COMPUTATIONAL COLOR STEREOVISION

A Thesis Submitted to the

College of Graduate Studies and Research

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in the

Department of Electrical Engineering

University of Saskatchewan, Saskatoon, CANADA

By

ANKAMUTHU SATHISH

October 1998

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Electrical Engineering

University of Saskatchewan

Saskatoon, Saskatchewan

CANADA, S7N 5A9.

ABSTRACT

The present study aims at developing a practical method of video image analysis for evaluating depth of scour in three-dimensional flow fields. The method used a stereoscopy scheme which involves acquiring two images of the scene with two different viewpoints. The stereoscopy scheme involved the use of an epipolar constraint and a relaxation technique to obtain corresponding points in the two images. In the relaxation technique, the edges obtained by standard edge detectors were used as features to determine the match. A correlation technique was developed to eliminate false matches. The scheme was first applied to grayscale images. In the next stage, the three important components of color namely Hue, Saturation and Intensity were taken into consideration, and adopted in the stereoscopy scheme. The depth of scour was determined using the parallax between the matched points.

ACKNOWLEDGEMENTS

I thank GOD for providing me with all the resources to complete my thesis work.

I thank my supervisors, Dr. Hugh Wood and Dr. Ram Balachandar for their support and guidance through out my research work.

I thank Dr. Pramanick for his advice and support in various forms during these times.

I thank my friends Navjeet Singh Sandhu, Narendranath NV, Raju Balsubramanian, Saravananan Subramaniam, Mark Francis Tachie, Jayakumar Srinivasan, Venkatesh Mahalingam, Vijay Anand and Thiruchelvan Ratnapuri for their encouragement, kind words and assistance.

I thank my parents Ankamuthu and Shantha, my elder brother Babu Ankamuthu for their all forms of support at all times.

I thank all the members in the Hydrotechnical laboratory, University of Saskatchewan for their support during my work.

DEDICATION

Dedicated to my younger brother Ashok Ankamuthu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

.bmp	Bitmap file format
.jpg	Joint Photographic Experts Group compressed file format
.pgm	Portable gray map file format
.tif	Tagged image file format
2-D	Two dimensions
3-D	Three dimensions
A	Matrix with the corresponding points
A^T	Transpose of matrix A
B	Distance between the ground and the view point
C_ℓ	Optical center of the left camera
C_r	Optical center of the right camera
CMY	Cyan, Magenta, Yellow
DG	Disparity gradient
d_1	Relative distance of the objects in the right view of the image
d_2	Relative distance of the objects in the left view of the image
e	Unit vector
ep_ℓ	Epipole of the first camera with respect to second camera
ep_r	Epipole of the second camera with respect to first camera
E	Essential matrix
G	Gaussian function
h	Height of the object from the surface
HSI	Hue, Saturation, Intensity
I	Image
I_ℓ	Left Image
I_r	Right Image
K	Distance between the viewpoints
ln_ℓ	Epipolar line in the right image for a point n_l
ln_r	Epipolar line in the left image for a n_r

LGN	Lateral Genticulate Nucleus
LOG	Laplacian of Gaussian
n_ℓ	A point in the left image
n_r	A point in the right image
$\hat{n}_\ell = (x_\ell, y_\ell, 1)$	Image vector of a point in the left image
$\hat{n}_r = (x_r, y_r, 1)$	Image vector of a point in the right image
RGB	Red, Green, Blue
S	Strength of match
YIQ	Luminance (Y) and Chrominance (I-color, Q-some luminance)
r	Correlation coefficient
t	token
σ	Standard deviation
∇	Laplacian operator
∂	disparity gradient value

1. INTRODUCTION

1.1. General

On March 10, 1995, the bridges on Interstate 5 across Los Gatos Creek near Coalinga, in California, United States, failed during a flood, and resulted in the loss of several lives. Exposure or undermining of bridge pier and abutment foundations by the erosive action of the flowing water was assessed to have caused the failure. Several such bridge failures have occurred in North America within the last 10 years and most of the failures were caused by scour around bridge piers or abutments (Transportation Research Board, 1997) [1].

Bridge failures are caused by scouring which involves sediment-transport and erosion processes that cause streambed material to be removed from the bridge vicinity (see Figure 1.1, http://wwwpah2o.er.usgs.gov/projects/bridge_scour/). In general, scour is separated into three components, namely: pier scour, abutment scour and contraction scour. Pier scour occurs when flow impinges against the upstream side of the pier, forcing the flow in a downward direction and causing scour of the streambed adjacent to the pier. Abutment scour happens when flow impinges against the abutment, causing the flow to change direction and mix with adjacent main-channel flow, resulting in scouring forces near the abutment toe. Contraction scour occurs when flood-plain flow is forced back through a narrower opening at the bridge, where an increase in velocity can produce scour. Total scour for a particular site is the combined effects of each component. While different materials scour at different rates, the ultimate scour attained for different materials is similar and depends mainly on the duration of peak stream flow acting on the material.

The scouring effect is best visualized on a real image of the New Mexico bridge in Figure 1.2. The combination of all the scouring effect makes the bridge weak and it will ultimately collapse.

As part of a program to eliminate bridge failures in the United States, the Federal Highway Administration (FHWA) has established a nationwide bridge inspection program. Every bridge over water is inspected and given a classification. Those bridges classified as "scour critical" are given immediate attention; they are either repaired, replaced, or monitored. An accurate estimate of local scour around hydraulic structures such as around bridge piers, would be very useful. In a laboratory setting, the amount of

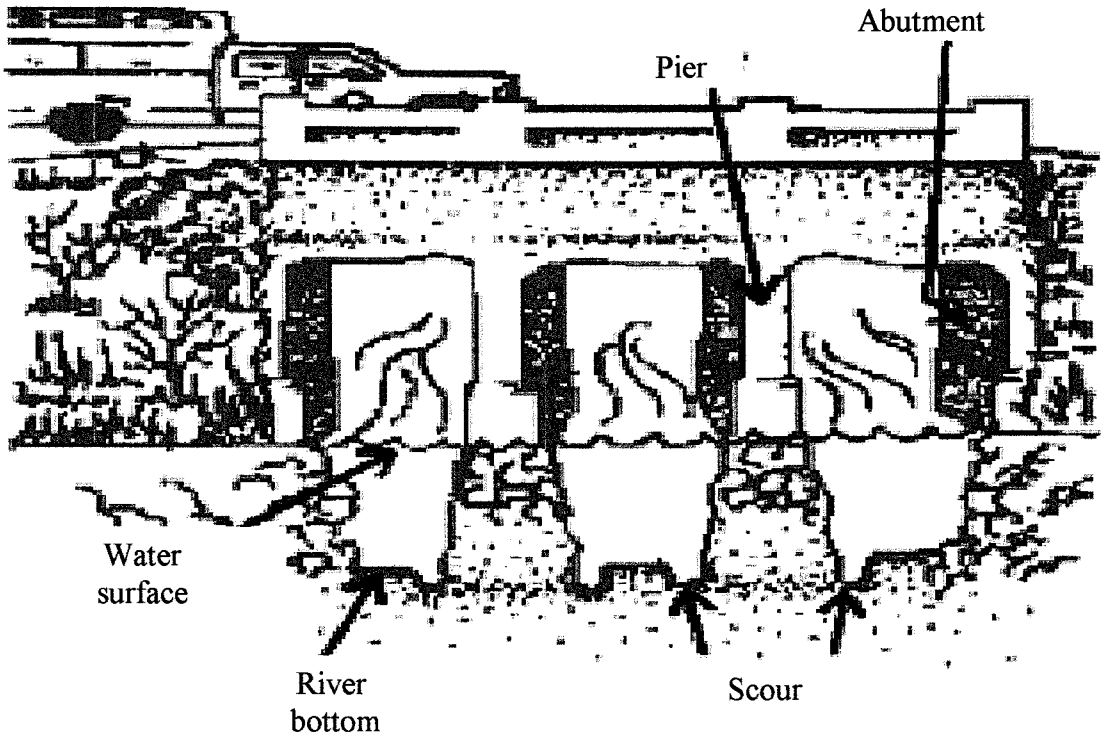

Figure 1.1. Bridge Scour

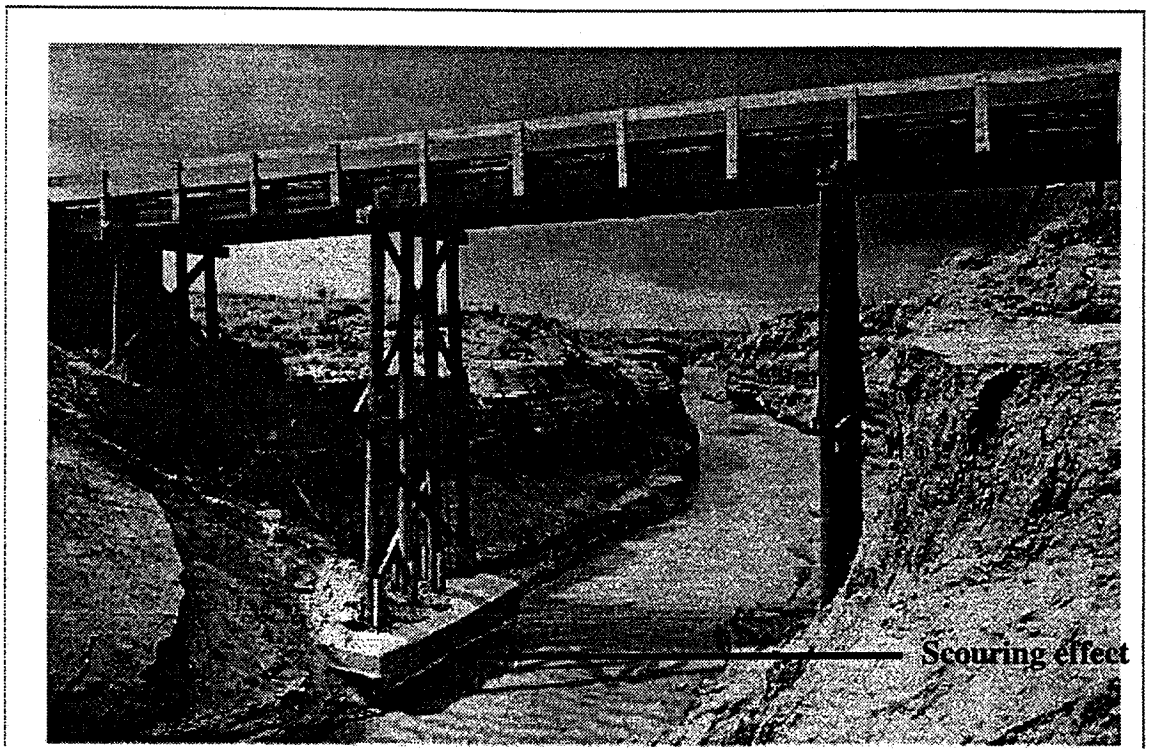

Figure 1.2. Bridge scour on Rio Puerco River, New Mexico. Photo courtesy of Steven D. Craigg, U. S. Geological Survey

erosion is generally quantified by noting the depth of scour at various locations in the bed. In steady flows, the use of conventional single point measurement devices to obtain the depth of scour yields fairly accurate results. However in unsteady flows, determining the depth is not a trivial task and involves a three-dimensional analysis of the scour field. The unsteady flow creates non-uniform surfaces all along the sand bed. Furthermore, it would be most useful if the depth measurement was non-intrusive. The laboratory data are extremely useful in developing computational models for predicting scour.

1.2. Two-dimensional imaging technique

Recently a non-intrusive scour depth measurement based on a video image analysis has been established for use in two-dimensional flow fields [2]. In this video image analysis, the scour field is captured using a Hi-8 video camera and stored as an image file. This image is later analyzed to determine the scour depth by applying thresholding values on the pixel parameters RGB (Red, Green and Blue components). The results obtained using this technique have shown to be very useful for obtaining the instantaneous water surface and the corresponding bed scour profiles for a two-dimensional flow field. But in real life, scour is not two-dimensional. This requires the development of a three-dimensional tool to analyze the scour field.

1.3. Three-dimensional imaging technique

There are various three dimensional imaging techniques available to obtain scour information. The 3-D shape can be determined from the grayscale intensity values; this concept is known as "shape from shading". A single camera is sufficient for this purpose and some apriori known position and knowledge of light source orientation is required. However, the depth measurement may not be accurate if there is noise in the image. The 3-D imaging can also be done by tomographic reconstruction which can produce a set of voxels (the 3-D analog of pixels) that represent the digitization of 2-D images. A variety of different signals can be used for tomography which including seismic waves, ultrasound, magnetic resonance, gamma rays, neutron beams, and electron microscopy to mention few.

The 3-D reconstruction can be obtained by serial sectioning of the images which is commonly used in biomedical applications. To achieve useful results using this technique, the slices of serial sections should be properly aligned and this is not a trivial task. Laser scanning and ranging techniques can also be used for 3-D mapping. Stereovision is one other technique that is widely used in the movie industry and satellite imaging. The term stereo is derived from the Greek word stereos which means firm or solid. This technique mimics the function of the human eye and brain to obtain depth information.

1.3.1 Stereovision

To obtain scour profiles in 3-D in a fairly accurate manner stereoscopy is suggested in this work. With stereovision one can perceive the object in three dimensions (x, y, z).

In our daily activities we unconsciously measure depth or judge distances to a vast number of objects about us through our normal process of vision. Determination of depth may be either a stereoscopic or a monoscopic process. Human beings with normal vision (those capable of viewing with both eyes simultaneously) are said to have binocular vision, and perception of depth through binocular vision is called stereoscopic viewing. A method of judging distances with only one eye is termed as monoscopic viewing. Monoscopic methods of depth measurement can be perceived on the basis of relative sizes of objects, shadows, shape of the objects, precedence, hidden objects, orientation of the object, shades, color, perspective and various other cues. Monoscopic methods of depth perception enable only rough impressions to be gained of distances to objects. On the other hand a much greater degree of accuracy can be obtained with stereoscopic viewing [3].

Depth perception is a difficult task because our initial visual representation is only two-dimensional. The fact that both eyes get slightly different views of the world is called disparity. The perception of three-dimensions can be produced when we view stereograms in which each eye receives the same image but in a slightly different view. The brain integrates these two images to see a single three-dimensional image. In other words, the disparity between the two scenes enables the observer to perceive depth.

1.4. Objectives

The present study aims at developing a practical video image analysis technique to estimate scour in three-dimensional flow fields. The developed algorithm was tested on standard grayscale images, standard color images and laboratory images.

1.5. Outline of the Thesis

The suggested method, to determine the depth of scour is based on stereoscopy. The left and right images are acquired by mounting a video camera at different viewpoints to simulate a stereoscopic viewing.

The stereoscopy scheme employed in the analysis involved the use of the only available geometric constraint called the epipolar constraint (given a point on the left image, the corresponding point on the right image is constrained to lie on a line called epipolar line). The epipolar constraint is used to reduce the problem from two-dimensions to one-dimension.

To determine the corresponding points in the images a robust relaxation technique is used. The features used in the relaxation technique are the vertical edges detected using standard edge detectors. The choice of edge detector depends on the quality of the images. A classical correlation technique is used to eliminate false matches. The depth of scour is determined using the disparity between the matched points.

