.03719) in radians

Perspective distance = 383.776

Internal camera orientation parameters

b = 102.052 c = -0.00626 d = 0.74720 e = 30.036

b' = -102.052 c' = -0.00837 d' = 1.33833 e' = -41.052

In a similar manner the digital images were calibrated and the results so obtained are shown in table 6.3. The calibration parameters are shown below.

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	119	192	23	195	611.6	613.9
2	149	187	58	189	650.2	650.0
3	177	190	83	191	638.1	632.7
4	208	198	108	199	597.3	597.7
5	232	193	138	193	630.9	637.2
6	231	157	139	158	650.1	658.0
7	234	126	139	127	640.4	644.8
8	236	94	139	94	634.7	638.7
9	239	61	137	61	607.2	614.7
10	207	61	112	62	657.7	656.0
11	179	60	84	62	646.1	653.3
12	151	59	55	62	640.9	644.0
13	121	56	23	60	623.0	628.6
14	120	90	20	94	601.4	609.8
15	123	123	32	126	652.8	660.6
16	120	158	27	161	627.1	639.4

RMS error in range = 6.2mm

Table 6.3 - Lego Model - Calibration Confirmation, Digital Images

Left Camera

Perspective centre (52.777, 114.631, 661.527)

Rotation (-0.07506, 0.02917, 0.03719) in radians

Perspective distance = 469.158

Right Camera

Perspective centre (187.452, 117.013, 648.653)

Rotation (-0.07344, 0.03862, 0.01589) in radians

Perspective distance = 459.797

Common internal camera orientation parameters

b = 147.072 c = -0.00594 d = -1.16797 e = 132.624

b' = -147.072 c' = -0.00508 d' = -0.08562 e' = 113.551

6.3.1.2 Application of Algorithm

The stereo ranging algorithm was applied to an area 37x37 pixels on the reference image, starting at (169,89), and the results of the algorithm when compared with those of photogrammetry are provided below.

RMS error in range = 6.4mm or approximately 1.4% of range

Average absolute error in range = 10.1mm or approximately 2.1% of range

Maximum error in range = 31.6mm or approximately 6.7% of range

6.3.2 Truck Model

The truck model consisted of a toy truck with loaded goods so that there was sufficient variation in range in the model. A stereo pair of images are shown in plate 6.2.

6.3.2.1 Calibration of Stereo Images

Upon calibration of the video hard copies the results obtained were as given in table 6.4. The calibration data is shown below.

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	0.000	0.000	0.000	0.000	591.8	589.6
2	26.675	3.545	30.423	2.836	630.5	622.9
3	50.151	2.902	53.871	1.659	618.4	615.3
4	74.949	-1.551	75.202	-3.357	577.6	579.6
5	96.729	1.979	100.387	-0.367	611.2	607.5
6	96.095	21.770	101.552	19.471	630.4	625.1
7	96.987	39.234	101.785	36.913	620.7	615.3
8	97.242	57.413	101.874	55.197	615.0	610.3
9	98.817	75.798	100.931	73.351	587.5	583.9
10	71.952	74.957	79.821	73.020	637.7	640.5
11	47.920	75.073	53.371	73.734	626.4	622.6
12	24.777	74.691	29.099	73.933	621.1	616.4
13	-1.387	75.822	1.321	75.190	603.3	599.6
15	2.963	39.173	7.850	38.740	633.1	631.1
16	1.569	19.605	2.678	19.135	607.4	600.9

RMS error in range = 4.4mm

Table 6.4 - Truck Model - Calibration Confirmation, Photogrammetric Images

Left Camera

Perspective centre (80.053, 116.757, 641.828)

Rotation (-0.06649, 0.03924, 0.00142) in radians

Perspective distance = 386.512

Internal camera orientation parameters

b = 36.708 c = -0.00038 d = 0.76587 e = 38.226

b' = -36.708 c' = 0.00050 d' = 1.30570 e' = -49.930

Right Camera

Perspective centre (177.867, 119.131, 654.032)

Rotation (-0.06318, 0.03489, 0.02537) in radians

Perspective distance = 405.236

Internal camera orientation parameters

b = 104.317 c = -0.00223 d = 0.74771 e = 37.548

b' = -104.317 c' = 0.00298 d' = 1.33742 e' = -50.528

In a similar manner the digital images were calibrated and the results so obtained are shown in table 6.5. The calibration data is shown below.

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	103	203	23	203	591.8	591.7
2	136	196	59	197	630.5	623.0
3	165	197	87	199	618.4	611.6
4	196	205	113	208	577.6	575.7
5	223	198	144	202	611.2	607.8
6	222	161	145	165	630.4	625.8
7	223	128	145	132	620.7	615.0
8	223	94	145	97	615.0	607.0
9	225	59	144	63	587.5	582.0
10	192	61	118	64	638.0	635.0
11	163	61	87	63	626.4	621.7
12	133	62	57	63	621.1	613.6
13	101	60	24	61	603.3	602.8
15	106	129	32	130	633.1	635.9
16	105	166	26	167	607.4	601.7

RMS error in range = 5.1mm

Table 6.5 - Truck Model - Calibration Confirmation, Digital Images

Left Camera

Perspective centre (80.053, 116.757, 641.828)

Rotation (-0.06649, 0.03924, -0.00179) in radians

Perspective distance = 478.545

Right Camera

Perspective centre (177.867, 119.131, 654.032)

Rotation (-0.06318, 0.03489, 0.02073) in radians

Perspective distance = 489.508

Common interior camera orientation parameters

b = 148.298 c = -0.00420 d = -1.16635 e = 130.555

b' = -148.298 c' = -0.00360 d' = -0.85738 e' = 111.935

6.3.2.2 Application of Algorithm

The stereo ranging algorithm was applied to an area 25x25 pixels on the reference image, starting at (150,131), and the results of the algorithm when compared with those of photogrammetry are provided below.

RMS error in range = 5.3mm or approximately 1.1% of range

Average absolute error in range = 11.6mm or approximately 2.4% of range

Maximum error in range = 27.4mm or approximately 5.7% of range

6.3.3 Robot Model

The robot model consisted of a toy robot with a large amount of detail in the design so that sufficient image quality could be obtained as well as range varaition. A stereo pair of images are shown in plate 6.3.

6.3.3.1 Calibration of Stereo Images

Upon calibration of the video hard copies the results obtained were as given in table 6.6. The calibration data is shown below.

Left Camera

Perspective centre (53.251, 116.631, 660.101)

Rotation (-0.08880, 0.02199, 0.03543) in radians

Perspective distance = 374.817

Internal camera orientation parameters

b = 22.131 c = 0.01018 d = 0.76826 e = 31.796

b' = -22.131 c' = -0.01325 d' = 1.30164 e' = -41.093

Right Camera

Perspective centre (180.185, 119.906, 653.278)

Rotation (-0.07927, 0.03213, 0.03709) in radians

Perspective distance = 384.738

Internal camera orientation parameters

b = 101.871 c = -0.00156 d = 0.73779 e = 29.994

b' = -101.871 c' = 0.00211 d' = 1.35540 e' = -40.870

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	0.000	0.000	0.000	0.000	611.5	609.4
2	24.176	3.264	29.582	2.637	650.2	648.1
3	46.245	2.019	50.134	1.584	638.1	635.7
4	71.270	-1.763	73.079	-2.747	597.3	594.7
5	90.553	1.273	95.477	-1.066	630.9	630.0
6	89.315	19.711	96.631	17.293	650.1	650.2
7	91.553	36.499	96.972	36.364	640.4	637.3
8	92.629	53.504	97.403	54.100	634.7	631.8
9	94.515	70.763	96.742	68.146	607.2	608.8
10	69.469	70.298	74.560	71.433	657.7	654.8
11	46.980	70.478	52.719	68.793	646.1	646.4
12	24.725	70.498	26.243	71.778	640.8	634.2
13	0.629	71.727	2.440	70.644	623.0	624.4
14	0.055	53.194	-1.272	52.610	601.4	600.6
15	2.609	36.627	7.610	36.019	652.8	649.5
16	1.180	18.113	2.886	17.686	627.1	622.6

RMS error in range = 2.8mm

Table 6.6 - Robot Model - Calibration Confirmation, Photogrammetric Images

In a similar manner the digital images were calibrated and the results so obtained are shown in table 6.7. The calibration data is shown below.

Left Camera

Perspective centre (53.251, 116.631, 660.101)

Rotation (-0.08880, 0.02199, 0.03543) in radians

Perspective distance = 468.227

Right Camera

Perspective centre (180.185, 119.906, 653.278)

Rotation (-0.07927, 0.03213, 0.02262) in radians

Perspective distance = 463.685

Common interior camera orientation parameters

b = 150.753 c = -0.00486 d = -1.16828 e = 138.643

b' = -150.753 c' = -0.00416 d' = -0.85596 e' = 118.673

Control Point	Left Image		Right Image		Range Value	
	x'	y'	x'	y'	True	Meas.
1	119	192	29	199	610.1	613.9
2	149	187	64	193	648.8	651.4
3	177	190	89	196	636.7	632.0
4	208	198	114	204	595.9	594.2
5	232	193	144	198	629.5	635.1
6	231	158	145	163	648.7	655.6
7	234	126	145	132	638.9	640.4
8	236	93	145	99	633.2	632.8
9	239	61	144	66	605.7	612.8
10	207	61	119	67	656.3	657.4
11	179	60	90	66	644.6	648.1
12	151	59	61	67	639.4	638.7
13	121	56	30	64	621.5	629.6
14	120	91	26	98	600.0	604.8
15	123	123	37	130	651.4	653.1
16	121	158	32	165	625.7	626.2

RMS error in range = 4.2mm

Table 6.7 - Robot Model - Calibration Confirmation, Digital Images

6.3.3.2 Application of Algorithm

The stereo ranging algorithm was applied to an area 25x25 pixels

on the reference image, starting at (164,105), and the results of the algorithm when compared with those of photogrammetry are provided below.

RMS error in range = 4.5mm or approximately 1.0% of range
Average absolute error in range = 9.3mm or approximately 2.0% of range
Maximum error in range = 23.7mm or approximately 5.0% of range

6.4 Discussion

From the results presented in this section it has been shown that reasonable results may be obtained from the stereo ranging algorithm, even with an image size of 256x256 pixels. The accuracy in general seems to be of the order of 1% of range since the average distance to the scene in all stereo models was just over 600mm and the RMS error in range was of the order of 6mm. This means that the maximum error is of the order of 5% since the maximum errors were of the order of 30mm. The maximum errors were probably due to stray points which were mis-matched and so this indicates that even though attempts were carried out to eliminate mis-matched points the procedures followed are not sufficient and further features should be matched such as gradient magnitude comparisons and so forth.

With the advent of the home video market, a drop in the cost of quality video systems has been noticed and hence it has become possible to purchase high quality video cameras and monitors at modest prices. This includes the dramatic reduction in

manufacturing costs of solid state cameras which are ideal for photogrammetric applications because of their small size, low power consumption, wide dynamic range and immunity to the various types of scanning distortions inherent in vidicon cameras to name but a few.

Plans to almost double the number of scan lines per frame of video for the next generation of television transmission standards has seen the introduction of high definition video cameras and monitors at reasonable cost. Therefore it should be possible to obtain better accuracies in measurement in the near future with only a moderate increase in system cost.

A further increase in accuracy can also be achieved by introducing zoom lenses into the stereo imaging system so that detailed surface inspection can be carried out. Considering all the afore mentioned factors, it should be possible to easily increase the accuracies achievable by a factor of 2 to 4.

Although the accuracies achievable using the stereo video system are poor when compared with conventional still picture photogrammetry, the flexibility, ease of use, short turn around times and low costs make the stereo video measurement system attractive in many situations including those requiring measurements in hazardous areas where conventional photogrammetry would not be possible.

The computational requirements of the stereo ranging algorithm are not extensive and a small stand alone micro-computer costing a few thousand pounds can easily suffice. Inclusion of

appropriate graphics software and a plotter would provide a very useful interactive measurement tool.

7.0 Conclusions

This thesis has been concerned with the application of stereo image processing techniques to photogrammetry to reduce the time required to obtain results from a stereo pair of images. As compared to conventional photogrammetry, where an operator is required to extract pairs of image coordinates from a stereo scene, the stereo ranging algorithm has been designed with the view of eliminating the operator. Although results cannot be achieved to the accuracy obtained in conventional photogrammetry, with a pair of video cameras, the algorithm designed enables the measurement system to be highly flexible with short turn around times (of the order of half an hour to execute the stereo ranging algorithm on a 40x40 pixel area). It is also possible to inspect several scenes by storing pairs of images from several scenes for later analysis.

A review of the various types of systematic errors that may occur in a stereo video system has been carried out and it has been shown how these errors can be mathematically modelled and corrected. A program has been developed to compute the various distortion parameters from known control points. The program developed enables calibration to be carried out using the collinearity condition but use of the coplanarity condition as well can reduce the number of control points required with known object space coordinates. Hence calibration using this alternative technique is highly recommended and the equations derived in chapter 2 implemented.

The City University stereo video measuring system has been described and it has been shown how measurements may be made with the aid of an operator. The system design to provide a minimal correction in systematic distortion has been fully described when an electronic floating light mark is introduced into the Interpretoskop. The next stage would be to introduce CCD cameras and incorporate the electronic floating mark into the control circuitry of the camera system. This would enable synchronisation of measurements made through both photogrammetry and use of the algorithm. The system would also be easier to interface with a computer and so the coordinate system of the Interpretoskop would be the same as that of the computer. Hence comparator coordinates of control points can be used to calibrate the photogrammetric system as well as the digital system, without requirement for a relative transformation between the two system measurement axes.

Finally, it has been shown in this thesis that it is possible to acquire three-dimensional data from a stereo pair of images using an iterative technique of transforming one stereo view into the other and then comparing. The technique starts with an initial approximation of the stereo model, in the form of a range image associated with one of the stereo images, and then improves upon this model iteratively. Upon each transformation of one view to the other, the transformed image is compared with the actual and an improvement of the stereo model is carried out. This process is repeated until the constructed model has converged.

The main assumptions of the algorithm have been to assume that the spectral characteristics of the scene being viewed are

uniform and that when the two stereo images are matched, the sum of residual intensity squared is a minimum. A further criterion of the algorithm has been to assume that the surface being inspected is smooth and does not possess any abrupt changes. This has been incorporated into the algorithm by allowing changes to the range image such that adjacent pixels have a close correlation of range values.

In the design of the algorithm it was found that pre-processing the incremental range image, by passing a variable smoothing template over it, improved the results. Although the results did improve, there must be an optimum selection method for the variable smoothing template size. In the algorithm the template size was selected on the basis of gradient analysis but a better predictor might be to use local statistical characteristics such as standard deviation. Where the standard deviation is large, the template size can be small and where the standard deviation is small, the template size can be large. A further point to consider is whether the relationship between the standard deviation and template size should be linear or non-linear.

During the design of the algorithm it has also been found that the performance can be improved by comparing more than just one feature from the images (such as intensity residual) and so investigations should be carried out in the use of matching further features such as gradient strength as well as gradient angle. The use of multiple features reduces the chances of error by minimising the dependence on any one property.

A further point to consider is the segmentation of the reference image into areas enclosed by edges and carrying out an initial search for corresponding areas. The algorithm may then be applied with a restriction that range values of points, within an area, are limited so that points within this area are restricted to map onto the corresponding area on the conjugate image. The extraction of edges is a different matter on its own but a recent article introduced the concept of extracting edges with the use of both an intensity image and its corresponding range image (Mitiche and Aggarwal, 1982) and this technique should be investigated with the aim of extracting edges from the initial estimated range image and its associated intensity image.

The design of the algorithm has been carried out using a set of computer generated images and it has been shown how the algorithm has been improved from a basic structure. The use of computer generated images has enabled the inspection of different variations of the algorithm without concern for image signal noise and photometric distortion. A tool which would be usefull in the refinement of the algorithm is the ability to display and output a three-dimensional image of the structure being analysed. It would then be possible to inspect performance of the algorithm from one iteration to another and any potential problems highlighted.

Having applied the algorithm to a set of real scenes the results obtained indicate that fairly accurate results are achievable (of the order of 1% of the camera to object distance of about 650mm, the focal length being about 460mm) for a 256x256 element image.

Increasing the focal length would enable a finer resolution for inspection of a smaller area of the scene.

Considering the hardware implementation, since the algorithm operates on a local optimisation approach to reach a global optimum it is suited to a parallel architecture. This form of architecture can have each processor dedicated to a single pixel or a small area of one stereo image with access to a small portion of the corresponding area in the second stereo image. This sub-image could be stored in each processor or made available to it by its neighbouring processors.

A further improvement would be to guide the stereo ranging algorithm from an knowledge based system with the view of increasing the accuracy from a knowledge of the type of structure being inspected. In addition the error of a false match may be reduced or detected before completion of the algorithm and hence corrective action taken from one iteration of the algorithm to another.

Glossary

Aberration - failure of an optical system to bring all light rays received from a point object to a single image point or to a prescribed geometric position.

Accomodation - the faculty of the human eye to adjust itself to give sharp images for different object distances.

Active system - a system which transmits an electromagnetic signal. A system with the capability to transmit, repeat or re-transmit electromagnetic information.

Adjustment - the determination and application of corrections to observations for the purpose of reducing errors or removing internal inconsistencies in derived results.

Calibration - the art or process of determining certain specific measurements in a camera, or other instrument or device by comparison with a standard, for use in correcting or compensating errors.

Camera station - the point in space occupied by the camera lens at the moment of exposure.

Comparator - an optical instrument for measuring rectangular or polar coordinates of points on any plane surface.

Comparator coordinates - the coordinates of an image point measured in the measurement instrument coordinate axes.

Corresponding points - conjugate points - the images of the same object point on two or more photographs.

Corresponding image rays - rays connecting each of a set of corresponding image points with its particular perspective centre.

Floating mark - a mark seen as occupying a position in three dimensional space formed by the stereoscopic fusion of a pair of photographs and used as a reference mark in examining or measuring the stereoscopic model.

Image coordinates - the coordinates of a point in the absolute geometric photographic image plane.

Incremental image - the image composed of minute incremental range changes which when added to the range image, improve the range image.

Individual match value - the residual intensity difference obtained when a point in the reference image is mapped onto a point in the sensed image.

Match value - the overall sum of absolute residuals squared when an area in the reference image is mapped onto the sensed image.

Observation equation - a condition equation which connects interrelated unknowns by means of an observed function, or a condition equation connecting the function observed and the unknown quantity whose value is sought.

Optical axis - in a lens element, the straight line which passes through the centres of curvature of the lens surfaces.

Parallax - the apparent displacement of the position of a body, with respect to a reference point or system, caused by a shift in the point of observation.

Passive system - a system which records energy emitted or reflected but which does not produce or transmit energy of its own.

Perspective centre - the point of origin or termination of bundles of perspective rays.

Principal distance - the distance measured along the optical axis from the rear nodal point of the lens to the plane of critical focus of a very distant object. Also known as the equivalent focal length.

Principal point - the foot of the perpendicular from the interior perspective centre to the plane of a photograph.

Range image - the image composed of range values associated with every pixel in the corresponding intensity image.

Reference image - the image used as a reference in the stereo matching algorithm.

Relative orientation - the geometric orientation between two camera stations.

Sensed image - the second image of a pair of stereo images used in the stereo matching algorithm. The reference image is transformed into this image.

Step size - the size of a step taken in inspection of changes to a range value in the range image.

Stereoscopy - the ability of the eyes to bring two images into superimposition for stereoscopic viewing.

Triangulation - the merging of corresponding rays from two camera stations in three dimensional object space.

References and Bibliography

ABDEL-AZIZ, Y.I. and KARARA, H.M. (1974)
Photogrammetric Potentials of Non-Metric Cameras.
Civil Engineering Studies, Photogrammetry Series. Number 36.
(Illonois: University of Illonois).

ARNOLD, R.D. and BINFORD, T.O. (1980)
Geometric Constraints in Stereo Vision.
In: IMAGE PROCESSING FOR MISSILE GUIDANCE, 1980. Proceedings.
Society of the Photo-optical Instrumentation Engineers.
Vol. 238: pp. 281-90.

BAKER, H.H and BINFORD, T.O. (1981)
Depth from Edge and Intensity Based Stereo.
In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE,
7th, 1981. Proceedings. Vol 2: pp. 631-6.

BALLARD, D.H. and BROWN, C.M. (1982)
Computer Vision.
(New Jersey: Prentice-Hall, Inc.)

BARNARD, S.T. and THOMPSON, W.B. (1980)
Disparity Analysis of Images.
Transactions on Pattern Analysis and Machine Intelligence.
Institute of Electrical and Electronic Engineers.
Vol. PAMI-2 (4): pp. 333-40.

BENARD, M. (1983)
Extracting 3-D Coordinates of an Object from a Digital
Stereopair: an automated method.

In: SIGNAL PROCESSING II: THEORIES AND APPLICATIONS, Erlangen, Germany, 1983. Proceedings. EUSIPCO pp. 227-30.

BENNETT, V.P. and BALASUBRAMANIAN, N. (1973)
Detection of Image Coincidence using Correlation Coefficient Measurement.
In: AMERICAN SOCIETY OF PHOTOGRAMMETRY, 39th, Washington D.C., 1973. Proceedings. pp. 1-19.

BROWN, D.C. (1971)
Analytical Calibration of Close-Range Cameras.
In: SYMPOSIUM ON CLOSE-RANGE PHOTOGRAMMETRY, Melbourne, Florida, 1971. Proceedings.

BROWN, D.C. (1972)
Calibration of Close Range Cameras.
In: CONGRESS OF THE INTERNATIONAL SOCIETY OF PHOTOGRAMMETRY, 12th, Ottawa, Canada, 1972.

BURR, D.J. and CHIEN, R.T. (1977)
A System for Stereo Computer Vision with Geometric Models.
In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 5th, 1977. Proceedings. pp. 583.

CARNAHAN, B. and WILKES,J.O. (1973)
Digital Computing and Numerical Analysis.
(New York: John Wiley and Sons, Inc.).
pp. 355-56 and 365-69.

CHONG, A.M., BECK, M.S., BROWN, M.B., MacKENZIE, I.S. and RITCHINGS, R.T. (1982)

Micro-computer Based Stereoscopic Rangefinding.

In: CONFERENCE, MANCHESTER 1982. Proceedings.

Institute of Electrical and Radio Engineers. pp. 95-102.

CONRADY, A.E. (1919)

Decentered Lens System.

Monthly Notices of the Royal Astronomical Society. Vol. 79 (5).

DOWMAN, I.J. and HAGGAG A. (1977)

Digital Image Correlation Along Epipolar Lines.

In: INTERNATIONAL SYMPOSIUM ON IMAGE PROCESSING INTERACTIONS WITH PHOTOGRAMMETRY AND REMOTE SENSING, GRAZ. 1977. Proceedings. pp. 47-9.

FALCONER, K.J. (1971)

A General Purpose Algorithm for Contouring Over Scattered Data Points.

(National Physical Laboratory: Division on Numerical Analysis and Computing Report No. NAC6).

FRANKE, G. (1966)

Physical Optics in Photography.

(London: Focal Press).

GILMORE, J.F. and ROWLAND, P.J. (1983)

Mean Neighbours - a Point Pattern Matching Algorithm.

In: SOUTHEASTCON CONFERENCE ON TECHNOLOGY CHALLENGES IN THE NEW SOUTH, Orlando, Florida 1983. Proceedings. pp. 317-21.

HALL, E.L., TIO, J.B.K. and MacPHERSON, C.A. (1982)

Measuring Curved Surfaces for Robot Vision.

Computer. Vol.15 (12): pp. 42-54.

HARDY, J.G. and ZAVODNY, A.T. (1981)
Automatic Reconnaissance-Based Target-Coordinate Determination.
In: TECHNIQUES AND APPLICATIONS OF IMAGE UNDERSTANDING, 1981.
Proceedings. Society of the Photo-optical Instrumentation
Engineers. Vol. 281: pp. 95-104.

HERMAN, M., KANADE, T. and KUROE, S. (1984)
Incremental Aquisition of a Three-Dimensional Scene Model from
Images.
Transactions on Pattern Analysis and Machine Intelligence.
Institute of Electrical and Electronic Engineers.
Vol. PAMI-6 (3): pp. 331-40.

JULESZ, B. (1971)
Foundations of Cyclopean Perception.
(Chicago: University of Chicago Press).

LANCZOS, C. (1967)
Applied Analysis.
(London: Sir Issac Pitman and Sons, Limited).
pp. 321-24.