The algorithm is first tested on several standard grayscale images. The algorithm is then extended for use on standard color images. Finally the algorithm is tested on non-standard color images acquired from the laboratory and the results are analyzed. The depth information is obtained by determining the parallax between the matched points.

2. BACKGROUND AND LITERATURE REVIEW

From two images of the same scene taken from slightly different view points, the three dimensional information can be extracted by fusing both images into a cyclopean image. The brain carries out the fusion process in quite a remarkable way. On the other hand, computer fusion of images is a difficult task, and requires locating the corresponding points in both images.

2.1. Human visual system

The basic concept of how a human eye views an object is best explained in Figure 2.1 (http://www.hhmi.org/senses/b/b150.htm). The distance between the two eyes is about 63 to 69 mm and is called the eye-base. The effective power of the eye is approximately 58.8 diopters (i.e., a focal length of approximately 17 mm). The anterior surface of the lens has a radius of curvature of approximately 10 mm, and the posterior surface has a radius of curvature of approximately 7 mm. The layer of light-sensitive nerve cells that line the inside surface is called retina. The optics of the eye forms a real image of the object on the retina. The retinal image that forms has an inverse orientation with respect to the object [4].

The initial event in the image perception process is absorption of light by visual pigment contained in the photoreceptors of the retina. There are two types of photoreceptors, the rods and the cones. They are named so because of their shapes. There are approximately 120 million rods and 6 million cones in the retina. The visual pigment associated with rods is rhodopsin. This consists of a lipoprotein called ospin and a chromophore (light absorbing chemical) called 11 cis-retinal. 11 cis-retinal has a spectral sensitivity peak at 500 nm. The cones contain three visual pigments with spectral sensitivity peaks at 450 nm (blue cones), 525 nm (green cones), and 555 nm

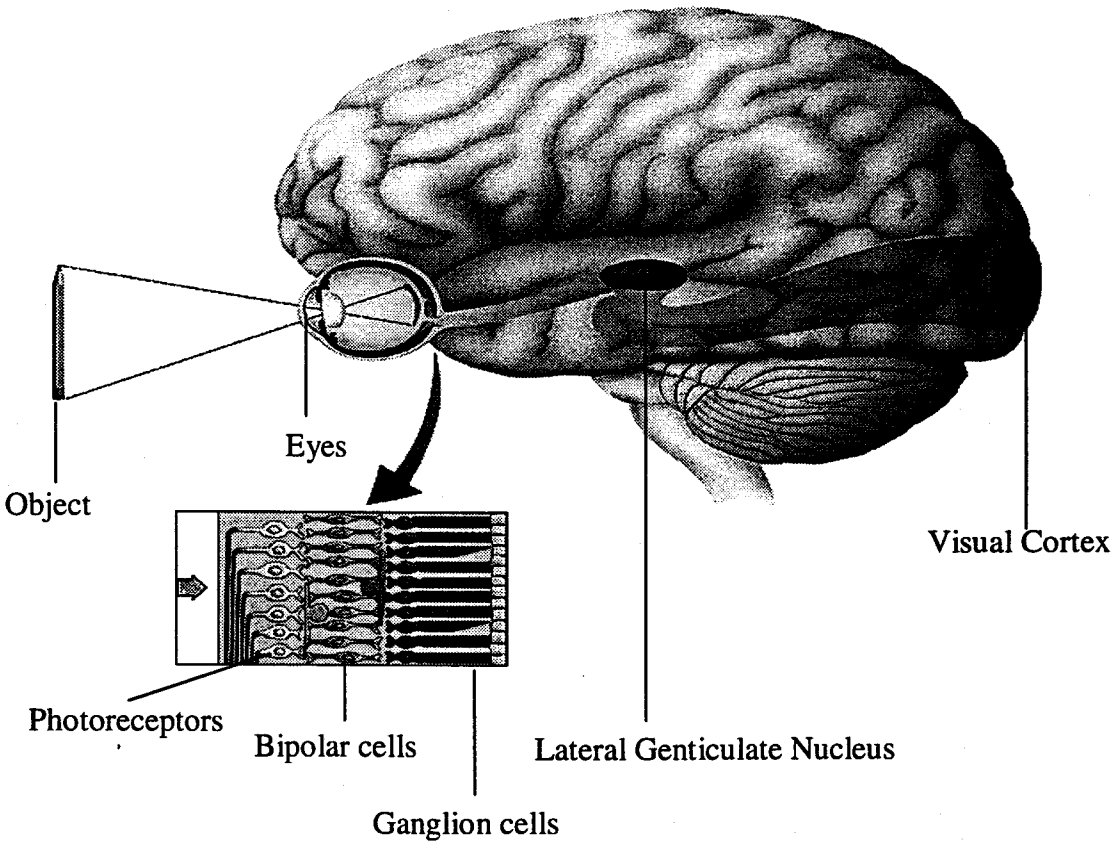

Figure 2.1. Visual pathway

(red cones). The three cone pigments share the same chromophore as the rods, and their different spectral sensitivities result from differences in the ospin levels. These photoreceptors generate a transmissible neural impulse whenever light falls on the

retina. These neural impulses pass ultimately to the ganglion cells, which are the final common pathway for the neural output from the eye.

The optic nerve is formed by the axons of the ganglion cells. There is considerable convergence of the photoreceptor output from 120 million photoreceptors to about 1 million ganglion cells. The optic nerves from the two eyes meet at the optic chaism [5]. The brain fuses these two images to form a single image called a cyclopean or virtual image. The fusion process is a response that involves several stages. The first response is the motor response that causes both eyes to look at the point of interest, or the visual axes to converge toward it. This process is called vergence, and the point on which the visual axes are converged is called the point of fixation. This means that the human eye finds the depth of a scene at one point at a time and not the entire scene at once.

The left eye captures the left view of the image and right eye captures the right view of the image as shown in Figure 2.2 (http://www.vision3d.com/stereo.html). The two images arrive simultaneously at the visual cortex. The brain combines the two images by matching up the similarities and adding in signals proportions to the small differences. The small differences between the two images, which is called retinal disparity, add up to a big difference in the final picture. Retinal disparity is the difference between the lateral position of object in the left and right eyes. The combined cyclopean image, is more than the sum of its parts; it is a three-dimensional stereo picture.

2.2. Stereoscopic depth perception

The displacement of the points in the two images formed by a physical point, termed parallax or disparity can be used to identify depth between the objects. The brain basically compares which one is closer by shifting attention from one feature to another. Figure 2.3 explains the principle of stereo fusion, in which the images from each eye are compared to locate the same feature in the left and right views. The optical muscles move the eye to bring this feature to the fovea, and the muscles provide the vergence information to the brain. This difference in the vergence angles from each view gives the relative distance to each feature in the left and the right view. The difference in the relative distances d_1 and d_2, in front of the eye called parallax can be used to identify the depth h [6].

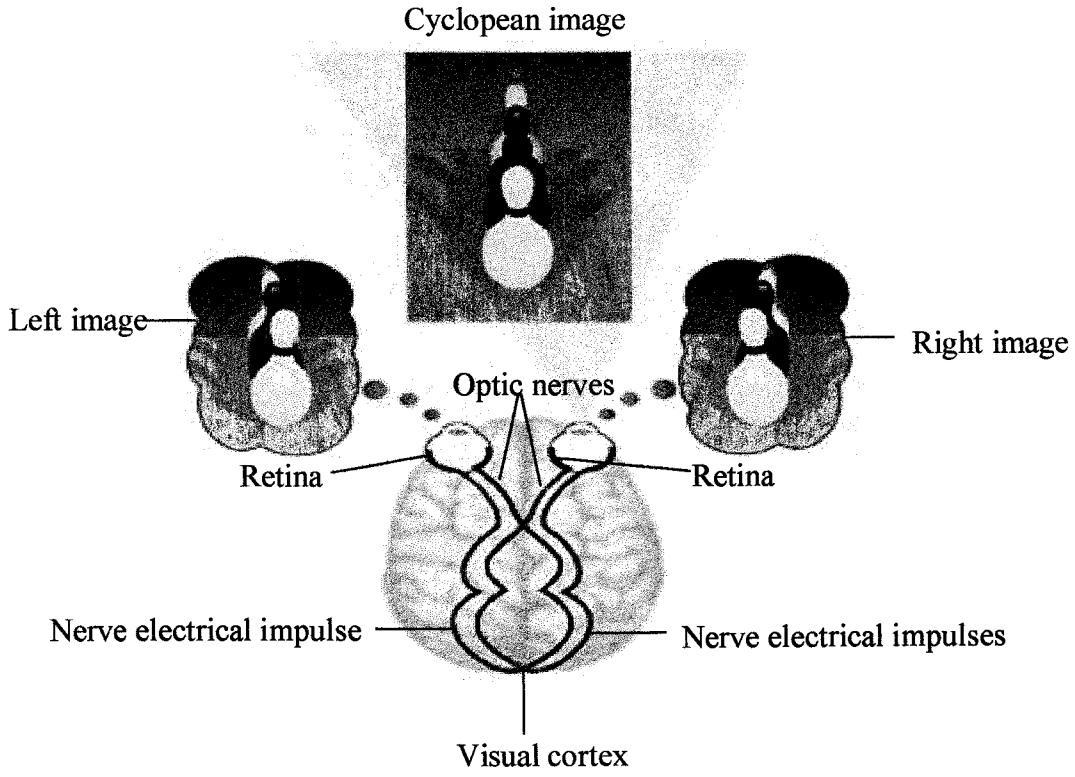

Figure 2.2. Human visual system

2.3. Correspondence imaging

As indicated earlier, to determine the parallax displacement, one needs to identify corresponding points in both the images. The process by which corresponding points are found is called correspondence imaging. This process still remains one of the bottlenecks in computer vision. Several algorithms have been proposed and they can be broadly classified into two categories: (1) Area based matching and (2) Feature based matching. Both the methods have advantages and disadvantages.

2.3.1. Area based matching

Under this category, a target pattern or template is shifted over each point in the image. The pixel values of the source are multiplied by the pixel values of the target that are overlaid to get a particular value. The pixel which shows the highest value is then selected as the best match. This technique works relatively well for textured images and if proper lighting conditions are ensured. The technique fails when the number of features are limited and if the viewing angles are wide. The area correlation method also requires much computational capacity.

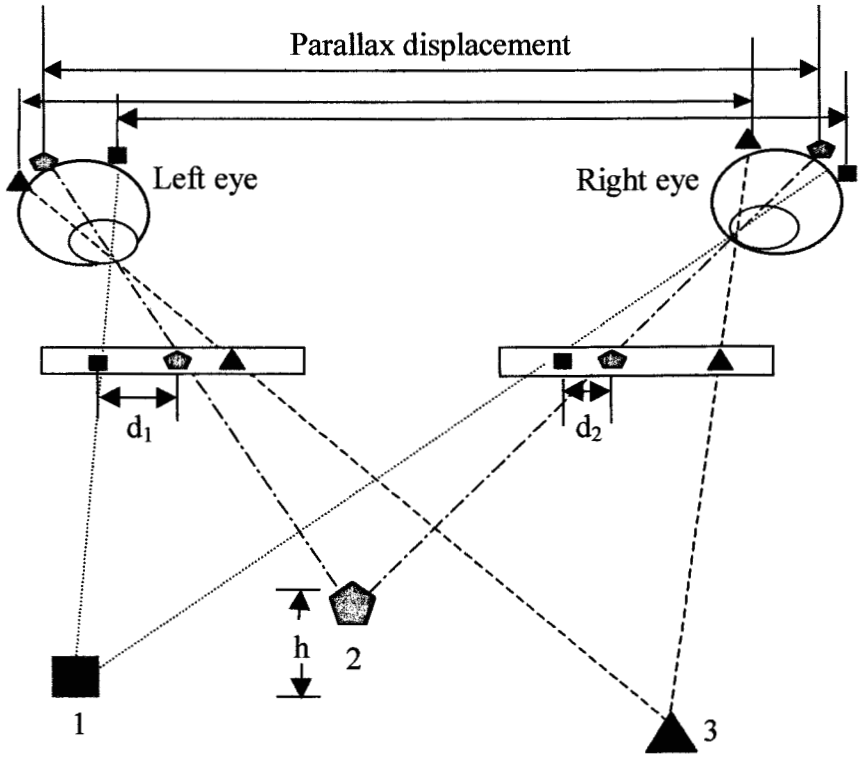

Figure 2.3. Stereoscopic depth perception

The back matching technique proposed by Hannah [7] increases the confidence level (correlation value) of matching but the computational complexity increases by two fold when compared with the conventional correlation technique. The method starts with a point in the image with a high resolution and moves towards lower resolution images

until it reaches the size of a correlation window. The corresponding point is then found in the small size window. In the next stage, the images are swapped and the same operation is repeated for only the previous corresponding points. This yields reliable results but increases the computational time enormously over the normal technique.

The two-stage cross-correlation method [8] is based on the idea of investing no time on positions which show no match at all. In this method, the results obtained are much faster than the results compared with the classical correlation technique which matches the template based on the correlation scores of the whole area. The method is also faster than the matching done with the FFT (Fast Fourier Transform) algorithm.

The patch based method [9] has been proposed to operate under arbitrary camera geometries, which means it can work even when there are large deformations in both the images. This algorithm also claims that it is not applied to narrow base line or parallel-axis situations. In this method, the image is sampled with a Gaussian filter and the outputs of the sampled points are compared to determine the match error. Using this method, the interimage transformation is computed by the depth, slant and tilt values which are obtained by highly accurate computer-controlled stepping motors. It should be remarked that it is not practical in many cases. Also the errors obtained vary with the patch size i.e., if the patch size is large, then the errors get reduced considerably but the computation increases.

2.3.2. Feature based matching

Under this category, the correspondence imaging is based on the features or tokens such as edges, lines, contours, and polygons from the two or more views. This approach is widely used nowadays as the method is faster than the area based approach and uses only a very small set of pixels. This method fails if the identified tokens are not reliable in the different view of the images.

Marr and Poggio [10] proposed a co-operative algorithm, which propagates two important constraints, uniqueness (each image point will participate in only one match). and continuity (based on the idea that the world is mostly made up of objects with smooth surfaces). The process of finding corresponding points is done iteratively and defined by a set of difference equations. The disparity based algorithm is described in

the form of a network, which has "inhibitory", and "excitatory" connections based on the disparity values. This algorithm works well for random dot stereograms but the results are not satisfactory for real images.

Pollard *et al.* [11] proposed a constraint algorithm based on the disparity gradient limit. This algorithm exploits the uniqueness constraint and the epipolar constraint (given a point on the left image, the corresponding point on the right image is constrained to lie on a line called epipolar line). The algorithm uses vertical edges as features and uses a disparity gradient limit value of one. Burt and Julesz [12] demonstrated that human eyes could fuse the object only if the disparity gradient value does not exceed the critical value of one. The disparity gradient is defined by the disparities of the neighboring objects divided by the cyclopean separation. The geometry of the stereograms is explained in terms of the imaginary dipoles that connect dot pairs as seen by left and right eye and a disparity gradient limit value of less than or equal to one is used. When the disparity gradient limit value exceeds a value of two, fusion will not occur. The Pollard, Mayhew and Frisby [11] algorithm normally converges within four or five iterations and correct matches are chosen based on the matching scores. The results were found to be accurate and reliable for both random dot stereograms and natural images.