LEVINE, M.D., O'HANLEY, D.A. and YAGI, G.M. (1973)
Computer Determination of Depth Maps.
Computer Graphics and Image Processing.
Vol. 2: pp.131-50.

LINDSEY, N.E., BABAEI-MAHANI, R. and SHAH, Y. (1985)
Development of a Stereo TV System for Recording and Measurement.

Optics and Lasers in Engineering

Vol. 6 (2): pp.111-117.

LLOYD, S.A. (1985)

A Dynamic Programming Algorithm for Binocular Stereo Vision.

GEC Journal of Research. Vol. 3 (1): pp. 18-24.

MAHEW, J.E.W. and FRISBY, J.P. (1980)

The Computation of Binocular Edges.

Perception Vol. 9: pp. 69-86.

MARR, D. and POGGIO, T. (1976)

Cooperative Computation of Stereo Disparity.

Science Vol. 194: pp. 283-7.

MARR, D. and POGGIO, T. (1977)

A Theory of Human Stereo Vision.

(MIT, Artificial Intelligence Lab., AI Memo 451).

MITICHE, A., GIL, B. and AGGARWAL, J.K. (1982)

On Combining Range and Intensity Data.

Pattern Recognition Letters. Vol. 1 (2): pp. 87-92.

NAGATA, S. (1982)

New Versatile Stereo (NS-type) Display System and Measurement of Binocular Depth Perception.

Japan, Journal of Medical Electronics and Biological Engineering. Vol. 20 (3): pp. 154-61.

NISHIHARA, H.K. and LARSON, N.G. (1981)

Towards a Real Time Implementation of the Marr and Poggio Stereo

Matcher.

Techniques and Applications of Image Understanding.

Society of Photo-optical and Instrumentation Engineers.

Vol. 281: pp. 299-303.

POSDAMER, J.L. and ALTSCHULER, M.D. (1982)

Surface Measurement by Space-Encoded Projected Beam Systems.

Computer Graphics and Image Processing Vol. 18: pp. 1-17.

REAL, R.R and FUJIMOTO, Y. (1984)

Digital Processing of Dynamic Imagery for Photogrammetric Applications.

Transactions on Instrumentation and Measurement.

Institute of Electrical and Electronic Engineers.

Vol. IM-33 (1): pp. 45-51.

SHAH, Y., CHAPMAN, R., BABAEI-MAHANI, R. and LINDSEY, N.E. (1985)

Extraction of Range Information from Stereo Images.

Optics and Lasers in Engineering.

Vol. 6 (2): pp. 125-127.

SIDNEY, L. (1981)

Geometric Constraints for Interpreting Images of Common Structural Elements: Orthogonal Trihedral Vertices.

In: TECHNIQUES AND APPLICATIONS OF IMAGE UNDERSTANDING, 1981. Proceedings. Society of the Photo-optical Instrumentation Engineers. Vol. 281: pp. 332-38.

WONG, K.W. (1968)

Geometric Calibration of Television Systems for Photogrammetric

Applications.

Civil Engineering Studies, Photogrammetry Series. Number 16.

(Illonois: University of Illonois).

WONG, K.W. (1980)

Basic Mathematics of Photogrammetry.

In: Manual of Photogrammetry, Fourth Edition,

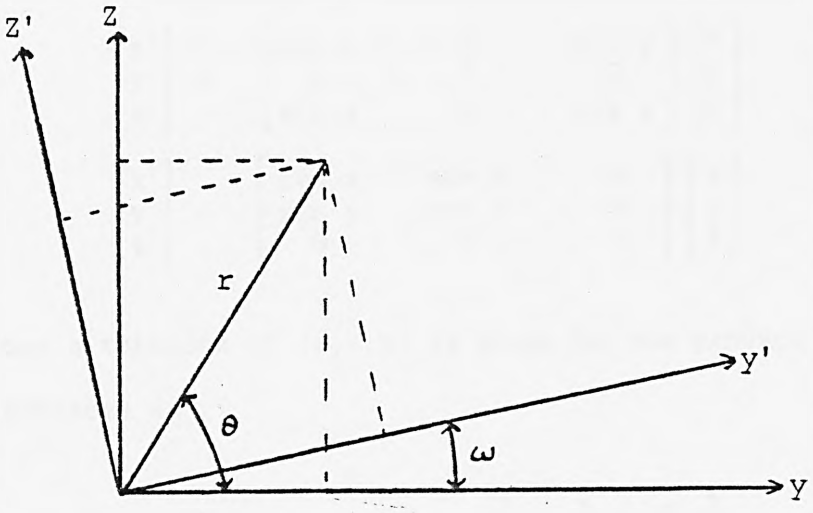
ed. by Chester C. Slama, American Society of Photogrammetry.

pp. 37-101.

Appendix A

Derivation of Rotation Matrix

The figure below shows a side view of the y - z axes in a right handed cartesian coordinate system together with the same axes rotated by ω about the x-axis, resulting in the y'- z' axes.



Rotation of ω about the x-axis

From simple geometry we have

$$y = r.\cos\theta \quad \text{and} \quad z = r.\sin\theta$$

In the new axes

$$y' = r.\cos(\theta - \omega)$$
$$= r.\cos\theta.\cos\omega + r.\sin\theta.\sin\omega$$
$$= y.\cos\omega + z.\sin\omega$$

and
$$z' = r.\sin(\theta - \omega)$$
$$= r.\sin\theta.\cos\omega - r.\cos\theta.\sin\omega$$

$$= z.\cos\omega - y.\sin\omega$$

The transformation is then expressed in matrix form as

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix} . \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Similarily a rotation of φ about the y-axis and a rotation of κ about the z-axis can be represented by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} . \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Therefore a rotation of (ω,φ,κ) is given by the product of these three matrices as

$$\text{rotation matrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

where

$$r_{11} = \cos\kappa.\cos\varphi$$

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

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

$$r_{21} = -\sin\kappa.\cos\varphi$$

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

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

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

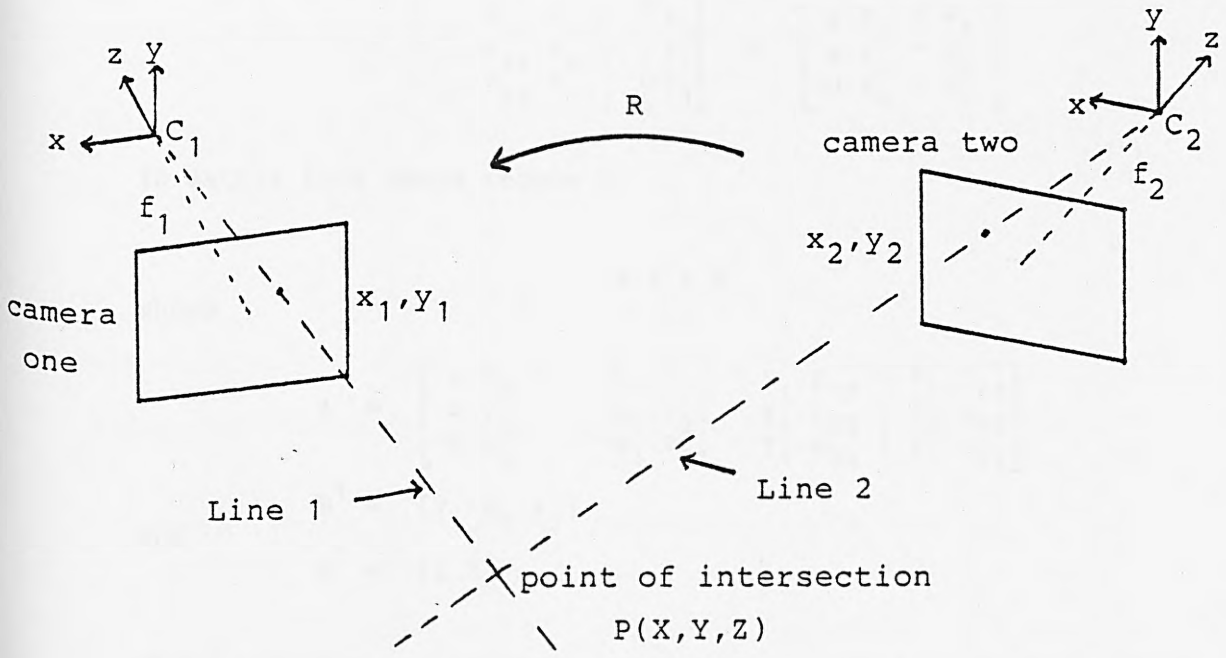
$$r_{32} = -\cos\varphi.\sin\omega$$

$$r_{33} = \cos\varphi.\cos\omega$$

Appendix B

Extraction of World Coordinates from Corresponding Stereo Points

Consider the figure illustrated below which depicts two cameras with the projection of a scene point, P(X,Y,Z) onto their respective image planes. The coordinates of the image coordinates being (x_1,y_1) for camera one and (x_2,y_2) for camera two.



Intersection of Conjugate Lines

The world coordinates of the image points may be calculated from the intersection of the two lines - one containing camera one image coordinates and perspective centre, the other containing camera two image coordinates and perspective centre. The equations of the two lines are

$$\text{line 1:} \quad \lambda.(x_1 i + y_1 j - f_1 k)$$

$$\text{line 2:} \quad \mu.(x_2 I + y_2 J - f_2 K) - (v_1 I + v_2 J + v_3 K)$$

where (i,j,k) is related to (I,J,K) by the relative rotation between the two cameras and is represented by

$$\begin{bmatrix} i \\ j \\ k \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} I \\ J \\ K \end{bmatrix}$$

The intersection point of the two lines is obtained by solving the three simultaneous equations

$$\lambda . \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix} = \begin{bmatrix} \mu . x_2 - v_1 \\ \mu . y_2 - v_2 \\ -\mu . f_2 - v_3 \end{bmatrix}$$

In matrix form these reduce to

$$A.X = B$$

where

$$A = \begin{bmatrix} \mu . x_2 & -x_1 . r_{11} - y_1 . r_{12} + f_1 . r_{13} \\ \mu . y_2 & -x_1 . r_{21} - y_1 . r_{22} + f_1 . r_{23} \\ -\mu . f_2 & -x_1 . r_{31} - y_1 . r_{32} + f_1 . r_{33} \end{bmatrix}$$

$$B^T = [v_1 \; v_2 \; v_3]$$

and

$$X^T = [\mu \; \lambda]$$

This equation can easily be solved for μ and λ using the pseudo inverse matrix which provides a least squares solution as

$$X = (A^T . A)^{-1} . A^T . B$$

The substitution of λ or μ into the appropriate line equation yields the world point P.

Appendix C

Solution of a Set of Non-Linear Equations

Consider the set of non-linear equations below, these equations may solved using the Newton-Raphson method (Carnahan et al., 1973).

$$\Sigma_i \, F_i(x_1, x_2, x_3 \ldots\ldots x_m) = \Sigma_i \, F_i(x) = 0$$

where i = 1 to n and n = number of equations.

The equation $F_i(x)$ may be approximated by a Taylor series upto the first differential as

$$F_i(x + \delta x) \approx F_{oi}(x) + \Sigma_j \, (\partial F_i / \partial x_j).\delta x_j$$

where j = 1 to m and m = number of unknown parameters, $F_{oi}(x)$ represents the evaluation of function $F_i(x)$ at the point x.

The set of equations may now be approximated to

$$\Sigma_i \, [F_{oi}(x) + \Sigma_j \, (\partial F_i / \partial x_j).\delta x_j] = 0$$

In matrix form these equations may be expressed as

$$J.\delta x = -F_o$$

where the column vector δx represents the improvement values for x, F_o represents the function evaluations at x and J is the Jacobean matrix composed of the function derivatives such that the (i,j) term is $\partial F_i / \partial x_j$.

The solution of the above equation can be carried out by the use of the Gauss-Jordan algorithm (Carnaham et al. 1973) to compute

the inverse of J. Therefore the improvement values, δx for the unknowns x can be used to provide a better approximation for the unknowns as

$$x_{new} = x + \delta x$$

Starting from an initial estimate of x, the above procedure may be repeated until the largest change in x is below a specified tolerance when the solution is assumed to have converged.