Deriche and Faugeras [13] approached the problem of matching the points between the two images by extracting edges in the images using a previously developed edge detector. The points are linked using a chain-based descriptor. The high curvature points are extracted by the rate of change of the angle of the edge-linked chain. The high curvature points are matched by using the epipolar constraint. Finally the curves are matched from the matched curvature points by using the figural continuity.

Medioni and Nevatia [14] matched the corresponding co-linear connected edge points called segments. The coordinates of the end points, the orientation and the strength (average contrast) describe these line segments. For global matching, a minimum differential disparity criterion was used. The exact coincidence of both the images was assumed so that the corresponding points lie exactly on the same horizontal line. For each possible match they computed an evaluation function and preferred the

matches with the lowest values. The evaluation function uses the disparity values, and matches with similar disparities were preferred to find the corresponding segment. The function was computed iteratively and the value was changed based on the new preferred matches. This algorithm cannot be guaranteed to attain convergence and was stopped after three iterations in any case. The algorithm was tested on many complex images and found to be performing well. However, it generates only a sparse disparity 3-D map based only at the points of the matched segments.

Chang and Aggarwal [15] applied a spatial and temporal grouping process to the corresponding line in two images. The Canny edge operator [16] was used to detect the edges and extract the lines using a pre-existing object recognition toolkit package. The improvement of this algorithm over the previous work is that it uses the spatial and temporal relations at the same time and use one relation as supporting evidence for the other. Initially, a line in the first image has multiple candidates in the second image. To resolve ambiguity and find the corresponding points, a relaxation labeling algorithm is used. The relaxation technique uses the information from the line's neighbors and propagates the constraint information. The line's neighborhood update process uses an operation which is obtained by combining the results from the temporal and spatial grouping process. The epipolar line constraint is not applied in this work. This algorithm is warranted where lines can be reliably extracted.

Horaud and Skordas [17] preferred feature grouping and maximal cliques to solve the correspondence problem. The feature they used was line segments extracted from pre-existing edge detector and edge linker. In their approach, they omitted the continuity constraint, which states that the surface has a smoothly varying disparity constraint. A relational graph is built from each image based on the position, orientation and relationships with a nearby segment. Then a correspondence graph is built from the relational graph. The largest maximal clique in the correspondence graph is identified as the best match. A benefit function is associated with each maximal clique. The maximal clique, which maximizes this benefit function, is chosen as the best match. This exhaustive search improves the quality of the results but tends to be very slow when compared to other methods.

Greenfield and Schenk [18] used polygonal approximation and Ψ - s representation of an edge to solve the correspondence problem. Here, Ψ represents one axis mapping cumulative orientation changes along the course of the curve, and the other axis s represents the cumulative curvilinear length of the curve. The primitives they use in the polygonal approximations are angles and orientations of the vertices, strength and zero crossing signs. The number of matches found by using this method varies as the sigma value of the Laplacian of Gaussian operator is varied. Greenfield and Schenk [17] tested the Ψ - s method on some limited number of edges and the found the results to be good. This method fails when the primitives are lines rather than curves because straight lines have the same Ψ - s representation, regardless of their orientation. They incorporated a voting scheme to resolve the ambiguity arising from incorrect or ambiguous matches. This matching scheme does not require any constraints such as epipolar constraints.

Hsieh *et al.* [19] used both area based and feature based approaches to obtain the 3-D information of complex aerial imagery. The epipolar geometry constraint was satisfied in their approach. The problem of selecting the control points i.e., the points of interest in the image which give good depth information, was addressed. An iterative selection algorithm was proposed to automatically select the control points. In the area based approach, they used the method of differences to find the corresponding points in the images. In the feature based approach, they used hierarchical waveform matching to find the corresponding points. A similarity function based on three components, intensity, shape and description, was defined. They combined the disparity map obtained from the area based and feature based approaches, and their results are an improvement over the individual methods. The control point selection problem was addressed, but in their scene registration, they considered manual selection of control points was still the best.

2.4. Camera Geometry

Camera geometry is one of the most important aspects considered in stereovision. Woods et al. [20] discussed the image distortions arising in stereoscopic video systems. They discussed keystone distortion, and lens distortion, and limitations on depth range. The keystone distortion arises when toed-in camera configuration is used. The lens radial distortion is caused by the use of spherical lens elements, resulting in the lens

having different focal lengths at various radial distances from the center of the lens. They developed a coordinate transformation from object space to image space coordinates to avoid some of these distortions by appropriate choice of system parameters. Parallel camera configuration is suggested over the toed-in camera configuration.

Yu *et al.* [21] discussed the various distortions and an accurate calibration method is suggested for stereoscopic system with a sub-pixel technique. Tsai's [22] universal method for calibration and correction of the coordinate system was applied.

2.5. Epipolar Geometry

Recently Zhang *et al.* [23] developed a robust technique for image matching by exploiting the epipolar geometry. The camera geometry and the camera parameters are unknown in their approach. The paper discusses how to recover the unknown epipolar geometry if two uncalibrated images are given. The points of interest on the two images are extracted by applying a corner detector. As a first step, the corresponding points are found by applying a classical correlation technique. For a correlation window in the left image, a corresponding search window is formed on the right image and the points are searched for correspondence. The points are selected based the correlation scores. The correlation score used in their approach is 0.8 and a correlation window of size 7 x 7 pixels is used. A point in the left image may be correlated with several points in the right image. In the second step, to resolve the ambiguities, a classical relaxation technique is used. In the relaxation process an energy function is defined to solve the ambiguity arising from the matches. The matches are updated with a new strategy which is called "some-winners-take-all". Next, the outliers are detected and the false matches are discarded using the least median of squares method. Finally, the fundamental matrix is found by using an eight-point algorithm given by Longuet-Higgins [24] which precisely recovers the epipolar geometry. By using the least median of squares method it has been ensured that the fundamental matrix is estimated correctly by avoiding false matches and bad locations. When tested on several images, the algorithm yielded very good results. But one pitfall using the eight-point algorithm is if the corresponding points identified are not accurate it gives inaccurate epipolar lines.

Hartley [25] suggested that the eight-point algorithm is extremely susceptible to noise and some times gives a useless fundamental matrix. He suggested that by using a very simple translation and scaling of the matched points, very good results could be obtained. The centroids of the points are translated to the origin and the points are then scaled so that the average distance from the origin is equal to $\sqrt{2}$. The theory has been verified by extensive testing on several real images.

2.6. Color as an added feature for stereovision

All the algorithms and techniques thus reviewed have used grayscale images. Color is widely used in many image processing applications and there are many commercial packages available for color processing. But the color information is not often used for stereovision because of the complexity (three channel Red, Green and Blue has to be separately processed) involved. Jordan and Bovik [26] considered the chromacity information for finding corresponding points. Using the red and green channels, they compared their color-matching algorithm with the intensity-matching algorithm and obtained useful results. For finding matches the sign and orientation of the color gradient was used.

Three different models of incorporating color in stereoscopy were investigated by Brockelbank and Yang [27]. The three models considered by them were: (i) the Three Channel Model - RGB channels were treated separately and edge detection was performed on each channel, (ii) the Trichromatic Stereopsis Model - edges were detected, separately, but they were immediately combined into one feature map, and (iii) the Opponent Color Stereopsis Model - the three channels were combined and then the edges were detected to get the feature map. The relaxation surface smoothing proposed has the ability to correct the assignment errors and claimed a slight increase in the percentage of correct matches.

The measurement of average travel time of a vehicle by object matching was tried by Dubuisson and Jain[28]. The color indexing based on histogram matching was used in their work of 2D matching of 3D moving objects in color outdoor scenes. Further, matching is done by finding the Hausdorff distance between two sets of edge points. The extraction of the object of interest was performed by fusing the multiple cues including motion, color, edges, and model information.

The block matching technique for color images were performed by A. Koschan *et al.*, [41]. In this technique, the similarity check was performed by finding the mean square error (MSE) between the left and right images. The problem of finding the corresponding points for homogenous image regions were addressed by illuminating the scene with color coded light. A rainbow like color spectrum was projected onto the scene which was synthetically generated. Considerable improvement in finding the corresponding points was shown in their work by using active color illumination.

3. VIDEO IMAGE ANALYSIS

3.1. Two-dimensional image analysis

To determine the effects of scouring, the amount of erosion occurring at any given period of time has to be determined. The use of a conventional point gauge to obtain the water surface and sand bed erosion profile yields fairly accurate results if the flow is steady. However, in unsteady and quasi-steady flows, the point gauge is not of much use. To get depth information even in unsteady flow a non-intrusive two-dimensional video image analysis procedure is first developed.

3.1.1. System hardware

The system hardware, as shown in Figure 3.1. essentially consisted of a Hi-8 CCD-Video camera (Sony TR-600), a TV monitor, a Hi-8 VCR (Sony EVC-200), and a computer fitted with a frame-grabber board. The video recording was carried out at a rate of 30 frames per second. The video images were stored on 8 mm tapes and played back on the Hi-8 VCR with a frame-by-frame advance feature to obtain the water surface and corresponding bed scour profiles at any instant. Throughout the video taping, the flow field was uniformly illuminated. Thereafter, selected frames of the video images were digitized using a commercially available image acquisition software (Ulead System) and stored as 640 x 480 pixel standard bitmap files.

The video camera captures an image in a color format which contains three color components, namely red, green and blue. The color recognized on a computer screen or a video monitor at each pixel is a combination of these three color components. From the digitized images of the flow field, the values of the pixel intensity can be characterized by the red (R), green (G) and blue (B) values. The values of R, G and B varies individually from zero to 255. Since each pixel is represented individually by all three values which are in the range of zero to 255, the combination of these three values

can give up to 256 x 256 x 256 colors. The list of colors formed by these components is given in Table 3.1.

Table 3.1: Pixel color definition

Red (R)	Green (G)	Blue (B)	Pixel Color
255	255	255	White
0	0	0	Black
255	0	0	Red
0	255	0	Green
0	0	255	Blue
Any value	Any value	Any value	Any color

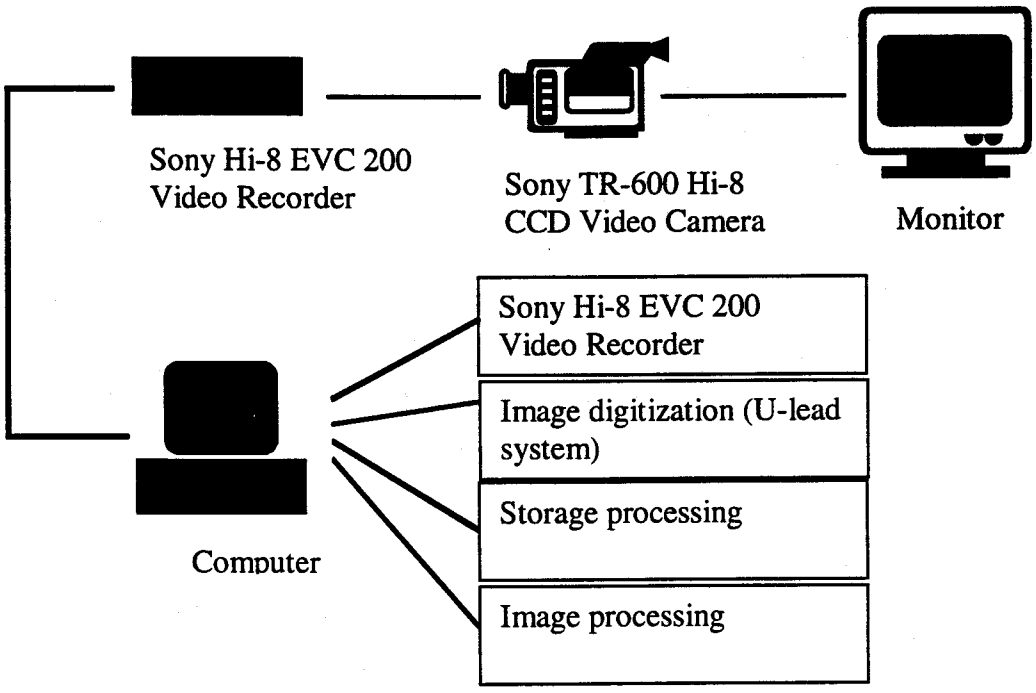

Figure 3.1. Schematic arrangement of video-imaging and analysis system

3.1.2. Thresholding technique

From the captured video images it was possible to qualitatively identify three zones: (i) the space occupied by air above the water surface, (ii) the water layer, and (iii) the sand bed. To analyze the digitized images and convert the pixel RGB information into useful data, a thresholding technique is used. To detect the features in an image and carry out various kinds of measurements, this technique is employed. One traditional way is to define a range of intensity values in the original image, and based on the intensity values, the pixels in the original image are classified into foreground and background pixels. These images are called binary or two-level images. In this study, based on the RGB values, the image were segmented into three zones, namely, air, water and sand. Appropriate selection of threshold values is very important for segmentation. There are various techniques to select the threshold value but the technique most frequently employed is proposed by Weska [29] for a grayscale image and explained in Fig 3.2. In this technique, a minimum value (between the maximum peaks) found is chosen as a threshold value. Here I is the intensity value and *f(I)* is some function of intensity value.

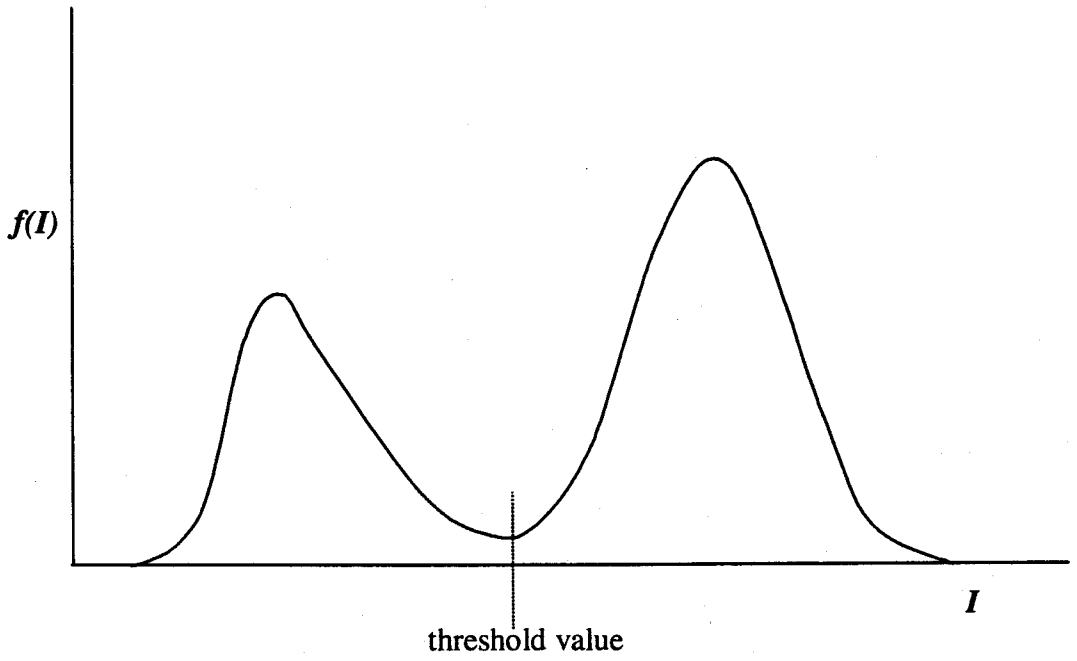

Figure 3.2. Histogram of pixel intensity grayscale levels in an image

In the ideal case as shown in the Figure 3.2 the smaller peak on the left is due to dark foreground objects and the large peak on the right is due to light background. The minimum of the distribution curve is an optimum intensity value to use as a threshold.

Following the determination of the threshold value in a standard image, the noise level of the subsequent images will affect the segmentation process, so the manual selection of the threshold values is preferred. Typical RGB values in the air, water and sand zones are as shown in Table 3.2.

Table 3.2: Typical RGB values in the three zones

Zone	Red (R)	Green (G)	Blue (B)
Air	~ 255	~ 255	≥ 232
Water	180 to 255	160 to 255	140 to 224
Sand	≤ 100	≤ 72	≤ 40

The required features of the program included: (i) identifying the relative location of each pixel with respect to the edge of a known non-erodible bed, (ii) acquiring the RGB information at each pixel, and (iii) using the RGB information to identify if the pixel location in the image referred to air, water or sand. The segmented image after applying thresholding is shown in Figure 3.3.

The current pixel coordinates are indicated as CurrX and CurrY. Below this display are indicated the pixel R, G and B values for the current cursor location. A quick survey of the image can be easily carried out by clicking the mouse at various locations and the lower limit values for the three zones can then be established. These values are thereafter used as threshold values to segment the image. The analysis can also be carried out only on a region of interest to reduce the run time significantly.

The use of video imaging and image processing using threshold values is very useful for obtaining instantaneous water surface and bed scour profiles in two-dimensional flow fields. This technique enables the measurement of the scour profiles in a non-intrusive manner at any instant of time during the erosion process and eliminates the

uncertainties associated with the traditional methods of obtaining such profiles. Moreover, important changes occurring in the profiles over short periods of time can be quantified very easily.

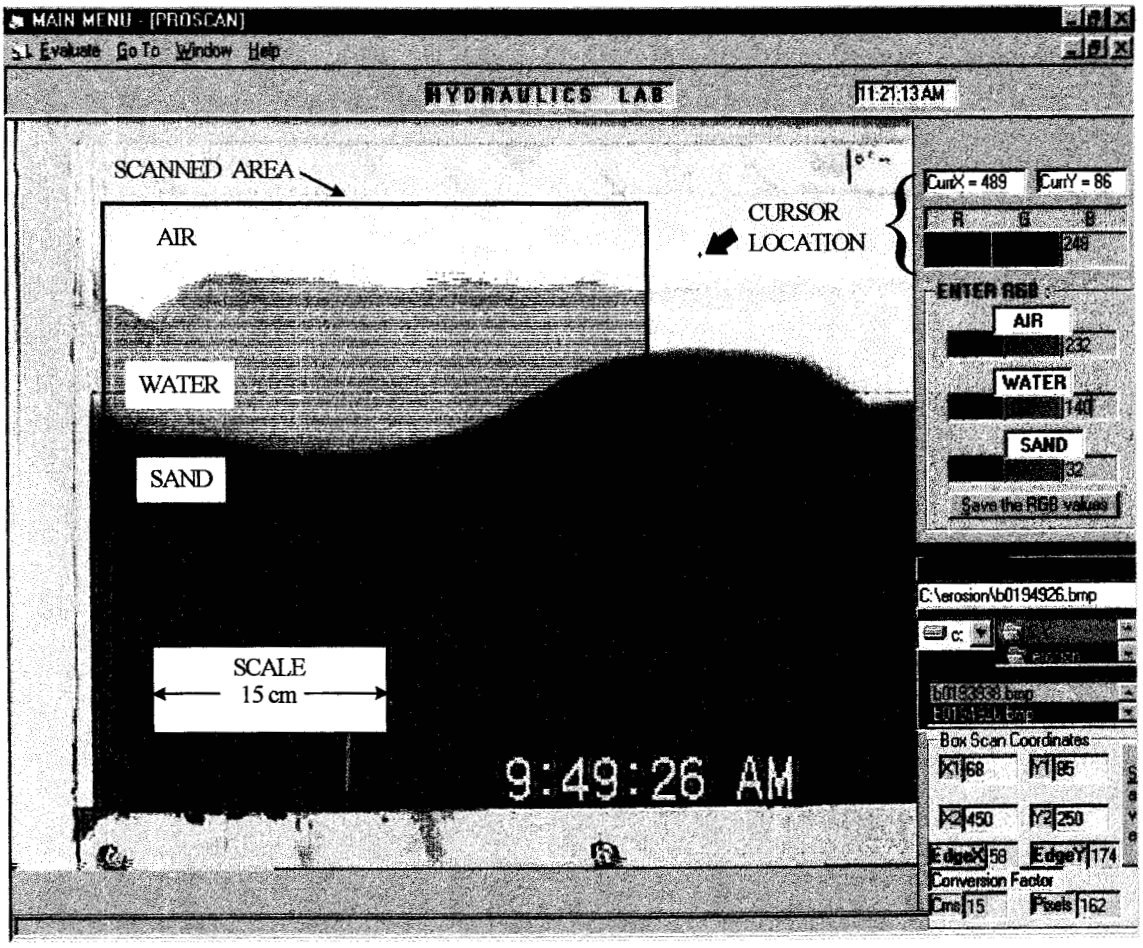

Figure 3.3. Typical segmented image after processing

3.2. Three-dimensional image analysis

The present study aims at developing a practical video image analysis for use in three-dimensional flow fields. To acquire images, two cameras with similar specifications were mounted over the model of the eroded bed and the two images were simultaneously captured. The stereoscopy scheme involved the use of an epipolar constraint and a relaxation technique to match corresponding points in the two images. A correlation technique was developed to eliminate false matches. The depth of scour was determined using the parallax between the matched points. This process is illustrated in Figure 3.4.

3.2.1. System hardware

The system hardware, as shown in Figure 3.5 essentially consisted of a Hi-8 CCD-Video camera (Sony TR-600), a TV monitor, camera(s) mountable stand, a Hi-8 VCT (Sony EVC-200), and a computer fitted with a frame-grabber board. In the camera mounting stand, provisions are made to mount the camera at different elevations over

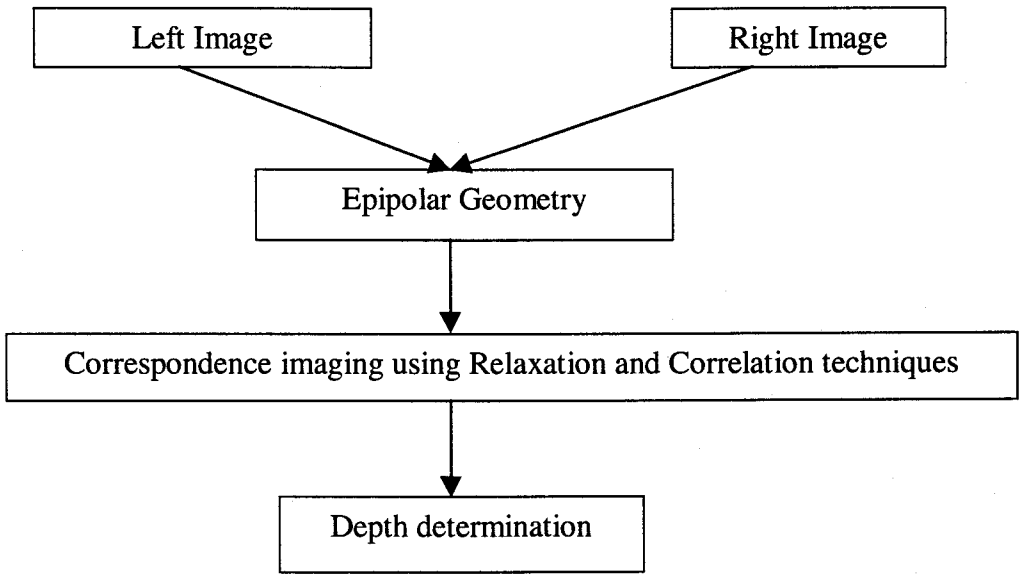

Figure 3.4. Three-dimensional imaging process

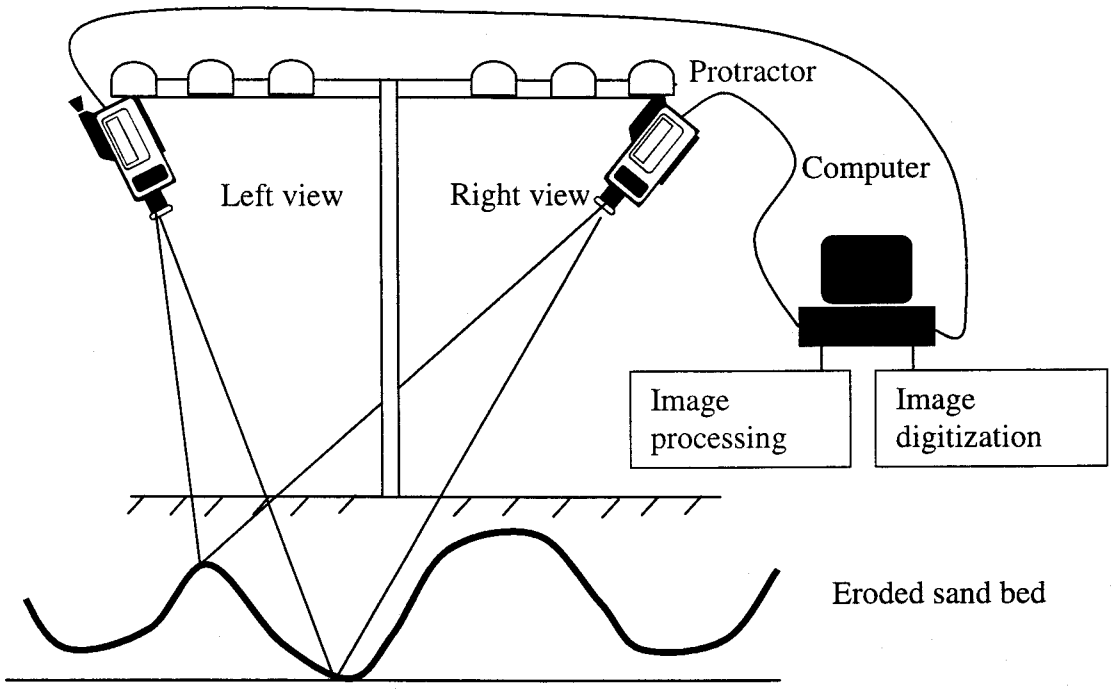

Figure 3.5. System hardware for three-dimensional imaging

the sand bed. The angle of view of each camera can also be varied. The system described in Figure 3.1 was used to analyze the images.

3.2.2. Epipolar geometry

To identify distances such as d_1 and d_2 in Figure 2.3 by comparing two images, one needs to locate the corresponding points in both images. For example, one needs to locate the square and polygon in both images to compute d_1 and d_2. In this work, to find the corresponding points in the two images the only available geometric constraint called the epipolar constraint [23] was exploited. The tremendous advantage of applying the epipolar constraint is realized in reducing the image matching problem from two-dimensions to a single dimension.

Consider two images I_ℓ (left) and I_r (right) as shown in Figure 3.6. Let C_ℓ and C_r be the optical centers of the left and right camera respectively. Given a point n_ℓ on the left

image, the corresponding point n_r in the right image is constrained to lie on a line (ln_ℓ) called the epipolar line of n_ℓ. The line ln_ℓ is the projection of the infinite-half line C_ℓ - N on the image plane I_r where N is the physical point. The point n_ℓ may correspond to any arbitrary point on the infinite-half line C_ℓ - N. The plane passing through the points n_ℓ, C_ℓ and C_r and intersecting the image plane I_r is known as the epipolar plane. All the epipolar lines of the left image pass through a common point called ep_r and this is called as epipole. ep_r is the intersection of the line C_ℓ - C_r with the image plane I_r.

The epipolar constraint is symmetrical and, for a point n_r in the right image the corresponding epipolar line is ln_r in the left image. The epipole of the first camera with respect to the second camera is ep_ℓ. The advantage of using epipolar line is that, to find corresponding points in both the images we need to search only along a single line and not the entire image. This reduces our problem from a two-dimensional search to a one-dimensional search.

3.2.3. Algorithm

The correspondence between a pixel and its epipolar line [24] is given by the expression

$$\hat{n}_r^T E \hat{n}_\ell = 0 \tag{3.1}$$

where,

$$\hat{n}_r = (x_r, y_r, 1) \tag{3.2}$$

is the image vector of point N on the right image,

$$\hat{n}_\ell = (x_1, y_1, 1) \tag{3.3}$$

is the image vector of point N on the left image,

E = Essential matrix, and (x_r, y_r) and (x_ℓ, y_ℓ) are the image coordinates of the point N in the right and left images respectively.

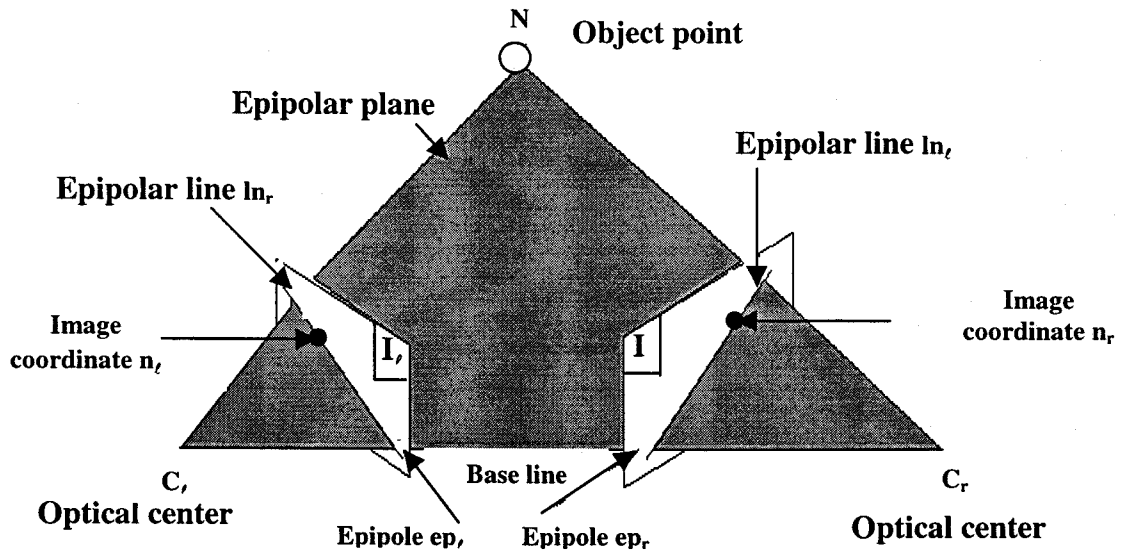

Figure 3.6. Epipolar geometry

Any pixel n_r on the epipolar line of ln_ℓ of n_ℓ which satisfies the equation (3.1) forms the corresponding epipolar line. There are various methods to solve the essential matrix [23, 24, 30, 31, 32]. The frequently used algorithm to solve the essential matrix is the eight-point algorithm and known as Longuet-Higgins algorithm [24].