Appendix D

Partial Differentials for Solution of Absolute Camera Orientation

The transformation equations to be solved, for the determination of the absolute orientation of a camera, are

$$F = x' - x_m' = 0$$

$$G = y' - y_m' = 0$$

where

$$x' = (x + \Delta x) + b$$

$$y' = c.(x + \Delta x) + d.(y + \Delta y) + e$$

$$\Delta x = p_1.(r^2 + 2.x^2) + x.(1 + acc\_a + 2.p_2.y)$$

$$\Delta y = p_2.(r^2 + 2.y^2) + y.(1 + acc\_a + 2.p_1.x)$$

$$x = -f.(R_1 / R_3)$$

$$y = -f.(R_2 / R_3)$$

$$R_1 = r_{11}.(X - X_c) + r_{12}.(Y - Y_c) + r_{13}.(Z - Z_c)$$

$$R_2 = r_{21}.(X - X_c) + r_{22}.(Y - Y_c) + r_{23}.(Z - Z_c)$$

$$R_3 = r_{31}.(X - X_c) + r_{32}.(Y - Y_c) + r_{33}.(Z - Z_c)$$

$$r_{11} = \cos \varphi.\cos \kappa$$

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

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

$$r_{21} = -\cos \varphi.\sin \kappa$$

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

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

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

$$r_{32} = -\sin \omega.\cos \varphi$$

$$r_{33} = \cos \omega.\cos \varphi$$

The partial differential equations required for the Jacobean matrix are

$$acc\_c = 2.p_1.y + 2.p_2.x + (acc\_b.x.y / r)$$

$$acc\_d = 6.p_1.x + acc\_a + 2.p_2.y + (acc\_b.x^2 / r)$$

$$\partial F/\partial X_c = acc\_c.f.(R_3.r_{21} - R_2.r_{31}) / R_3^2 + acc\_d.f.(R_3.r_{11} - R_1.r_{31}) / R_3^2$$

$$\partial F/\partial Y_c = acc\_c.f.(R_3.r_{22} - R_2.r_{32}) / R_3^2 + acc\_d.f.(R_3.r_{12} - R_1.r_{32}) / R_3^2$$

$$\partial F/\partial Z_c = acc\_c.f.(R_3.r_{23} - R_2.r_{33}) / R_3^2 + acc\_d.f.(R_3.r_{13} - R_1.r_{33}) / R_3^2$$

$$\partial F/\partial \omega = acc\_c.f.(R_2.\partial R_3/\partial \omega - R_3.\partial R_2/\partial \omega) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \omega - R_3.\partial R_1/\partial \omega) / R_3^2$$

$$\partial F/\partial \varphi = acc\_c.f.(R_2.\partial R_3/\partial \varphi - R_3.\partial R_2/\partial \varphi) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \varphi - R_3.\partial R_1/\partial \varphi) / R_3^2$$

$$\partial F/\partial \lambda = acc\_c.f.(R_2.\partial R_3/\partial \lambda - R_3.\partial R_2/\partial \lambda) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \lambda - R_3.\partial R_1/\partial \lambda) / R_3^2$$

$$acc\_c = 6.p_1.x + acc\_a + 2.p_2.y + (acc\_b.x^2 / r)$$

$$acc\_d = 2.p_1.y + 2.p_2.x + (acc\_b.x.y / r)$$

$$\partial F/\partial f = -(R_1 / R_3).acc\_c - (R_2 / R_3).acc\_d$$

$$\partial F/\partial b = 1$$

$$\partial F/\partial c = 0$$

$$\partial F/\partial d = 0$$

$$\partial F/\partial e = 0$$

$$\partial F/\partial a_1 = x.r^2$$

$$\partial F/\partial a_2 = r^2.\partial F/\partial a_1$$

$$\partial F/\partial a_3 = r^2.\partial F/\partial a_2$$

$$\partial F/\partial p_1 = r^2 + 2.x^2$$

$$\partial F/\partial p_2 = 2.x.y$$

$$acc\_c = d.(6.p_2.y + acc\_a + 2.p_1.x + (acc\_b.y^2 / r))$$

$$acc\_d = d.(2.p_2.x + 2.p_1.y + (acc\_b.x.y / r))$$

$$\partial G/\partial X_c = acc\_c.f.(R_3.r_{21} - R_2.r_{31}) / R_3^2 + acc\_d.f.(R_3.r_{11} - R_1.r_{31}) / R_3^2 + c.\partial G/\partial X_c$$

$$\partial G/\partial Y_c = acc\_c.f.(R_3.r_{22} - R_2.r_{32}) / R_3^2 + acc\_d.f.(R_3.r_{12} - R_1.r_{32}) / R_3^2 + c.\partial G/\partial Y_c$$

$$\partial G/\partial Z_c = acc\_c.f.(R_3.r_{23} - R_2.r_{33}) / R_3^2 + acc\_d.f.(R_3.r_{13} - R_1.r_{33}) / R_3^2 + c.\partial G/\partial Z_c$$

$$\partial G/\partial \omega = acc\_c.f.(R_2.\partial R_3/\partial \omega - R_3.\partial R_2/\partial \omega) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \omega - R_3.\partial R_1/\partial \omega) / R_3^2 + c.\partial G/\partial \omega$$

$$\partial G/\partial \varphi = acc\_c.f.(R_2.\partial R_3/\partial \varphi - R_3.\partial R_2/\partial \varphi) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \varphi - R_3.\partial R_1/\partial \varphi) / R_3^2 + c.\partial G/\partial \varphi$$

$$\partial G/\partial \lambda = acc\_c.f.(R_2.\partial R_3/\partial \lambda - R_3.\partial R_2/\partial \lambda) / R_3^2 + acc\_d.f.(R_1.\partial R_3/\partial \lambda - R_3.\partial R_1/\partial \lambda) / R_3^2 + c.\partial G/\partial \lambda$$

$$acc\_c = 6.p_2.y + acc\_a + 2.p_1.x + (acc\_b.y^2 / r)$$

$$acc\_d = 2.p_2.x + 2.p_1.y + (acc\_b.x.y / r)$$

$$\partial G/\partial f = -d.((R_2 / R_3).acc\_c + (R_1 / R_3).acc\_d) + c.\partial F/\partial f$$

$$\partial G/\partial b = 0$$

$$\partial G/\partial c = x + \Delta x$$

$$\partial G/\partial d = y + \Delta y$$

$$\partial G/\partial e = 1$$

$$\partial G/\partial a_1 = r^2.(x.c + y.d)$$

$$\partial G/\partial a_2 = r^2.\partial G/\partial a_1$$

$$\partial G/\partial a_3 = r^2.\partial G/\partial a_2$$

$$\partial G/\partial p_1 = 2.x.y.d + (r^2 + 2.x^2).c$$

$$\partial G/\partial p_2 = (r^2 + 2.y^2).d + 2.x.y.c$$

where

$$\partial R_1/\partial \omega = (\cos \omega.\sin \varphi.\cos \kappa - \sin \omega.\sin \kappa).(Y - Y_c) + (\cos \omega.\sin \kappa + \sin \omega.\sin \varphi.\cos \kappa).(Z - Z_c)$$

$$\partial R_2/\partial \omega = (\sin \omega.\cos \kappa + \cos \omega.\sin \varphi.\sin \kappa).(Y_c - Y) + (\cos \omega.\cos \kappa - \sin \omega.\sin \varphi.\sin \kappa).(Z - Z_c)$$

$$\partial R_3/\partial \omega = \cos \omega.\cos \varphi.(Y_c - Y) - \sin \omega.\cos \varphi.(Z - Z_c)$$

$$\partial R_1/\partial \varphi = \sin \varphi.\cos \kappa.(X_c - X) + \sin \omega.\cos \varphi.\cos \kappa.(Y - Y_c) - \cos \omega.\cos \varphi.\cos \kappa.(Z - Z_c)$$

$$\partial R_2/\partial \varphi = \sin \varphi.\sin \kappa.(X - X_c) - \sin \omega.\cos \varphi.\sin \kappa.(Y - Y_c) + \cos \omega.\cos \varphi.\sin \kappa.(Z - Z_c)$$

$$\partial R_3/\partial \varphi = \cos \varphi.(X - X_c) + \sin \omega.\sin \varphi.(Y - Y_c) - \cos \omega.\sin \varphi.(Z - Z_c)$$

$$\partial R_1/\partial \lambda = \cos \varphi.\sin \kappa.(X_c - X) + (\cos \omega.\cos \kappa - \sin \omega.\sin \varphi.\sin \kappa).(Y - Y_c) + (\sin \omega.\cos \kappa + \cos \omega.\sin \varphi.\sin \kappa).(Z - Z_c)$$

$$\partial R_2/\partial \lambda = \cos \varphi.\cos \kappa.(X_c - X) - (\cos \omega.\sin \kappa + \sin \omega.\sin \varphi.\cos \kappa).(Y - Y_c) + (\cos \omega.\sin \varphi.\cos \kappa - \sin \omega.\sin \kappa).(Z - Z_c)$$

$$\partial R_3/\partial \lambda = 0$$

$$acc\_a = 1 + (a_1 + (a_2 + a_3.r^2).r^2).r^2$$

$$acc\_b = 2.r.(a_1 + (2.a_2 + 3.a_3.r^2).r^2)$$

$$r = \sqrt{(x^2 + y^2)}$$

For the second stage of the solution, the determination of the transformation from comparator-to-image coordinates, the equations to be solved are