Let $X_i = (x_i, y_i, 1)^T$, $X'_i = (x'_i, y'_i, 1)^T$, i = 1, 2, . . ., n, be the corresponding image vectors of n ($n \geq 8$) points and x_i, y_i = image coordinates in the left image and x'_1, y'_1 = image coordinates in the right image. Let

$$A = \begin{bmatrix} x_1x_1' & x_1y_1' & x_1 & y_1x_1' & y_1y_1' & y_1 & x_1' & y_1' & 1 \\ x_2x_2' & x_2y_2' & x_2 & y_2x_2' & y_2y_2' & y_2 & x_2' & y_2' & 1 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ x_nx_n' & x_ny_n' & x_n & y_nx_n' & y_ny_n' & y_n & x_n' & y_n' & 1 \end{bmatrix} \quad (3.4)$$

and let,

$$\mathbf{e} = (e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9)^T \quad (3.5)$$

The unit vector **e** has to be solved such that

$$\|A\mathbf{e}\| = \text{min.} \quad (3.6)$$

The solution is the eigenvector of A^TA associated with the smallest eigenvalue. The essential matrix E is determined by

$$E = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \quad (3.7)$$

In the present study, the method suggested by Zhang *et al.* [23] was used to compute the essential matrix from the two images. The concept of epipolar geometry is best visualized from Figure 3.7 and Figure 3.8. These are aerial photographs of the White House, Washington D.C., U.S.A. taken from two views (left and right view) (http://www.ius.cs.cmu.edu//idb/html/stereo/whouse/index.html). The width of epipolar line is set by setting the value in the right hand side of the equation (3.1) which is denoted as the epiwidth value. The essential matrix for these images is given by a 3 X 3 matrix,

$$E = \begin{bmatrix} 9.320398e-07 & -2.007319e-04 & 2.160543e-02 \\ 2.166033e-04 & 2.696489e-05 & -1.533646e-01 \\ -2.404708e-02 & 1.417785e-01 & 9.774114e-01 \end{bmatrix}$$

For the small white circles in Figure 3.8 (left image), the white lines (right image) represent corresponding epipolar lines in the right image.

In Figure 3.8, instead of a value of zero in equation (3.1), the epiwidth value is set between 0 and 0.1. So there are possibilities that for a point in left image, the corresponding epipolar line may not exactly pass through the corresponding point in the right image. The eight-point algorithm is very sensitive to noise. To avoid this problem, the width of the epipolar line may be increased by changing the epiwidth value in equation (3.1). The effect of increasing the epiwidth value from 0.1 to 0.3 is shown in Figure 3.8.

The estimation of epipolar line using the essential matrix was tested on several real images and found to be working well when visibly investigated the epipolar lines. However, if the corresponding points used to estimate the essential matrix is not accurate then it tends to produce false epipolar lines.

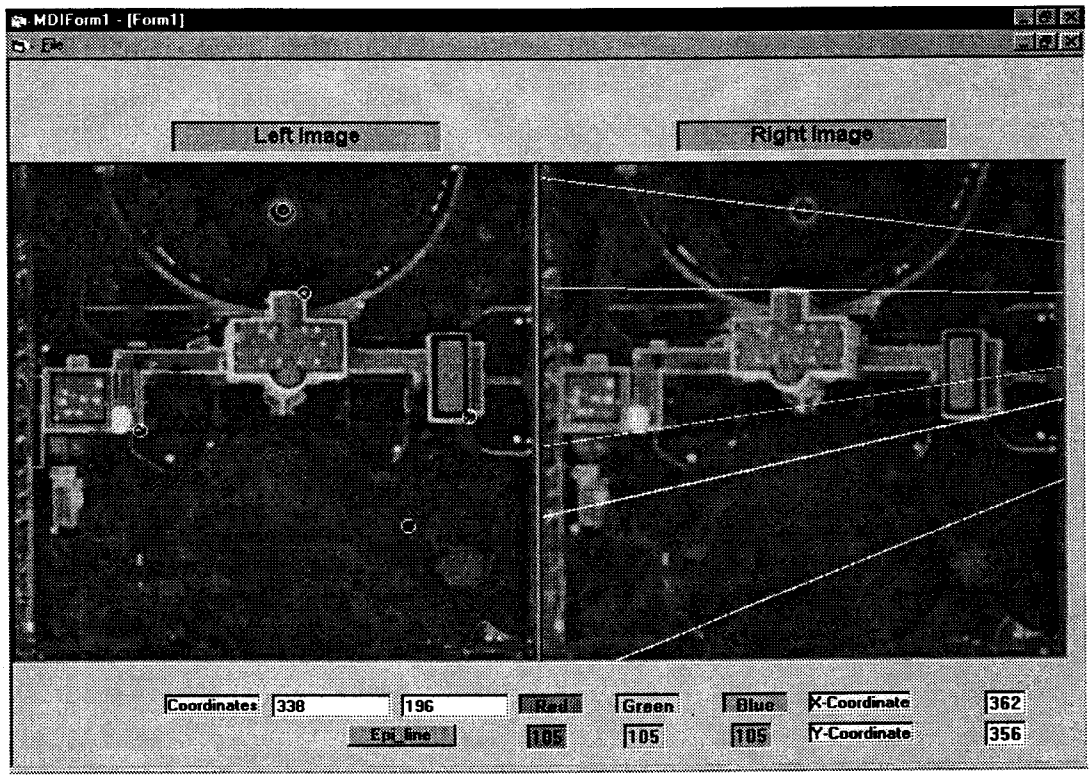

Figure 3.7. Epipolar lines for the corresponding point with an epiwidth value between 0 and 0.1 inclusive

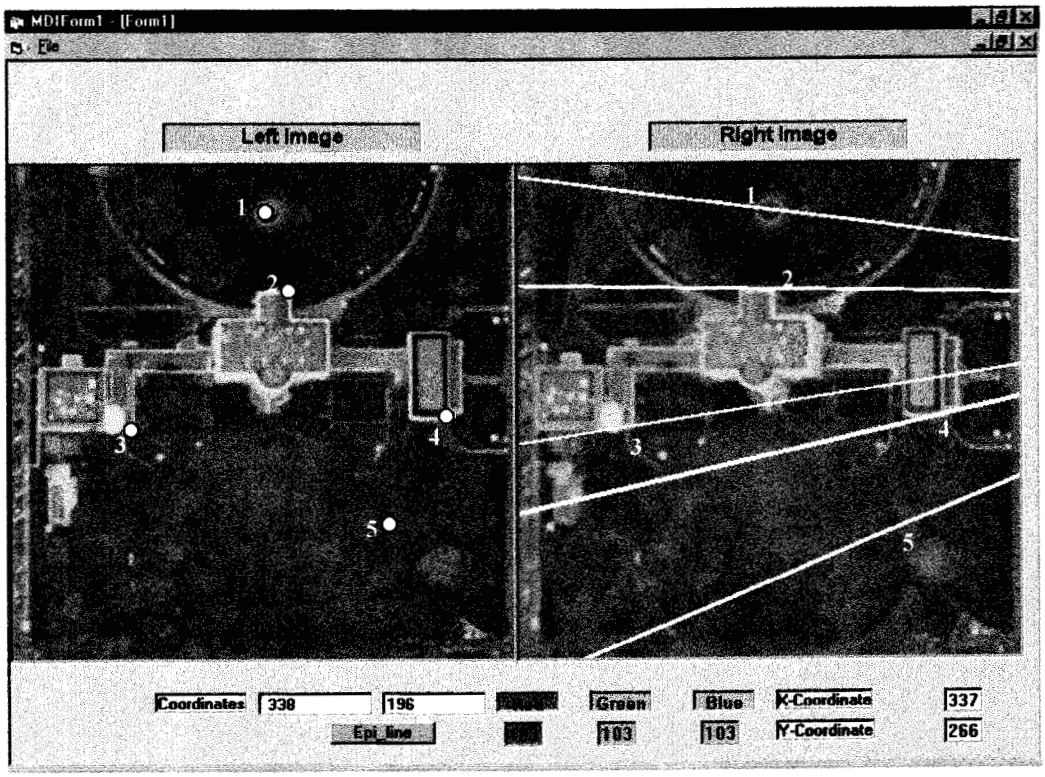

Figure 3.8. Epipolar lines for the corresponding point with an epiwidth value between 0 and 0.3 inclusive

4. CORRESPONDENCE IMAGING

Depth perception is attained by the fact that both eyes get slightly different views of the world from the same object. The disparity between the two scenes enables the observer to perceive depth. The corresponding points should be determined in the left and right images in order to calculate disparity. As discussed earlier, the process of determining the corresponding points in two images can be broadly classified into two categories, namely area based and feature based matching. In this thesis both methods have been used. The steps involved in finding the corresponding points in illustrated is Figure 4.1.

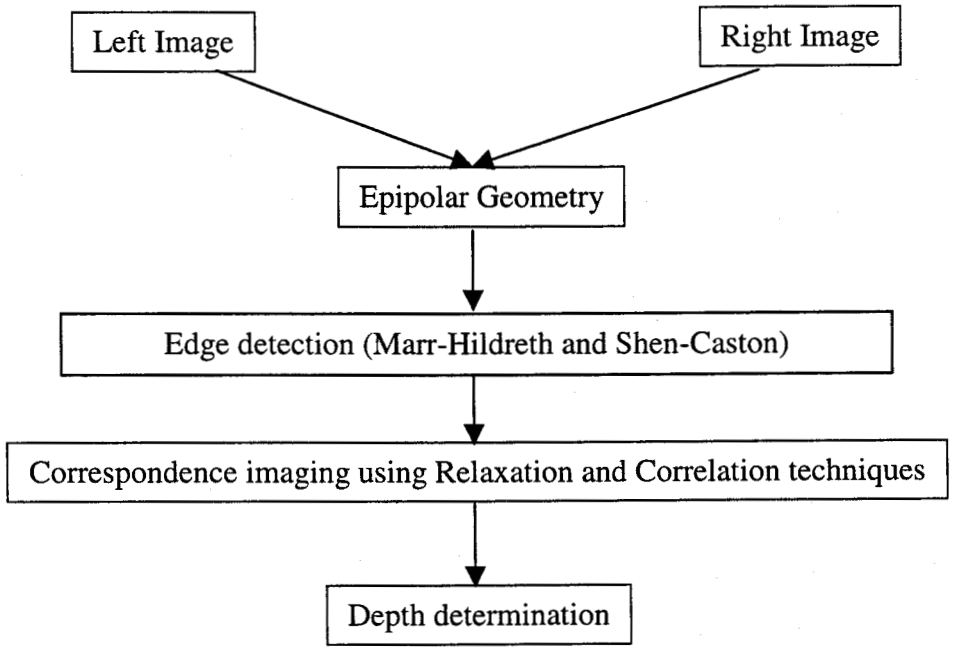

Figure 4.1. Three-dimensional imaging process

4.1. Edge detection

An edge may be regarded as a boundary between two dissimilar regions in an image. For example, an edge may be a boundary between light and shadow falling on a single surface. An edge is usually easy to find since the differences in pixel values between regions are relatively easy to calculate by considering gradients. Edges are very important to any vision system because they help in the recognition process by providing strong visual cues. Edges can be used to detect corresponding points in the left and right images. Edges, however, are affected by the noise present in the image since they are determined using differences between pixel values in order to obtain the gradients. A number of edge detectors have proposed over the years. The edge detectors used in this work are the Marr-Hildreth [33] and the Shen-Caston [34] detectors.

4.1.1. Marr-Hildreth edge detector

Edge detection is basically performed by the spatial filtering phenomenon in the retina and visual cortex. The Marr and Hildreth edge detection scheme has received much attention as it is based on biological vision. The edges are detected by identifying zero-crossings of pixel values in the image. The properties of a theoretical image are shown in Figure 4.2a. One way of locating the position of the edge in this image would be to take the first derivative of the intensity function, and then mark the edge location at the peak of the derivative (Figure 4.2b). Alternatively one can find the second derivative and mark the zero-crossing as the detected edge (Figure 4.2c). Real images rarely have a well-behaved intensity gradation at points that we subjectively would identify as a clean edge. A real image intensity function with noise may take the form as shown in Figure 4.3a. The first derivative for the noisy image (Figure 4.3b) may have several peaks and the highest peak may not be the edge. The second derivative will be even worse with many spurious zero-crossings, and the higher the frequency of the noise, the worse the problem gets (Figure 4.3c).

One possible solution to this problem is to pre-filter the image with a convolution operation that blurs out the fine detail which is presumably due to noise (Figures 4.4a & 4.4b), and then proceed with the differentiation. The problem, then, is to choose the degree of blur. Too much blur can result in missing the edge, while, very little can result

in false edges from the false peaks in the derivatives and zero-crossings. The pre-filtering is performed by convolving the image in two dimensions which can be expressed by

$$I * G(n,m) = \sum_i \sum_j I(i,j)G(n-i, m-j), \qquad (4.1)$$

where,

I = image being convolved,

(i,j) = image coordinates

and

(n,m) = Gaussian kernel size.

d = first derivative and d^2 = second derivative.

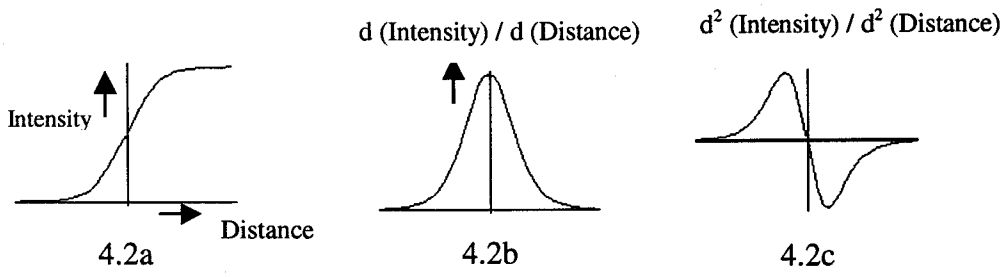

Figure 4.2a. Intensity as a function of distance measured along a line in the image, **4.2b.** First derivative of the intensity function, **4.2c.** Second derivative of the intensitv function

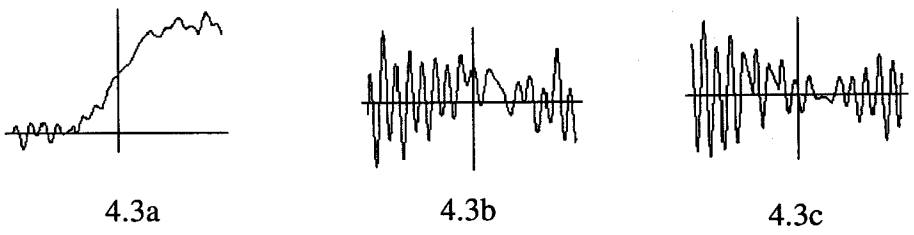

Figure 4.3a. A more realistic example of intensity function (noisy image), **4.3b.** First derivative of the noisy image, **4.3c.** Second derivative of the noisy image

4.4a 4.4b

Figure 4.4a. Before filtering, **4.4b.** After applying Marr-Hildreth filter

The function G being convolved with the image is a two-dimensional Gaussian function (Mexican hat function), given by

$$G_{\sigma}(x,y) = \sigma^{2} e^{-(x^2+y^2)/\sigma^2} \tag{4.2}$$

where σ = standard deviation of the kernel which is the radius (in pixels).

After convolution, the Laplacian operator is applied to detect the zero-crossings and this operator is expressed by,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{4.3}$$

A simple symmetrical Laplacian operator derived from equation (4.3) is given by [35],

$$\begin{bmatrix} & -1 & \\ -1 & +4 & -1 \\ & -1 & \end{bmatrix}$$

This subtracts the grayscale values of each of the neighboring pixels from the central pixel. In a region of the image that has uniform grayscale value, the result of applying this kernel produces zero grayscale value. If there is a discontinuity present in the neighborhood in the form of a point, line, or edge, the result of applying the operator produces a non-zero value. When the value changes from negative to positive, or from positive to negative this effectively means a zero-crossing has occurred. This zero-crossing point is marked as an edge and given a constant numerical value.

In this way the kernel is passed over the entire image and the zero-crossings are identified. The edges are identified, marked and the resulting left and right images are stored for further processing.

4.1.2. Shen-Caston edge detector

Shen and Caston [34] suggested an optimal smoothing filter for an edge detector. For the optimal filter, they developed an *infinite symmetric exponential filter* (ISEF). In their edge detection method the image is convolved with a 1-D filter first in the x direction and then in the y direction. They identified this filter as a *recursive filter*. The use of recursive filtering speeds up the convolution greatly. After filtering, the edges are located by finding the zero crossings of the Laplacian, a process similar to that undertaken in the Marr-Hildreth algorithm.

4.2. Relaxation technique

The corresponding points from the edges can be determined by applying a classical relaxation technique. The relaxation technique uses the features in the images to identify the corresponding points in the stereo images. In this work, edges were identified as the features to be used in the relaxation technique. Several techniques are available for use and a relaxation technique based on cooperative computation [10] has been quite effectively used for random dot stereograms. This technique is rather unsatisfactory when applied to real images. Another useful technique called the PMF algorithm has also been used [11]. The PMF algorithm is primarily based upon a constraint called the *disparity gradient* [35]. It also propagates other constraints such as *continuity, uniqueness and non-reversal ordering constraints* which allows for the candidate matches to reorganize themselves. The similar form of algorithm used in our work is described in the following sections.

4.2.1. Disparity gradient

The disparity gradient is given by the ratio of the disparity difference to the cyclopean separation (based on the difference between the average distances of x and y coordinates in the left and right images). Figure 4.5 depicts the basics of the disparity gradient. Given two points $P_1(X, Y, Z)$ and $P_2(X, Y, Z)$ their disparity gradient DG (equation (4.4)) is

$$DG = \frac{\left| \left(x_2^{lt} - x_1^{lt} \right) - \left(x_2^{rt} - x_1^{rt} \right) \right|}{\left\| \left(\mathbf{p}_2^{c} - \mathbf{p}_1^{c} \right) \right\|} \tag{4.4}$$

Here, $p_1^{lt}(x_1^{lt},y_1^{lt})$ and $p_1^{rt}(x_1^{rt},y_1^{rt})$ are the projections of the point $P_1(X, Y, Z)$ on the left and right image and $p_2^{lt}(x_2^{lt},y_2^{lt})$ and $p_2^{rt}(x_2^{rt},y_2^{rt})$ are the projections of the point $P_2(X, Y, Z)$ on the left and right image respectively. Furthermore, $\mathbf{p_1}^c = (p_1^{lt} + p_2^{rt})/2$ and $\mathbf{p_2}^c = (p_2^{lt} + p_2^{rt})/2$ and, f and b are the focal length and the distance between the two cameras respectively. In the case of a simple camera geometry $y_1^{lt} = y_1^{rt} = y_2^{lt} = y_2^{rt}$.

To fuse the points in the left and right images, it has been shown (based on human vision) that the disparity gradient should not exceed a value of one [12]. The disparity gradient can be used to reveal various stereo-matching constraints and a brief summary is as follows,

$DG > 2$	–	violation of non-reversal ordering constraint
$DG < 1.1$	–	disparity-gradient limit
$DG \ll 1$	–	figural continuity constraint

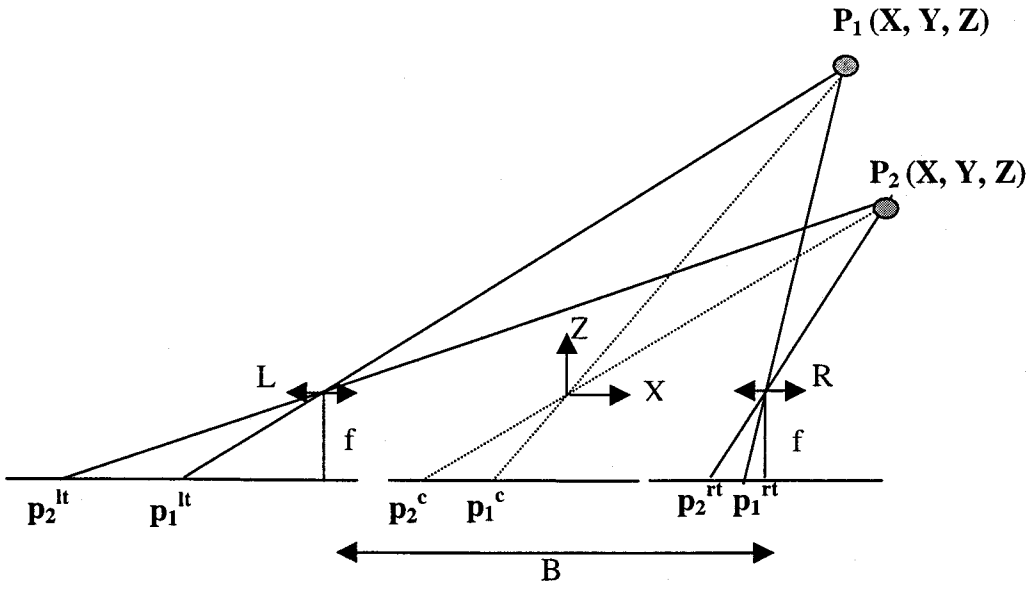

Figure 4.5. Disparity gradient definition

Ordering constraint: Consider a 3-D point **P** and its projections p_1 and p_2 in retinas 1 and 2 respectively as shown in Figure 4.6. Also consider a point **Q** and its projections q_1 and q_2 in retinas 1 and 2 respectively. Q is chosen in such a way that it lies in the cone defined by **P**, C_1, C_2 (cross-hatched in Figure 4.6) containing the base line $C_1 - C_2$. The order of the images on the epipolar lines $\mathbf{E_1}$ and $\mathbf{E_2}$ is represented by ($\mathbf{E_1}$, p_1, q_1) for retina 1 and ($\mathbf{E_2}$, p_2, q_2) for retina 2. This implies that the images of **P** and **Q** appear in the same order along the epipolar lines when Q is in the other cone. This cross-hatched zone in Figure 4.6 is known as the forbidden zone attached to **P** [36].

Consider a case when the violation of ordering constraint occurs. If the point Q bulges out of the forbidden zone then the images along the epipolar lines $\mathbf{E_1}$ and $\mathbf{E_2}$ is

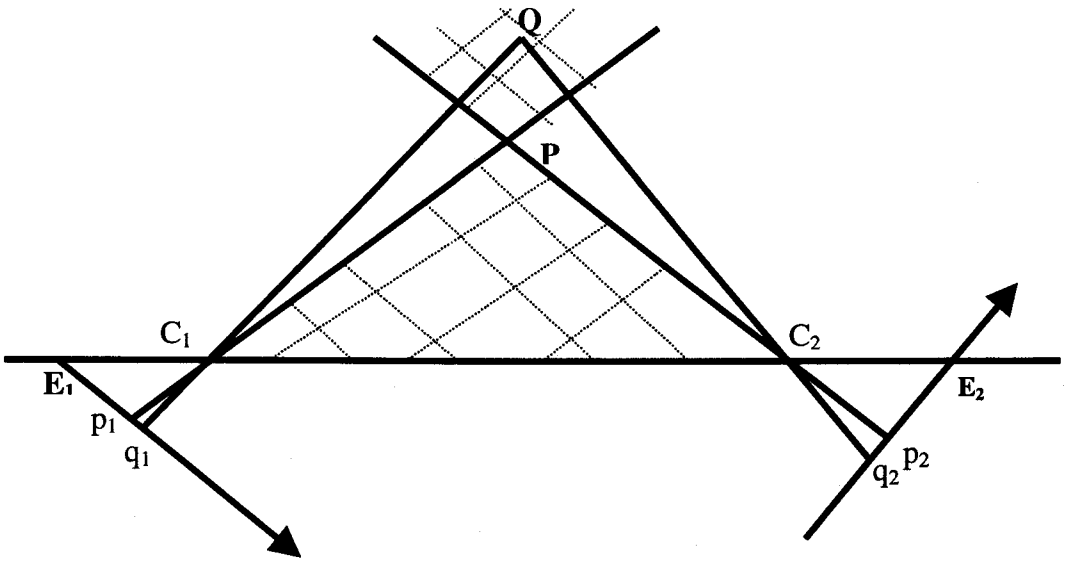

Figure 4.6. Forbidden zone attached to P

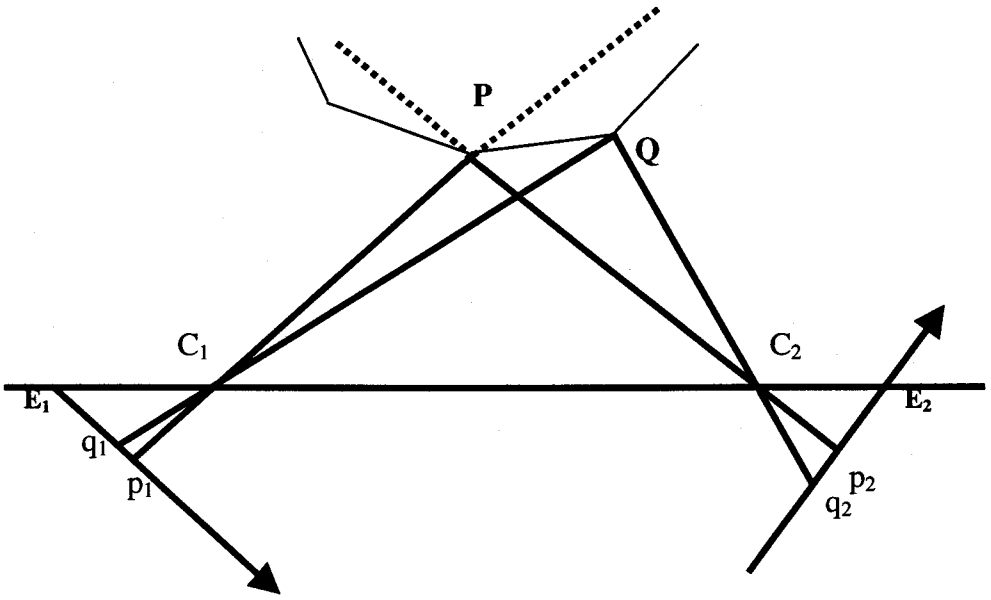

Figure 4.7. The object bulges out of the forbidden zone, and the ordering constraint applies

represented by ($\mathbf{E_1}$, q_1, p_1) for retina 1 and ($\mathbf{E_2}$, p_2, q_2) for retina 2 (Figure 4.7). The value DG > 2 essentially eliminates any order-reversal that occurs.

Uniqueness constraint: The uniqueness constraint states that each image point will participate in only one match. This constraint is based on the assumption that an item corresponds to something that has a unique physical position.

Figural-continuity constraint: The figural constraint is based on the idea that the world is mostly made up of objects with smooth surfaces. For a pair of matched pixels for a 3-D point the reconstruction function is smooth almost everywhere. The reconstruction function is a function of disparity and the disparity varies smoothly almost everywhere.

4.2.2. Algorithm

After applying the Marr-Hildreth [33] edge detector, relaxation technique is used to find the corresponding points in the left and right image. In this thesis, we find the corr-

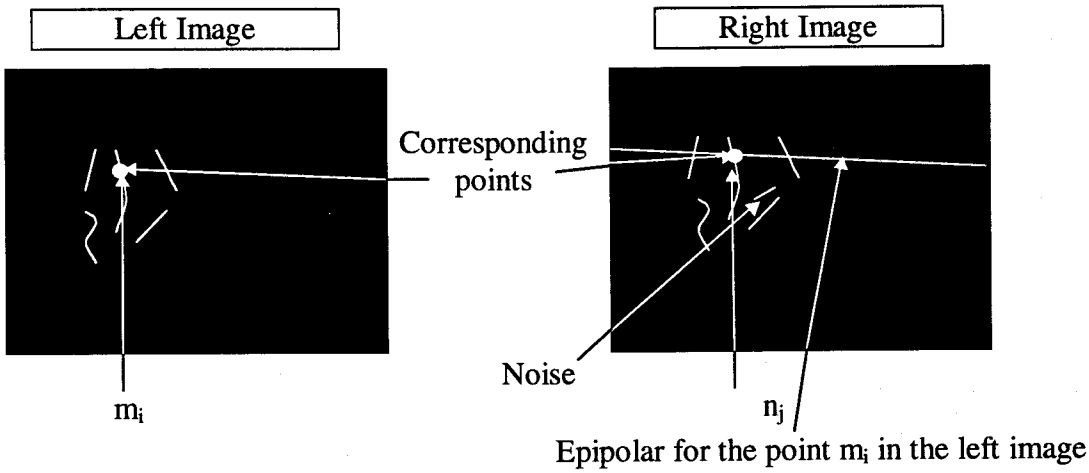

Figure 4.8. Determination of matching strength in presence of noise

-sponding points by matching the zero crossings of the left and right image. The algorithm developed is similar in form to the PMF algorithm [11, 37]. For each token(feature) m_i located at the pixel in the left image, the best possible match n_j in the right image is computed (Figure 4.8). All tokens t_k located at pixels m_k within some neighborhood (V_i) of m_i are considered. For each such token the matches t_l located at pixels n_ℓ in the right image satisfying the constraint:

$$DG(m_i, n_j, m_k, n_\ell) < 1 \tag{4.5}$$

are considered as candidates for supporting the match (m_i, n_j). In the neighborhood V_i, the strength of match S is defined as:

$$S = \sum_{t_k \text{ in } V_i} \frac{1}{dist(m_i, m_k)} \max_{\text{all } t_i} \frac{1}{dist(n_j, n_\ell)} \partial[DG(m_i, n_j, m_k, n_\ell)] \tag{4.6}$$

where,

$$\partial[DG(m_i, n_j, m_k, n_\ell)] = 1 \text{ if } DG(m_i, n_j, m_k, n_\ell) < 1$$
$$= 0 \text{ otherwise} \tag{4.7}$$

where,

dist = distance and

∂ = disparity gradient value.

Here, the distance measure (dist (m_i, m_k)) can be any distance measure (Euclidean distance, City Block distance, etc.,) or pixel strength, orientation etc., Tokens are then matched as a winner-take-all procedure. This definition is non-symmetric, and a similar expression can be computed from the right image to the left image. At each iteration, the matches for which the strength is a maximum is chosen as the correct match. Following the uniqueness constraint, all the other matches associated with the two tokens in each chosen match are eliminated from further consideration. This allows further matches to be selected as correct provided that they now have the highest

strengths for both tokens. Convergence is usually achieved within a few iterations. For noisy images, more iterations are needed. In this algorithm, only vertical edges are considered. The reason being the relaxation is based on the disparity gradient limit which is obtained from the shift between the vertical edges in the stereo images. The matching strength is computed for all the zero-crossing points along the epipolar line for a particular match of interest.