$$H = x + \Delta x' - x_c = 0$$

and

$$I = y + \Delta y' - y_c = 0$$

where

$$\Delta x' = p_1'.(r^2 + 2.x^2) + x.(acc\_a + 2.p_2'.y)$$

$$\Delta y' = p_2'.(r^2 + 2.y^2) + y.(acc\_a + 2.p_1'.x)$$

$$x = x' + b'$$

$$y = c'.x' + d'.y' + e'$$

The partial differential equations required for the Jacobean matrix are

$$acc\_c = acc\_a + 4.x.p_1' + 2.y.p_2'$$

$$acc\_d = (x.acc\_b + 2.r.p_1') / r$$

$$\partial H/\partial b' = acc\_c + acc\_d.x$$

$$acc\_c = 2.x.p_2'$$

$$acc\_d = acc\_d.y$$

$$\partial H/\partial c' = x'.(acc\_c + acc\_d)$$

$$\partial H/\partial d' = y'.(acc\_c + acc\_d)$$

$$\partial H/\partial e' = acc\_c + acc\_d$$

$$\partial H/\partial a_1' = x.r^2$$

$$\partial H/\partial a_2' = r^2.\partial H/\partial a_1'$$

$$\partial H/\partial a_3' = r^2.\partial H/\partial a_2'$$

$\partial H/\partial p_1' = r^2 + 2.x^2$

$\partial H/\partial p_2' = 2.x.y$

$acc\_c = (acc\_b.y + 2.r.p_2') / r$

$acc\_d = 2.y.p_1'$

$\partial I/\partial b' = acc\_c.x + acc\_d$

$acc\_c = acc\_c.y$

$acc\_d = acc\_a + 4.y.p_2' + 2.x.p_1'$

$\partial I/\partial c' = x'.(acc\_c + acc\_d)$

$\partial I/\partial d' = y'.(acc\_c + acc\_d)$

$\partial I/\partial e' = acc\_c + acc\_d$

$\partial I/\partial a_1' = y.r^2$

$\partial I/\partial a_2' = r^2.\partial I/\partial a_1'$

$\partial I/\partial a_3' = r^2.\partial I/\partial a_2'$

$\partial I/\partial p_1' = 2.x.y$

$\partial I/\partial p_2' = r^2 + 2.y^2$

where

$acc\_a = 1 + (a_1' + (a_2' + a_3'.r^2).r^2).r^2$

$acc\_b = 2.r.(a_1 + (2.a_2 + 3.a_3.r^2).r^2)$

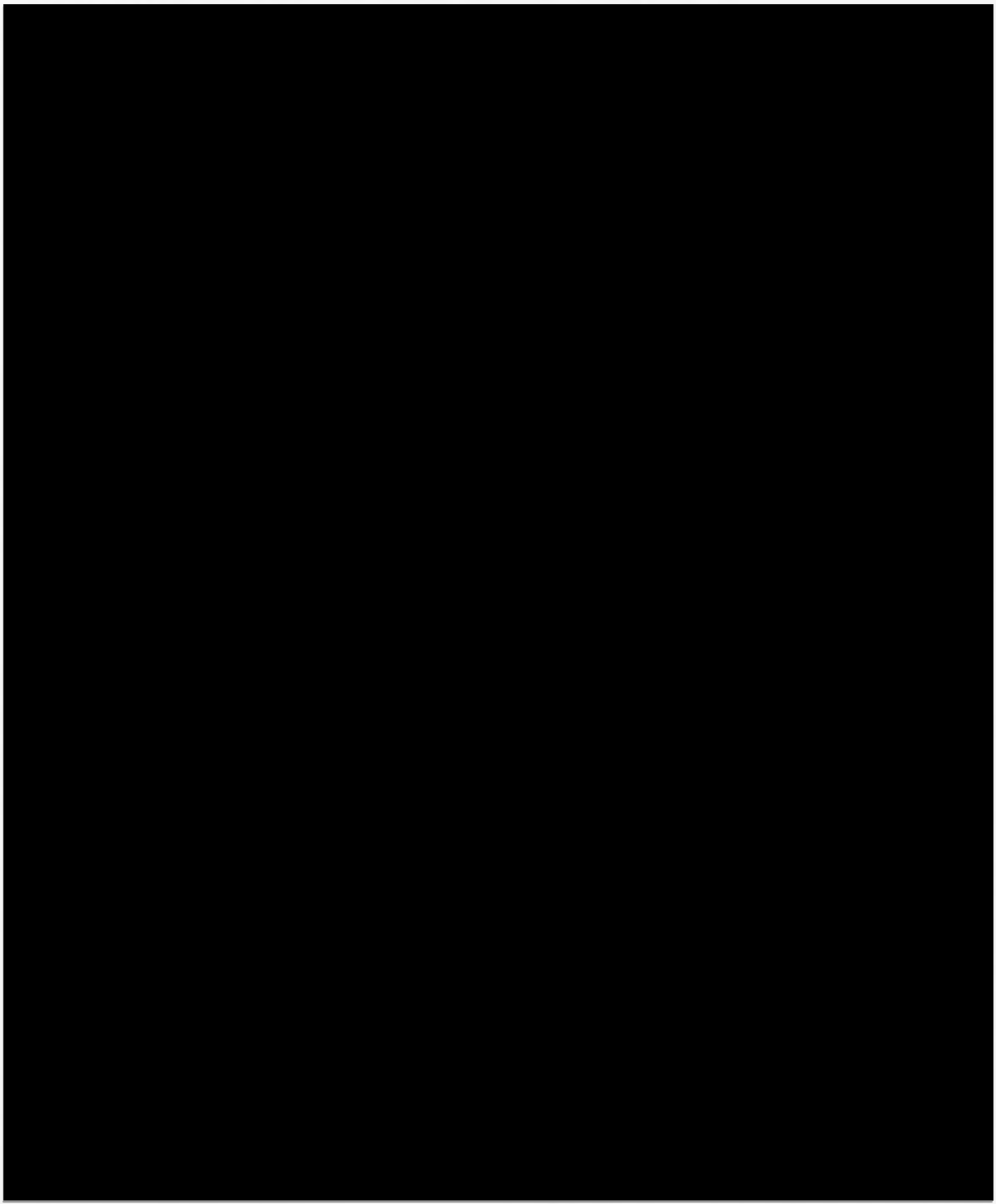
$r = \sqrt{(x^2 + y^2)}$

Appendix E - Camera Calibration Program Listing (pp.167-177) has been removed for copyright reasons

Appendix E

Camera Calibration Program Listing

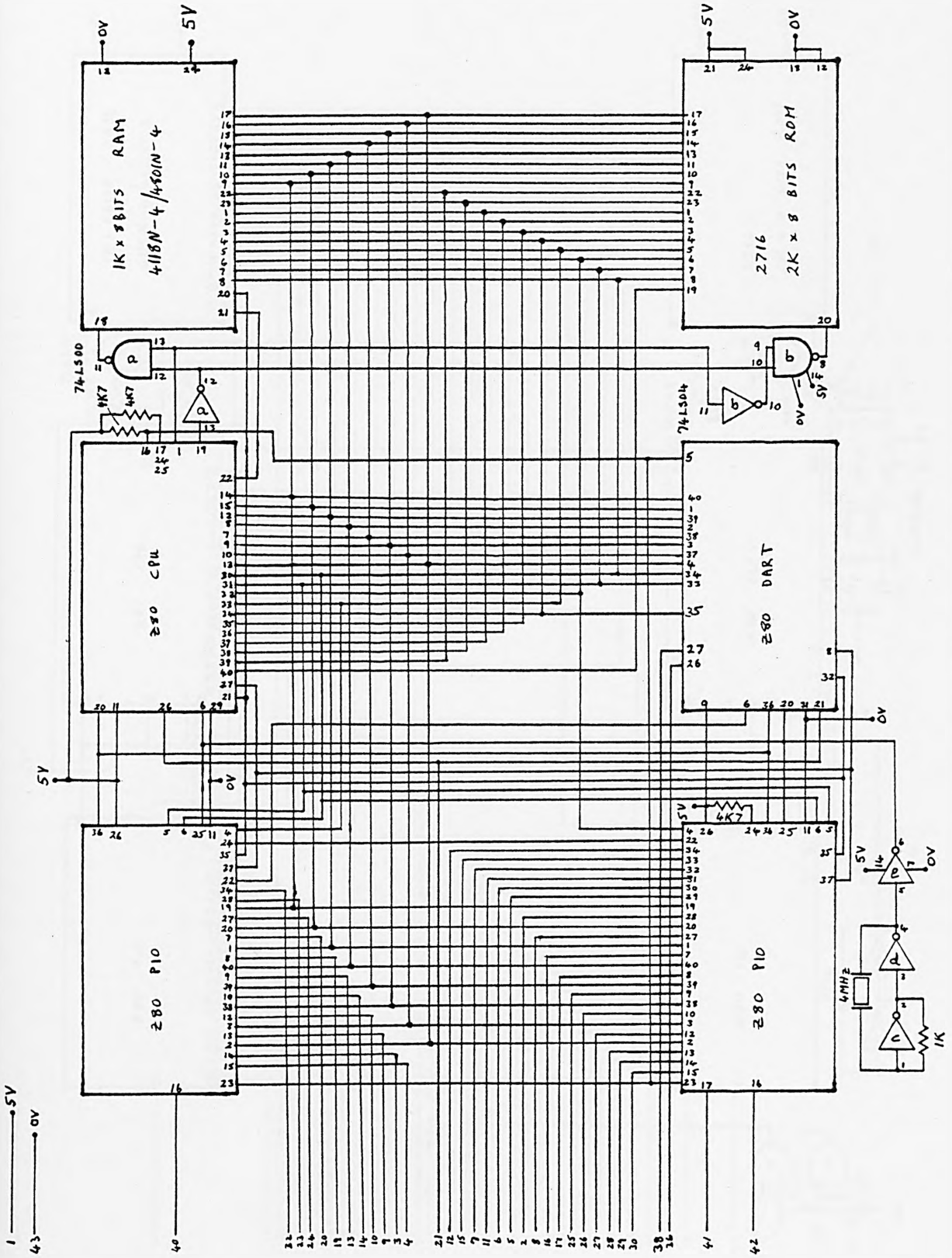
This is a listing of the camera calibration program which computes the internal and exterior geometric parameters of a camera.