For a point in the left image, the corresponding point in the right image lies only within a certain window size of the point in the left image. This condition is applied and a window size of one-eighth of the window size of the image is normally selected after experimentally trying with several other window sizes. The search space, in effect, applying the epipolar geometry constraint has been greatly reduced to a short segment. The matching strength is computed from the candidate matches found in the neighborhood. A disparity-gradient limit of less than 1 is used in the matching process. Normally a value of 0.3 - 0.6 proves to work well. By choosing this value, all the constraints are satisfied.

The presence of noise (Figure 4.8) will give a low matching strength and some times no matching strength. This can be counteracted by increasing the number of iterations.

4.3. Correlation technique

In this technique, a target pattern or template is shifted over each point in the image. The pixel values of the source are multiplied by the pixel values of the target that are overlaid to get a particular value called the *correlation coefficient.* The pixel which shows the highest value is then selected as the best match. Since this technique needs more computational time than the relaxation technique, this technique is used only to eliminate the false matches computed from the relaxation technique.

4.3.1. Image parameters

The images can be analyzed in two formats, *grayscale images* or *color* images. The pixel values for grayscale images are given by, zero for Black and 255 for White. All the other pixel values in the grayscale images lies between zero and 255. The values for color images are represented by three components namely Red, Green, and Blue, and their values are given in Table 3.1.

Over the years, there have been approaches to solve the problem of correspondence imaging using correlation and most of the research has been done only on grayscale images. The reason for the color being neglected is that three channels namely RGB (Red, Green, Blue) have to be taken into consideration when compared to gray scale images where only one channel is taken into consideration. RGB components in the form of cube are shown in Figure 4.9. Much research has been done toward the development of color measurement techniques, but there is no one color coordinate system that is universally accepted. There are several color coordinate systems, each having its particular merits. The color model RGB is used in hardware applications like PC monitors, cameras and scanners. The CMY (Cyan, Magenta and Yellow) color model is used in color printers, and the YIQ model is used in television broadcast.

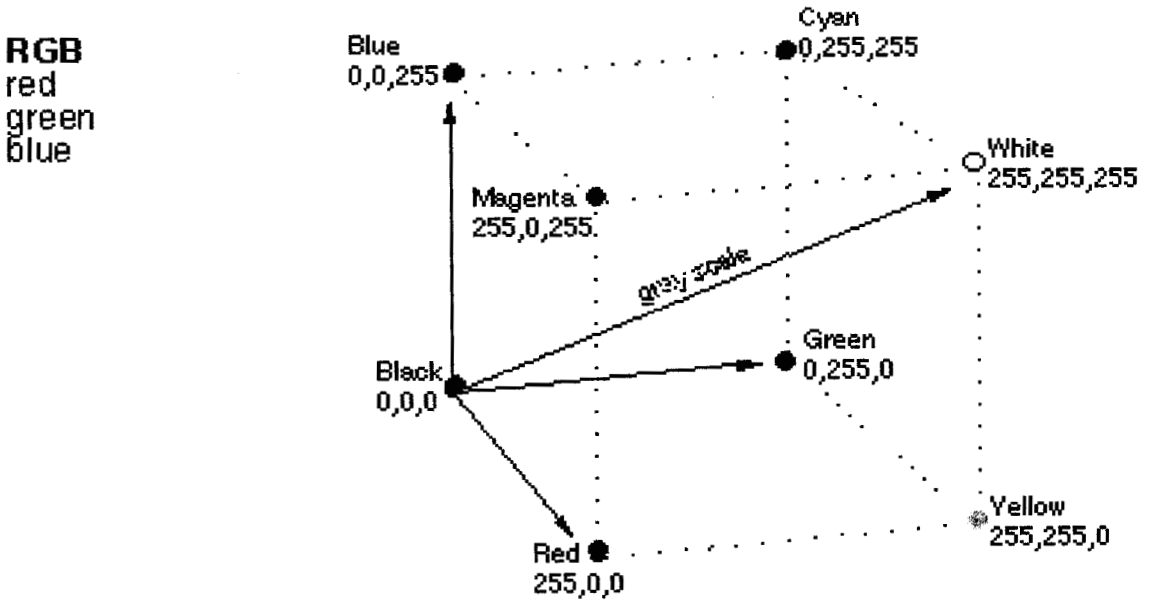

Figure 4.9. RGB cube

The YIQ system is the color primary system adopted by National Television System Committee (NTSC) for color TV broadcasting. NTSC luminance Y, contains the gray scale information, and chrominance (I and Q) contains the hue and saturation

information. In color image manipulation, a model widely used is HSI (Hue, Saturation, Intensity or Value). HSI color space is represented in Figure 4.10.

In this model, hue is the color as described by wavelength, for instance the distinction between red and yellow. Saturation is the amount of the color that is present and aids for instance to distinguish between red and pink. The intensity is the amount of light, the distinction between a dark red and light red or between dark and light gray. HSI is supposed to be more intuitive and to follow the actual color sensitivity of the human eye [38].

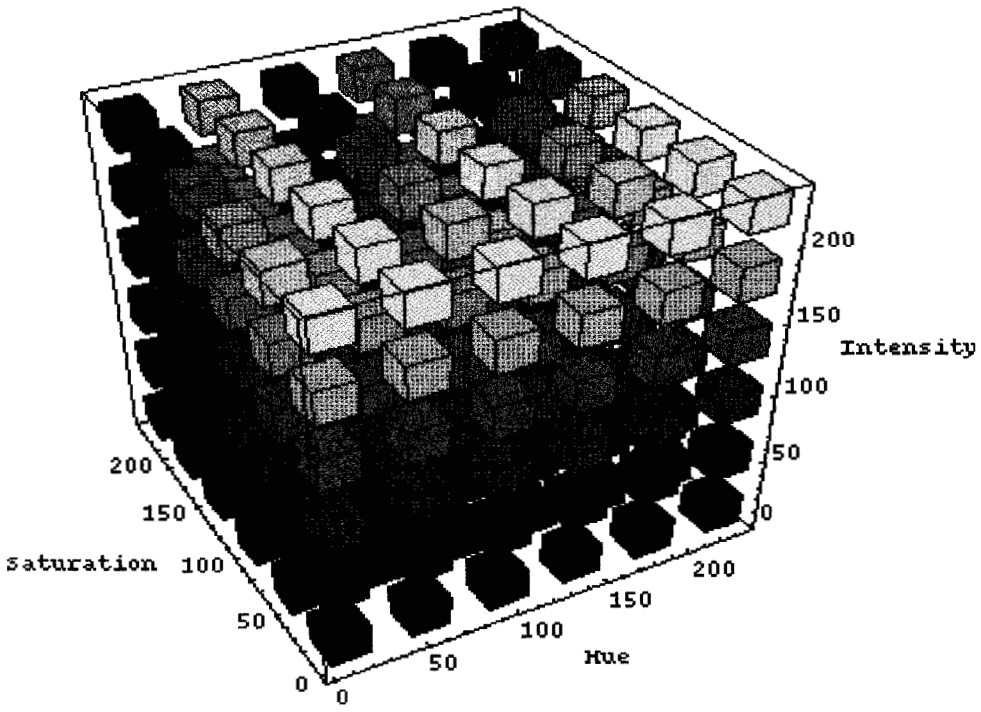

Figure 4.10. HSI color space

The conversion of RGB space into HSI space is given by [39],

$$H = \arctan\left[\frac{-1/3.R + 2/3.G - 1/3.B}{2/3.R - 1/3.G - 1/3.B}\right] \tag{4.8}$$

$$S = \sqrt{\left((2/3.R - 1/3.G - 1/3.B)^2 + (-1/3.R + 2/3.G - 1/3.B)^2\right)} \tag{4.9}$$

$$I = \frac{R + G + B}{3} \tag{4.10}$$

The present study aims at also using the color values for template matching instead of grayscale values. The values considered are HSI because that closely reflects the perception of color.

4.3.2. Correlation coefficient

The correlation coefficient r is defined as:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2\right]^{1/2}} \tag{4.11}$$

Here, x_i = (H or S or I value) in the left image, y_i = (H or S or I value) in the right image. A value of r = +1 represents an absolute positive correlation, while r = 0 and r = -1 represent no correlation and a negative correlation respectively. The false matches are detected and eliminated based on the correlation coefficient from the left and right image HSI values. A 3 x 3 matrix is constructed around the corresponding points in the left and right image as shown in Figure 4.11. A value of 0.85 is set as the correlation

coefficient value. If the value found is greater than or equal to 0.85 then it is marked as a correct match else it is disregarded as a false match. Typically each color image helps to compute three sets of results.

The dark outlined pixels represent the corresponding points in the left and right image. The pixels P1 – P8 represent the neighboring pixels surrounding the pixels of interest in the two images.

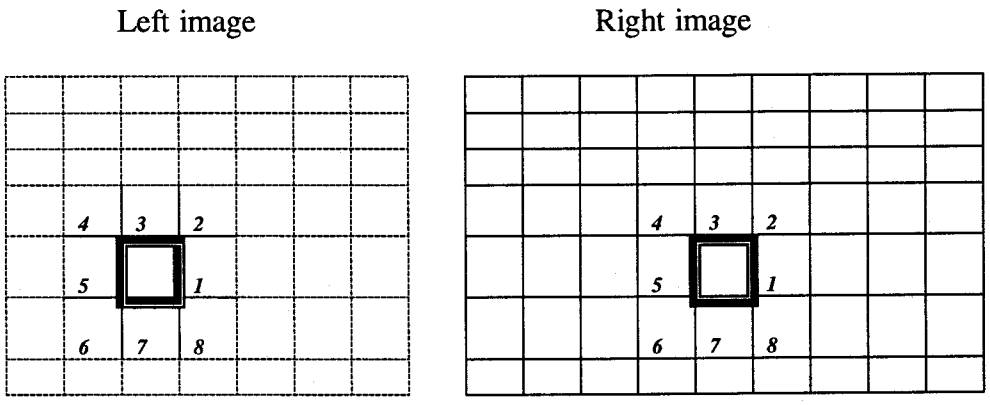

Figure 4.11. Correlation technique

4.4. Depth determination

Determining the depth from left and right images taken from different views is similar in principle to using stereo pair images taken from aircraft or satellites to measure the elevation of topographic features on the earth's surface [38]. The measurement geometry to determine the depth depends on the method used to determine the depth. The depth relationship between the measured parallax (the apparent displacement of points in the left and right image) and the relative elevation of the two points on the surface is as shown Figure 4.12. The height h [40] is given by

$$h = B.\frac{d_1 - d_2}{K} \qquad (4.12)$$

where,

h = height of the object from the surface,

B = distance between the ground and the view points,

d_1 = relative distance of the objects in the right view of the image,

d_2 = relative distance of the objects in the left view of the image,

and,

K= distance between the view points.

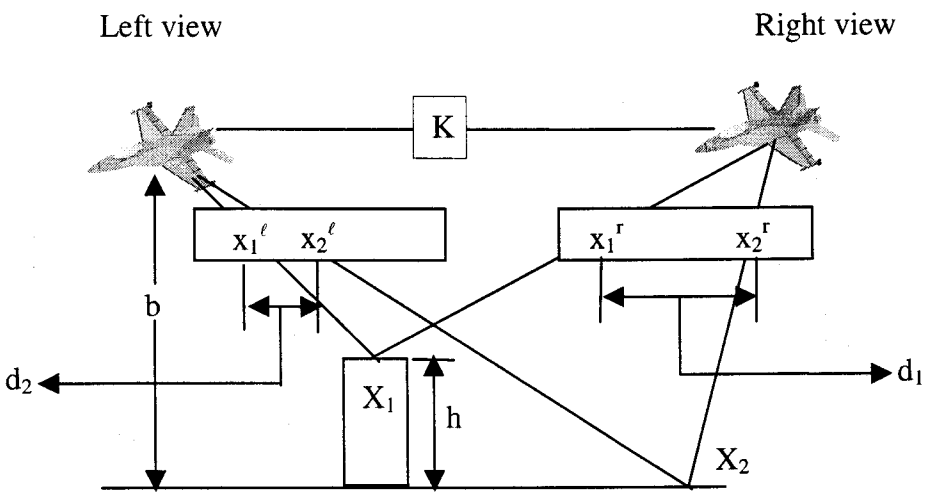

Figure 4.12. Depth determination using parallax

5. RESULTS AND DISCUSSION

The stereoscopy algorithms developed were tested on both gray scale and color images. For color images the HSI values were used. The essential matrix used to find the epipolar line was computed using the eight-point algorithm and the software developed by Zhang *et al.* [23]. Using the essential matrix the epipolar lines for the point of interest was found. The points of interest considered were the zero-crossings which were established using two types of edge detectors: Marr-Hildreth and Shen-Caston edge detectors. The choice of edge detector was based on the quality of the images.

The stereo algorithm software was developed using Visual Basic 5.0©. The area of interest where depth to be determined can be either manually or automatically selected. To this effect two software programs were developed. One was set-up for manual selection of points of interest and the other was set-up for automatic selection. The format of image files considered are .pgm, .jpg and .tif files. After finding edges, the relaxation technique was used to find the best match. To eliminate any false matches detected from the relaxation technique, the correlation technique was used. The values considered in correlation technique were HSI values detected from the color images. A template of size 7 x 7 pixels was used to perform the correlation technique. A correlation coefficient value of between 0.9 and 0.95 (based on experimental results) depending on the image quality was found to be appropriate in most cases. The matchpoints were stored in a separate file for further use.

5.1. Grayscale images

The edge detected, left and right image of the White House images (Figure 3.7) using Marr-Hildreth edge detection is shown in Figure 5.1. The edge detection was carried out using the Marr-Hildreth edge detector. Since the images have good features, the number of edges detected is quite large. The results of applying the relaxation and correlation

techniques for grayscale images are shown in Figure 5.2. In these images the control points are manually selected and the corresponding points are found along the epipolar line. A correlation coefficient value of 0.95 was used. By visual inspection, one can observe a single mismatch (the point number 15). Otherwise, the corresponding points were found by applying the techniques are good.