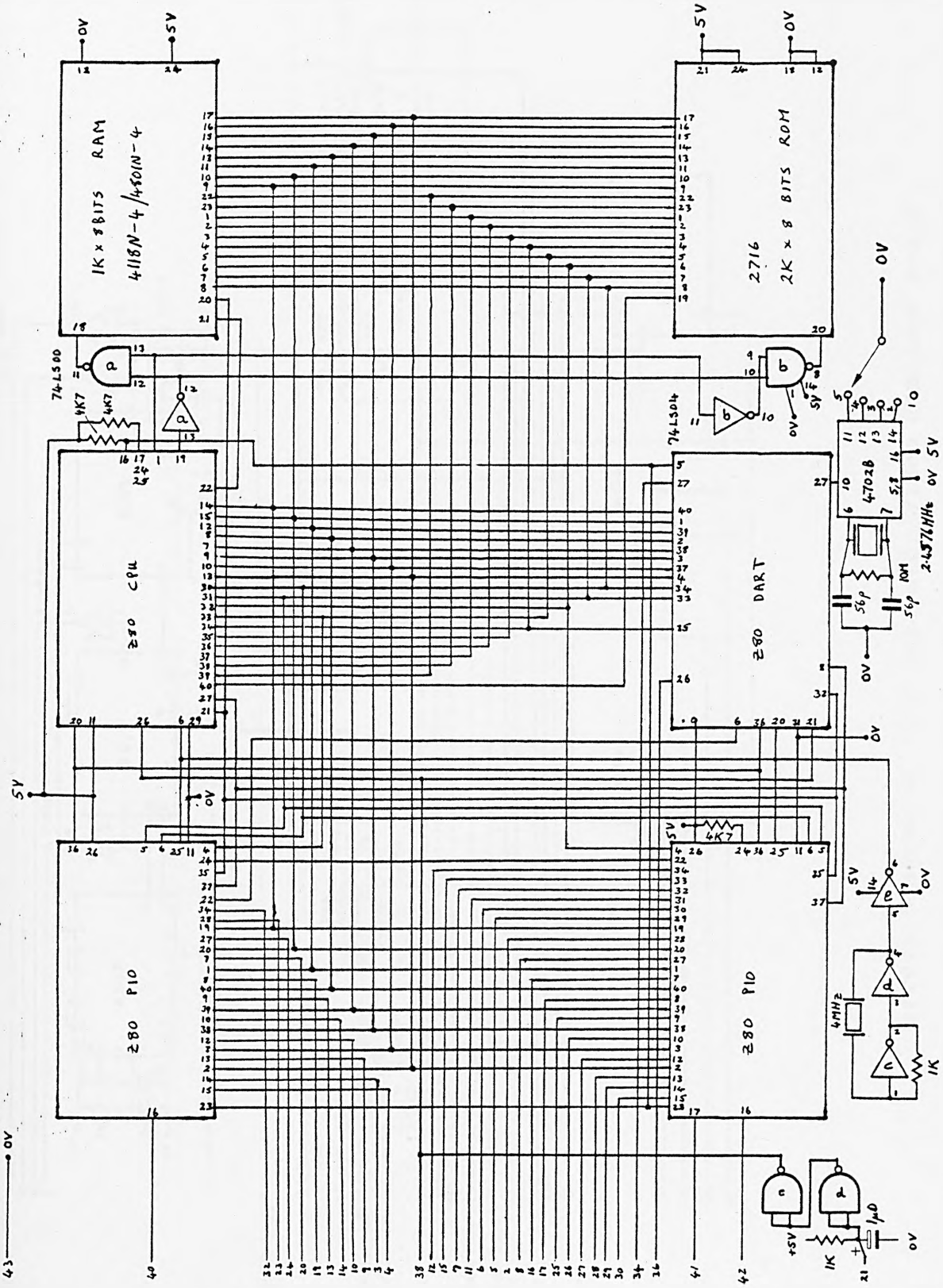
Appendix F

Circuit Diagrams of Marker Generator/Tracker

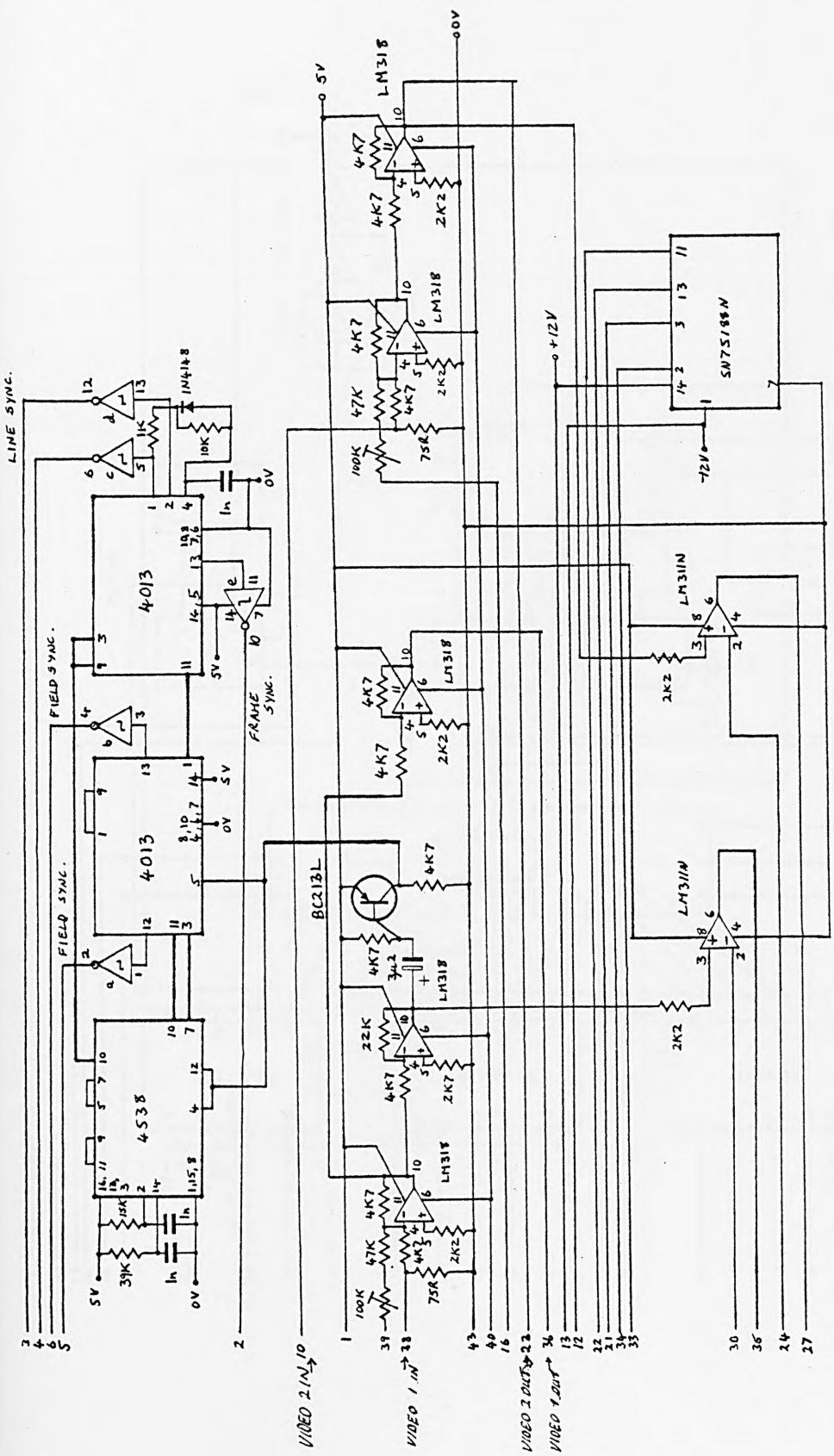
This appendix provides a full details of the marker generator/tracker circuit diagrams. The unit comprises of five basic circuit boards with a power supply board and a wiring diagram for interconnection of the individual boards is also provided.



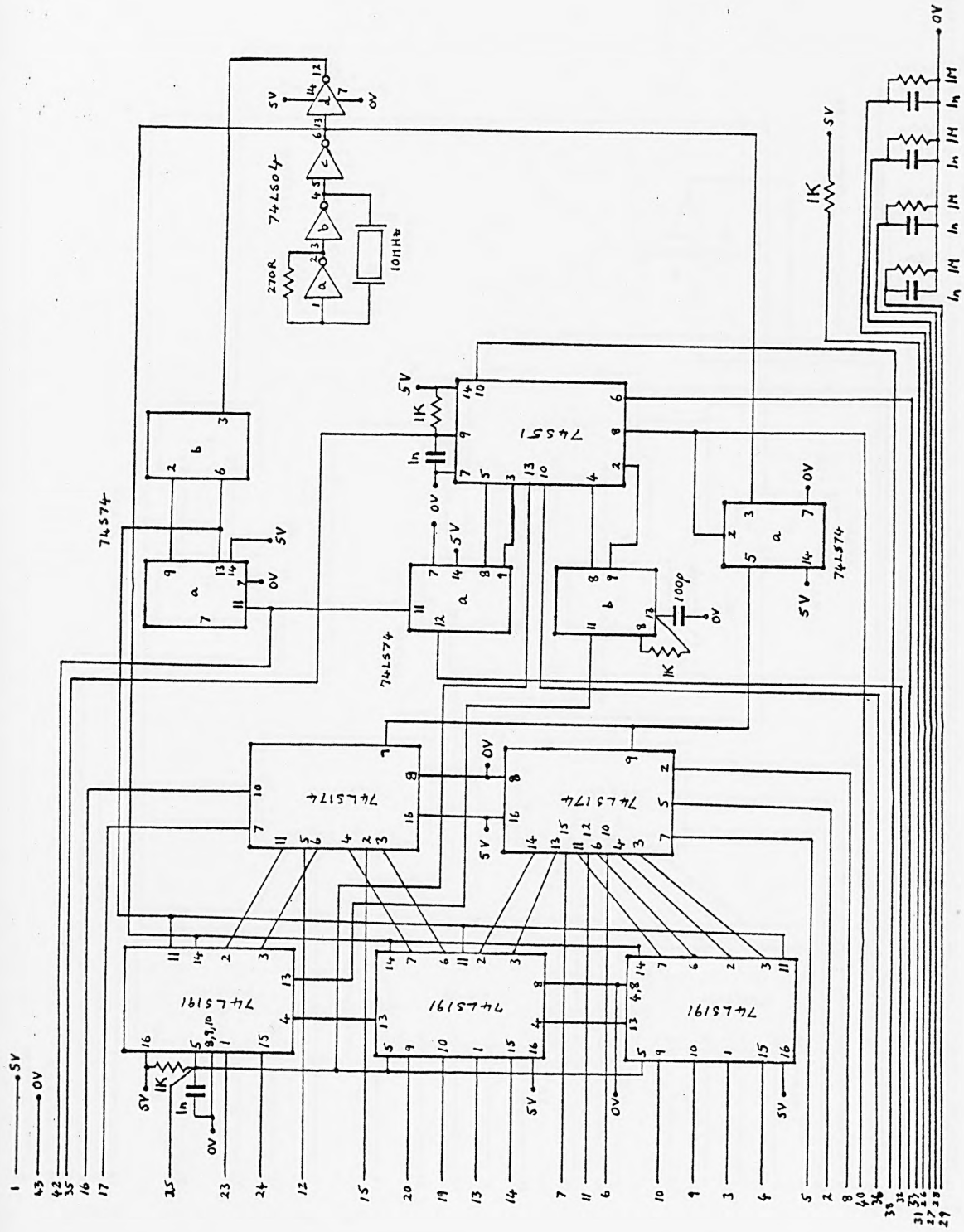
Circuit Diagram of Second Micro-processor Board



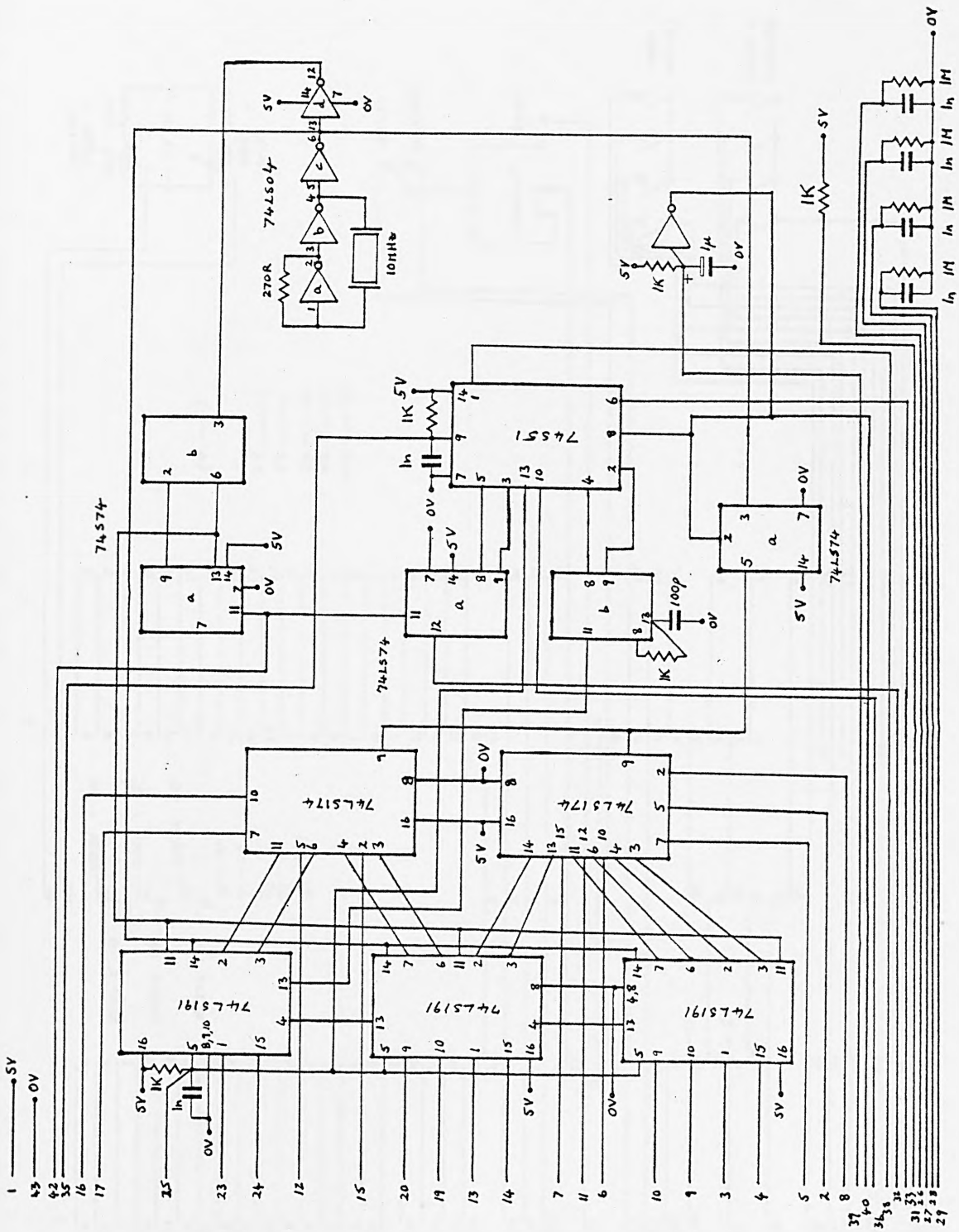
Circuit Diagram of First Micro-processor Board



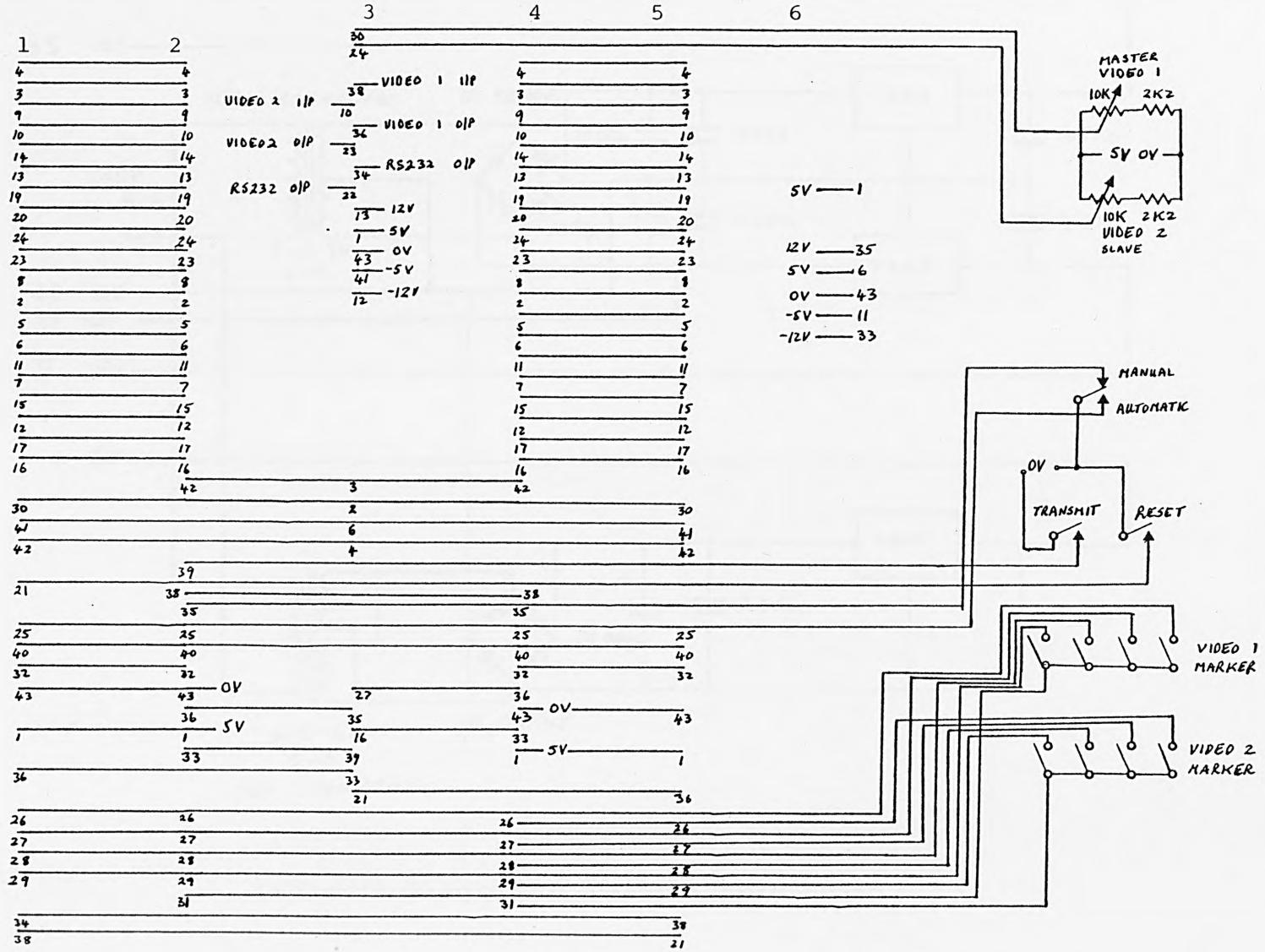
Circuit Diagram of Sync Separator/Video Combiner Board



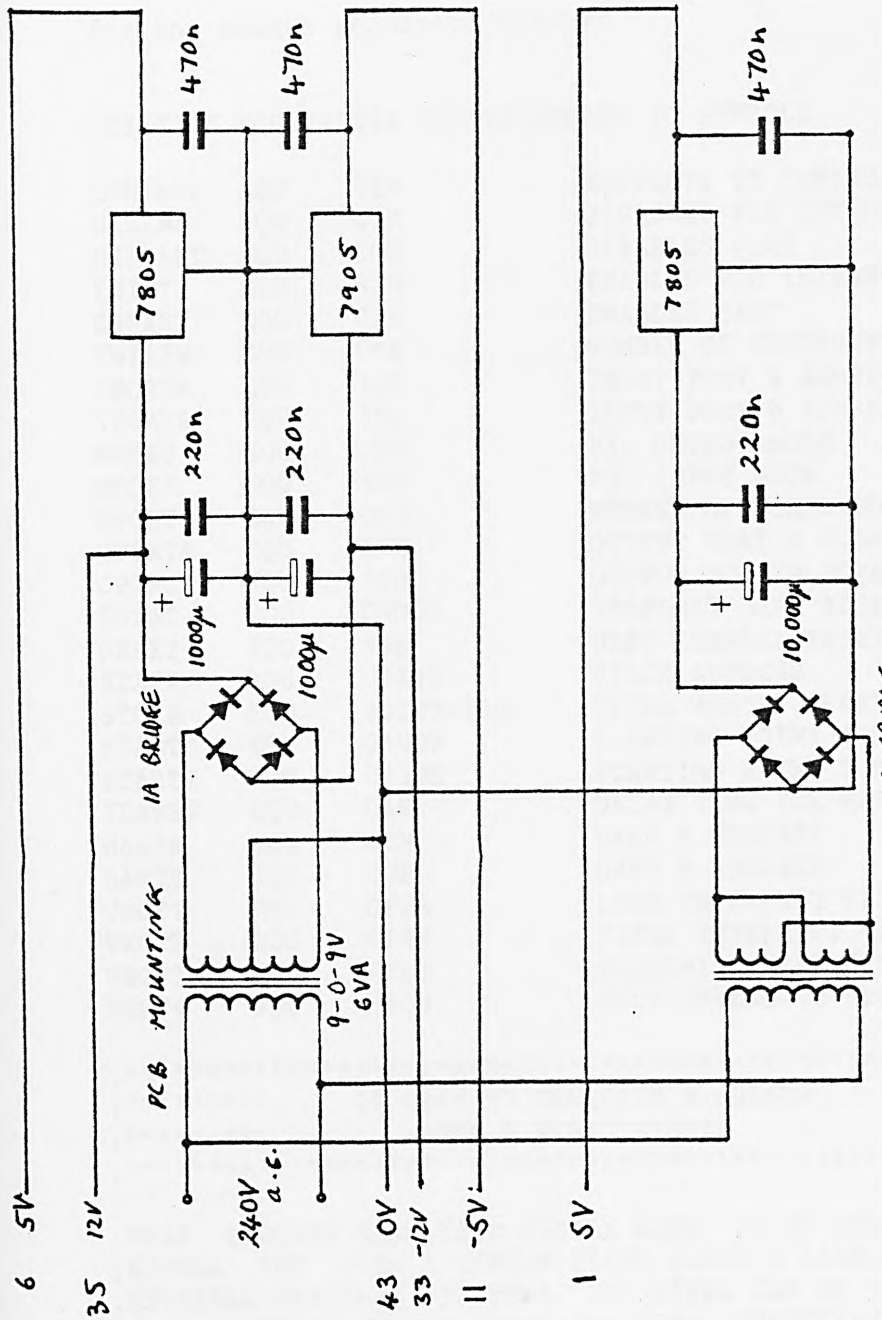
Circuit Diagram of Micro-processor Interface Board



Circuit Diagram of Micro-processor Interface Board



Wiring Diagram of Marker Generator/Tracker Circuit Boards



Circuit Diagram of Power Supply Board

Appendix G - Z80 Assembler Listing of Marker Generator/Tracker (pp.186-191) has been removed for copyright reasons

Appendix H - Stereo Ranging Algorithm Program Listing (pp.192-208) has been removed for copyright reasons

Appendix G

Z80 Assembler Listing of Marker Generator/Tracker

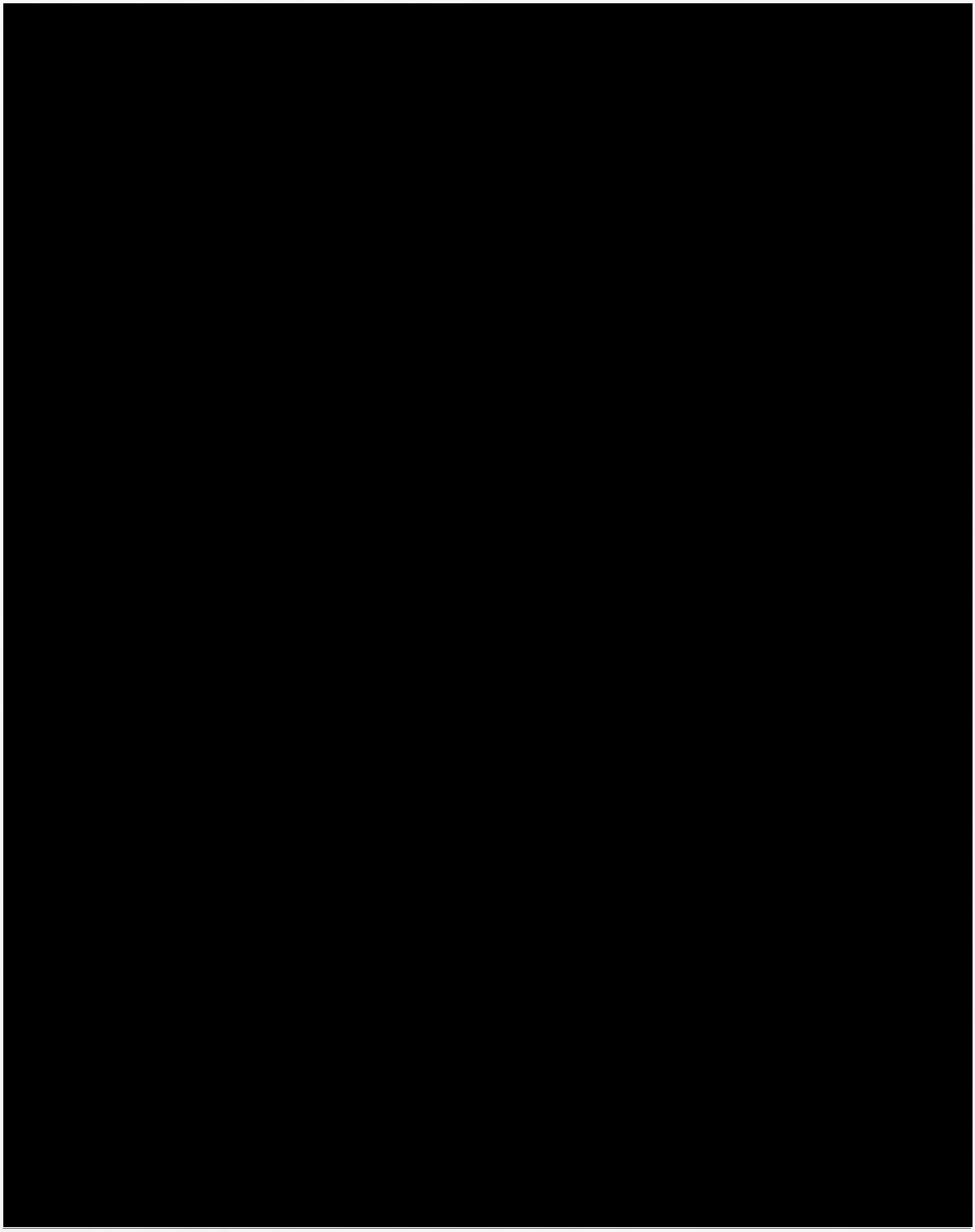
This is a full listing, in Z80 assembler, of the program written for the marker generator/tracker.

```
;*********************************************************
;*********     PROGRAM TO GENERATE A CURSOR     *********
;*********          ONTO A VIDEO SIGNAL         *********
;*********************************************************

;THIS  PROGRAM ENABLES A SINGLE LINE  TO BE SELECTED FROM A VIDEO
;SIGNAL  AND  ALSO A SINGLE PIXEL ALONG A LINE WITH  THE  AID  OF
;EXTERNAL HARDWARE COUNTERS. THE PIXEL CAN BE SELECTED BY THE USE
;OF A JOYSTICK SWITCH WHICH PROVIDES APPROPRIATE DATA SIGNALS FOR
;THE PROGRAM.

;THE  PROGRAM OPERATES COMPLETELY ON  INTERRUPTS AND SO THE VIDEO
;LINE  AND  FIELD  SIGNALS ARE  USED  FOR  INTERRUPT  GENERATION.
;ANOTHER  INTERRUPT  IS  GENERATED WHEN  TRANSMITTING  OF  CURSOR
;COORDINATES IS REQUIRED.
```

Appendix H

Stereo Ranging Algorithm Program Listing

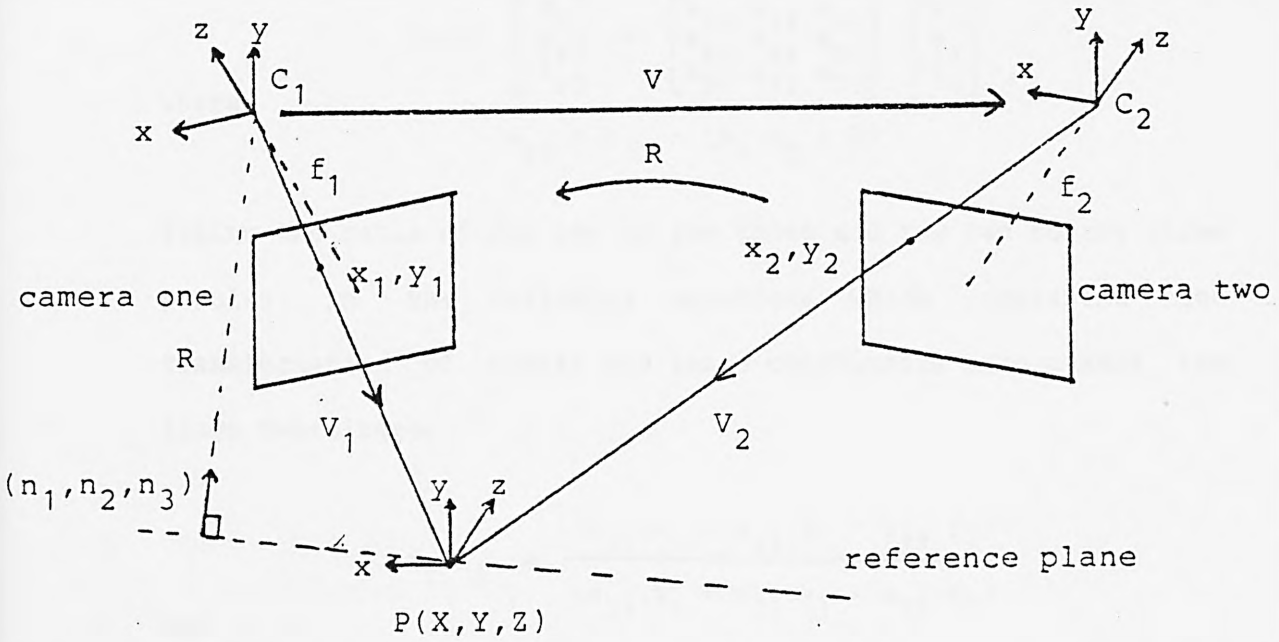
This is a full listing of the stereo ranging algorithm with associated subroutines.



Appendix I

Mapping of Coordinates from One Image onto Another

The figure below shows a pair of camera stations separated by direction vector V. The vector from the scene point, P(X,Y,Z) to camera one perspective centre is V_1 and the vector from the scene point to camera two perspective centre is V_2. We can therefore write



Translation of Image Coordinates

$$V_2 = V_1 - V$$

In matrix form the above equation becomes

$$\mu \cdot \begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \lambda \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix} - \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

We know that the normal vector of the reference plane passing through the scene point, P is (n_1, n_2, n_3) so

$$V_1 . N = R$$

that is

$$\lambda . (n_1 . x_1 + n_2 . y_1 - n_3 . f_1) = R$$

hence

$$1 / \lambda = (n_1 . x_1 + n_2 . y_1 - n_3 . f_1) / R$$

The above equation reduces to

$$(\mu/\lambda) . \begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \left[\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} - (1/R) . \begin{bmatrix} v_1 . n_1 & v_1 . n_2 & v_1 . n_3 \\ v_2 . n_1 & v_2 . n_2 & v_3 . n_3 \\ v_3 . n_1 & v_3 . n_2 & v_3 . n_3 \end{bmatrix} \right] . \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

where

$$(\mu/\lambda) . \begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} . \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

$$a_{ij} = r_{ij} - (v_i . n_j / R)$$

Taking the ratio of row one to row three and row two to row three results in the following equations which represent the transformation of camera one image coordinates onto camera two image coordinates.

$$-x_2/f_2 = \frac{(a_{11} . x_1 + a_{12} . y_1 - a_{13} . f_1)}{(a_{31} . x_1 + a_{32} . y_1 - a_{33} . f_1)}$$

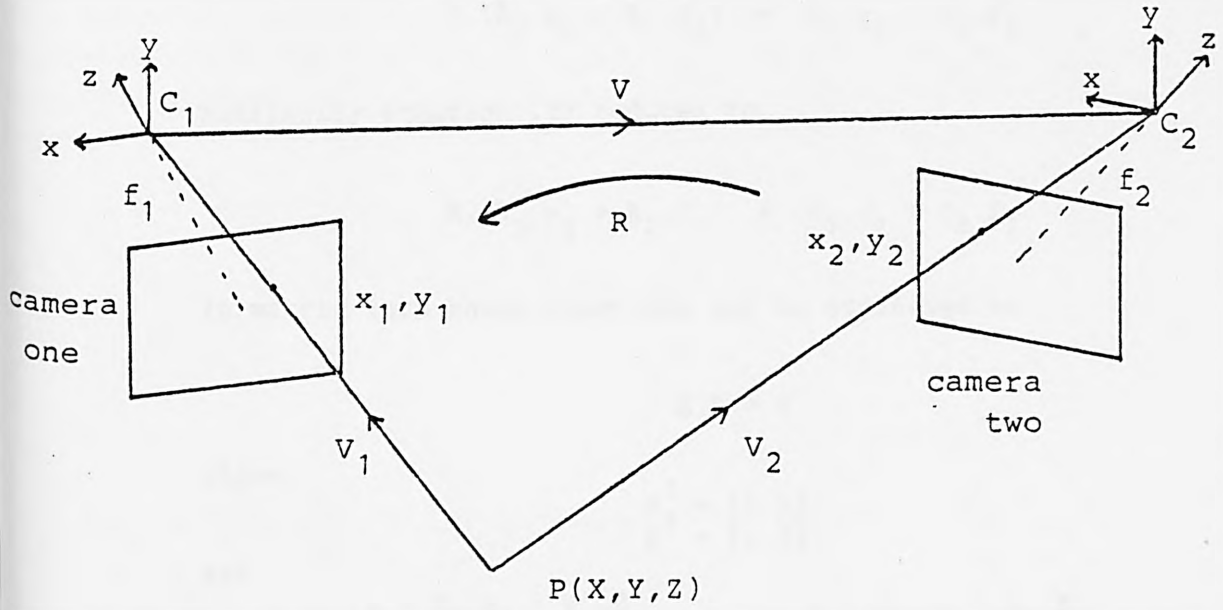
and

$$-y_2/f_2 = \frac{(a_{21} . x_1 + a_{22} . y_1 - a_{23} . f_1)}{(a_{31} . x_1 + a_{32} . y_1 - a_{33} . f_1)}$$

Appendix J

Computation of Range Value from Corresponding Stereo Points

The figure shown below illustrates the projection of a world point, P(X,Y,Z) onto two stereo images. The image coordinates are (x_1,y_1) on camera one and (x_2,y_2) on camera two. In addition camera one has associated with it a range value, R.



Extraction of Range Value from Conjugate Points

The transformation of camera one image point (x_1,y_1) onto camera two image point (x_2,y_2) is represented by the equation derived in Appendix I as

$$\begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

where

$$a_{ij} = r_{ij} - (v_i . n_j \; / \; R)$$

Obtaining two equations from the ratio of rows one and two to row

three gives

$$-x_2/f_2 = (R.R_1 - c_1) / (R.R_3 - c_3) \qquad (1)$$

$$-y_2/f_2 = (R.R_2 - c_2) / (R.R_3 - c_3) \qquad (2)$$

where

$$R_i = r_{i1}.x_1 + r_{i2}.y_1 - r_{i3}.f_1$$

$$c_i = v_i.(n_1.x_1 + n_2.y_1 - n_3.f_1)$$

Equation (1) can be reduced to

$$R.(R_3.x_2 + R_1.f_2) = c_3.x_2 + c_1.f_2$$

similarily equation (2) reduces to

$$R.(R_3.y_2 + R_2.f_2) = c_3.y_2 + c_2.f_2$$

In matrix form these equations may be expressed as

$$A.X = B$$

where

$$A^T = [a\ b]$$
$$B^T = [c\ d]$$

and

$$a = R_3.x_2 + R_1.f_2 \qquad b = R_3.y_2 + R_2.f_2$$

$$c = c_3.x_2 + c_1.f_2 \qquad d = c_3.y_2 + c_2.f_2$$

Use of the pseudo inverse matrix enables computation of the range value using a least squares solution.

$$R = (A^T.A)^{-1}.A^T.B$$

$$= (a.c + b.d) / (a^2 + b^2)$$

Appendix K

Derivation of Partial Derivatives $\partial x_2/\partial R$ and $\partial y_2/\partial R$

The function for mapping a point in image one, (x_1, y_1) onto a point in image two, (x_2, y_2) is represented by

$$\begin{bmatrix} x_2 \\ y_2 \\ -f_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ -f_1 \end{bmatrix}$$

where

$$a_{ij} = r_{ij} - (v_i . n_j / R)$$

Dividing the first and second rows by the third row gives

$$-x_2/f_2 = (R.R_1 - c_1) / (R.R_3 - c_3) \qquad (1)$$

$$-y_2/f_2 = (R.R_2 - c_2) / (R.R_3 - c_3) \qquad (2)$$

where

$$R_i = r_{i1}.x_1 + r_{i2}.y_1 - r_{i3}.f_1$$

$$c_i = v_i.(n_1.x_1 + n_2.y_1 - n_3.f_1)$$

differentiate (1) with respect to R gives

$$\partial x_2/\partial R = -f_2.(c_1.R_3 - c_3.R_1) / (R.R_3 - c_3)^2$$

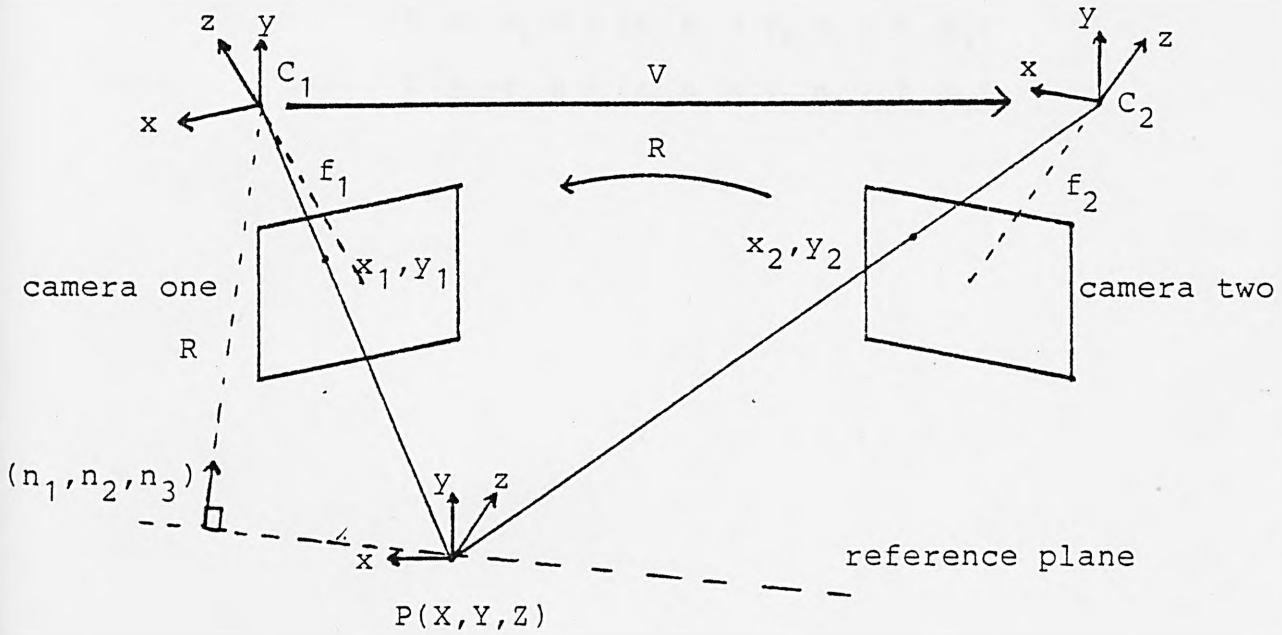
and differentiate (2) with respect to R gives

$$\partial y_2/\partial R = -f_2.(c_2.R_3 - c_3.R_2) / (R.R_3 - c_3)^2$$

Appendix L

Obtaining Absolute Coordinates of a Point

The figure below shows the projection of a world point, P(X,Y,Z) onto an image with image coordinates (x_1,y_1) and range value, R. The determination of the absolute coordinates from the image coordinates and associated range value is as follows.



Obtaining World Coordinates from Conjugate Points

The equation of the reference plane passing through the world point is

$$X.n_1 + Y.n_2 + Z.n_3 = R$$

and the equation of the line passing through the image point and the perspective centre is

$$\lambda.(x_1 i + y_1 j - f_1 k)$$

Substitution of the coordinates of a point on the line, into the plane equation yields

$$(\lambda.x_1).n_1 + (\lambda.y_1).n_2 - (\lambda.f_1).n_3 = R$$

therefore

$$\lambda = R / (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

Hence the world coordinates are given by

$$X = x_1.R / (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

$$Y = y_1.R / (x_1.n_1 + y_1.n_2 - f_1.n_3)$$

$$Z = -f_1.R / (x_1.n_1 + y_1.n_2 - f_1.n_3)$$