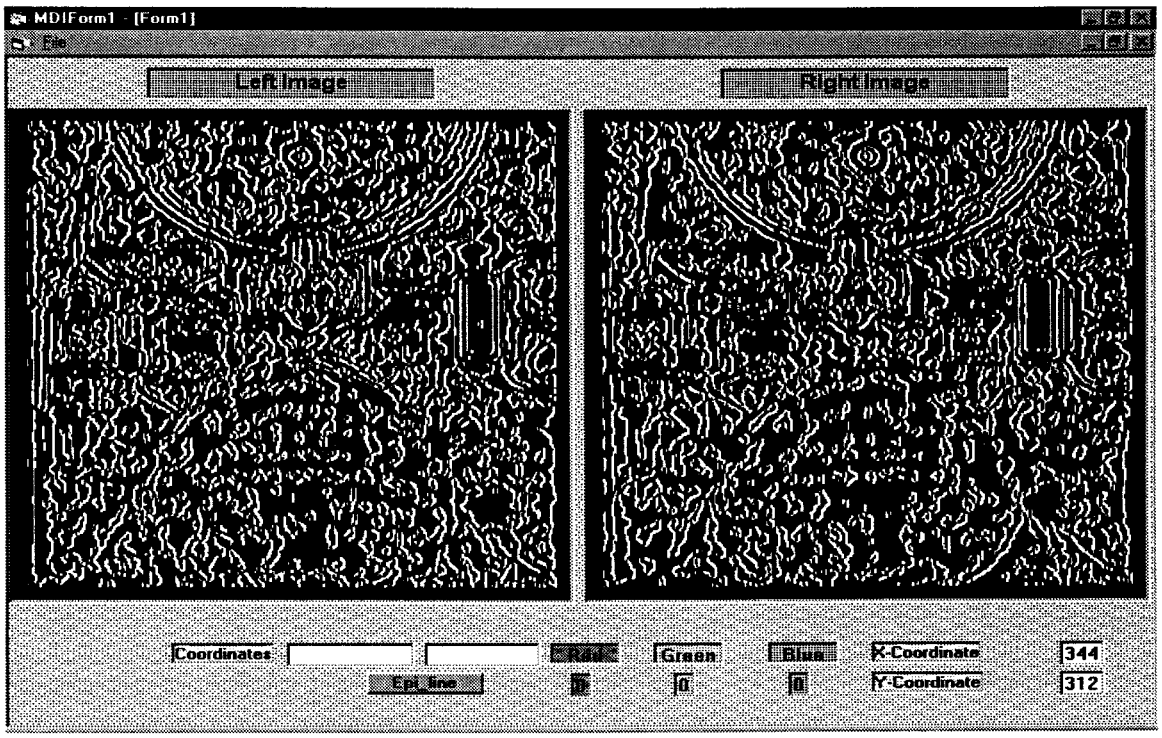

Figure 5.1. Edge detected image of left and right view of White House, Washington D.C., U.S.A.

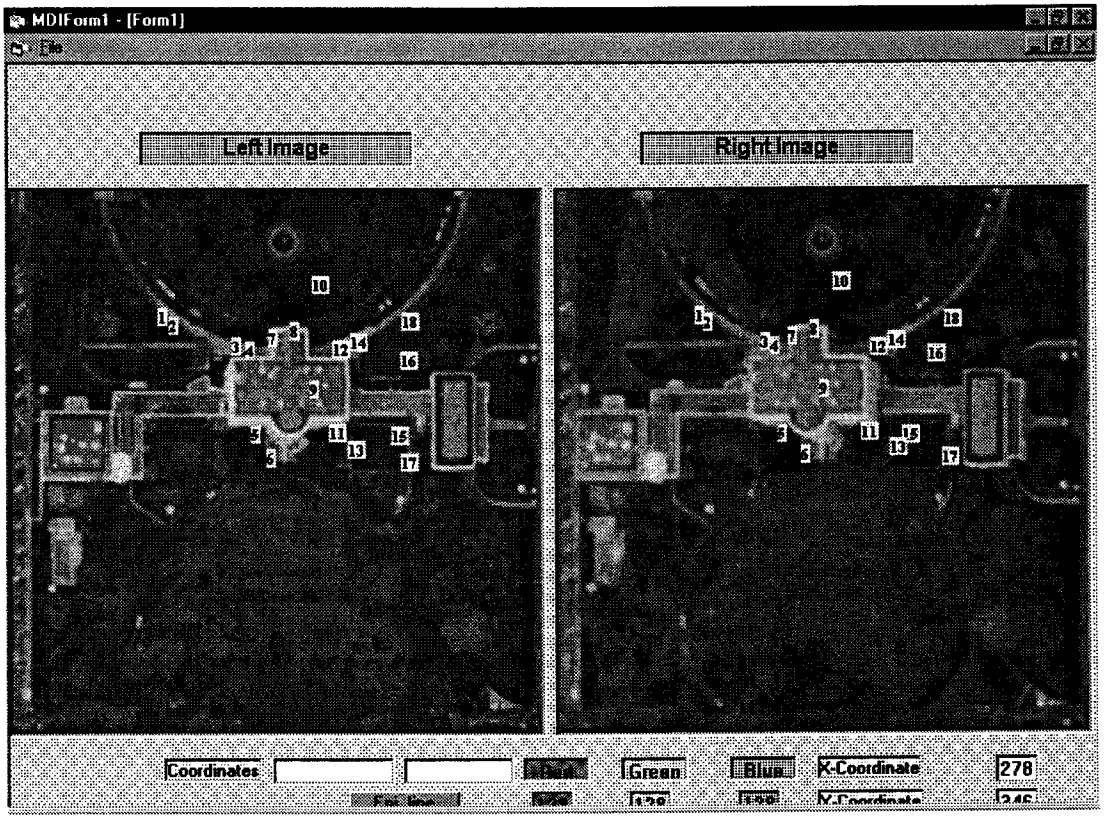

Figure 5.2. Corresponding points on the stereo images

The stereo images in Figures 5.3a and 5.3b were taken on the surface of Mars (http://www.iki.rssi.ru/jplmirror/mars/ops/sol71.html). This stereo image pair was taken on the afternoon on Sol 17 (September 14) shows Sojourner leaving the "Rock Garden," an assemblage of large rocks behind and to the right of the rover. Sojourner's last target was the rock "Stimpy," seen at the far right. For these images the Shen-Caston edge detector was used (Figures 5.3c and 5.3d). In these images, the bottom half of the image was considered as the area of interest and the corresponding points were determined. A correlation coefficient value of 0.9 was used. No mismatches were identified for these stereo images Figures 5.3e and 5.3f.

Figure 5.3a

Figure 5.3b

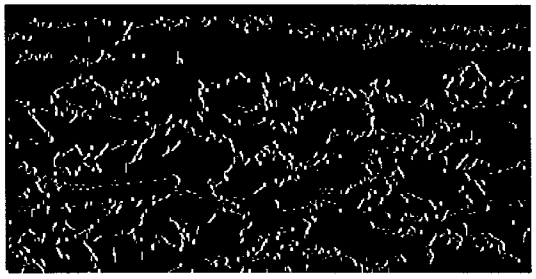

Figure 5.3c

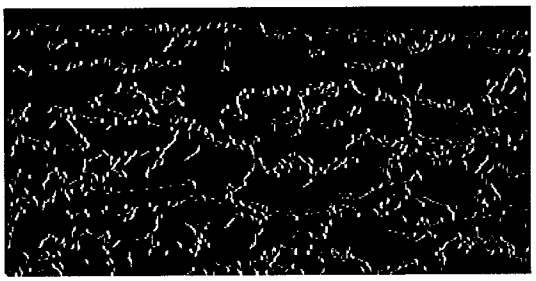

Figure 5.3d

Figures 5.3a. and 5.3b. Stereo images taken on the surface of the Mars,
Figures 5.3c. and 5.3d. Edge detected images

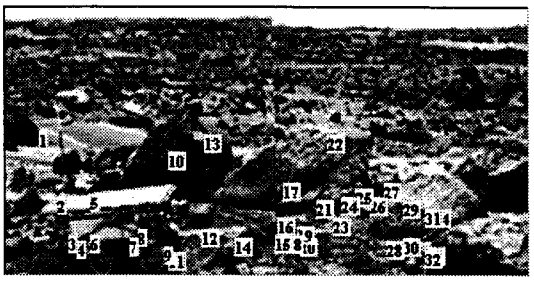

Figure 5.3e

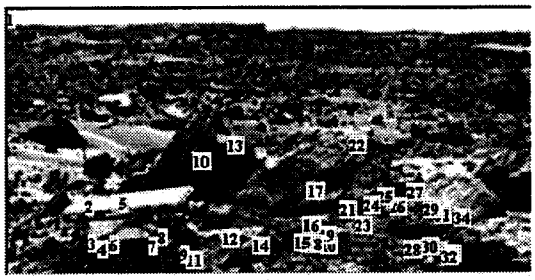

Figure 5.3f

Figures 5.3e. and 5.3f. Corresponding points on the stereo images

5.2. Color images

5.2.1. Case 1

The algorithm was tested on several color images. As indicated earlier, the parameters considered here are HSI. For example, Figures 5.4a. and 5.4b. show a pair of stereo images of clouds and mountains (http://www.cs.cmu.edu/afs/cs/project/ sensor_9/ftp/ images /stills/stereo/). The edge detected images are shown in Figures 5.4c and 5.4d. Automatic selection of the points of interest based on zero-crossings was used for these images. The Hue detected images and corresponding points are shown in Figures 5.5a and 5.5b, 5.5c and 5.5d respectively. The search was restricted along the epipolar line.

The matches found using Hue as the value in the correlation technique was found to be good, but only a few matches were obtained. The points of interest like the peak of the mountain and points along the mountain were not determined in this technique. But when the Saturation (Figures 5.6a and 5.6b) and Intensity (Figures 5.7a and 5.7b) values were used, more matches were found. In all these images, none of the matches were found inside the cloud and this can be easily attributed to the lack of edges found. There

were no matches found on the trees and this is probably because of the varying HSI values in the left and right images. This was one of the shortcoming of using the correlation technique to eliminate false matches. This is because of the wide variation in HSI values near the trees. The relaxation technique may have established several matches on the trees and on applying the correlation technique, there would have been rejection of these matches. In general the matches found for these types of images were very good (Figures 5.6c and 5.6d, and Figures 5.7c and 5.7d).

Figure 5.4a

Figure 5.4b

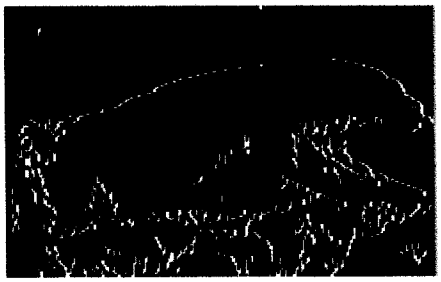

Figure 5.4c

Figure 5.4d

Figures 5.4a. and 5.4b. Stereo images of the mountain and clouds, **Figures 5.4c. and 5.4c.** Edge detected images

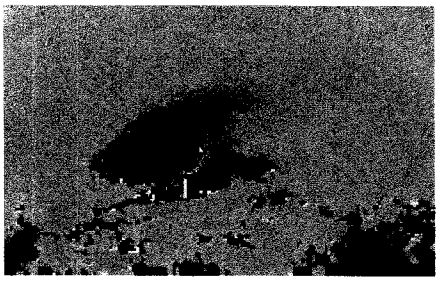

Figure 5.5a

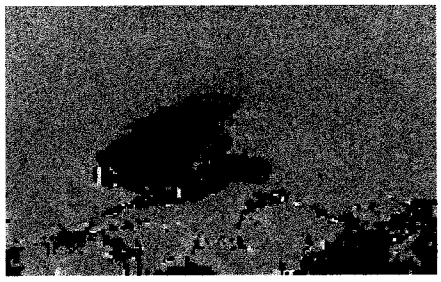

Figure 5.5b

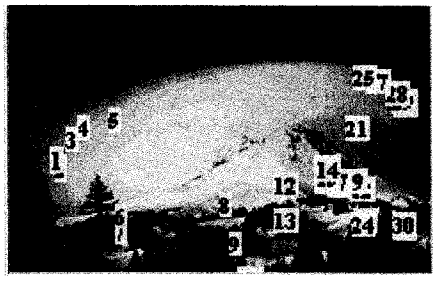

Figure 5.5c

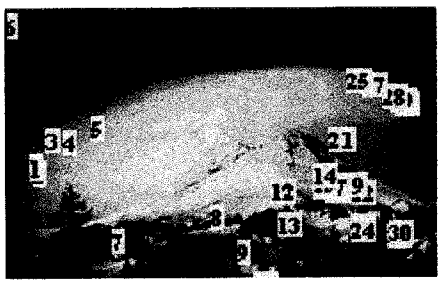

Figure 5.5d

Figures 5.5a. and 5.5b. Hue detected images,

Figures 5.5c. and 5.5d. Corresponding points on the stereo images

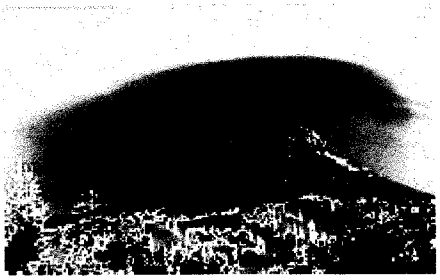

Figure 5.6a

Figure 5.6b

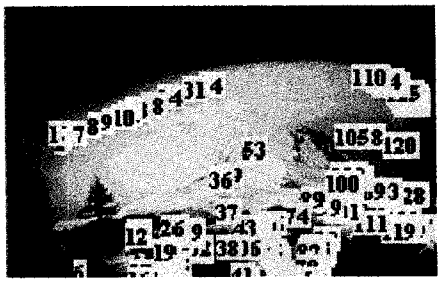

Figure 5.6c

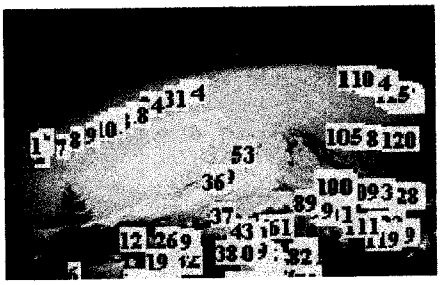

Figure 5.6d

Figures 5.6a. and 5.6b. Saturation detected images,

Figures 5.6c. and 5.6d. Corresponding points on the stereo images

Figure 5.7a

Figure 5.7b

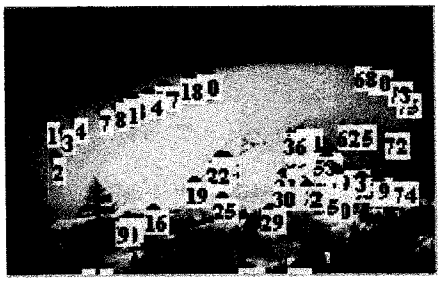

Figure 5.7c

Figure 5.7d

Figures 5.7a. and 5.6b. Intensity detected images,

Figures 5.7c. and 5.7d. Corresponding points on the stereo images

5.2.2. Case 2

The correspondence imaging technique is applied to the stereo images of a human image, Figures 5.8a and 5.8b (http://www.cs.cmu.edu/afs/cs/project/ sensor_9/ftp/ images /stills/stereo/). The edge detected images are shown in Figures 5.8c and 5.8d. These images appear to be perfect without much deformation when one views the face.

When the relaxation and correlation techniques were applied to the Hue value (Figures 5.9a and 5.9b) detected images there were only few a matches and the points 1 and 3 were not correlated (Figures 5.9c and 5.9d). Using the Saturation value (Figures 5.10a and 5.10b) point 2 is totally a mismatch and for points 5 and 6, either point 5 or 6 is a false match. Points 1 and 3 may or may not be false matches (Figures 5.10c and 5.10d). With Intensity values (Figures 5.11a and 5.11b) the number of points detected was large and there were some false matches like points 6, 8 and 4 (Figures 5.11c and 5.11d).

The reason for mismatches can be attributed to the epipolar lines. As mentioned earlier, the eight-point algorithm is sensitive to noise and the essential matrix computed for these images was not accurate. This can be explained from the Figure 5.12. These images were subjected to much shift between viewpoints which can be noticed by observing the shift in the branch behind the head. This would have resulted in an inaccurate estimation of the essential matrix. For example for points 2 and 3 in the left image the epipolar lines in the right image passes at the wrong points. However, for points 1 and 4, the epipolar lines pass exactly through the corresponding points. This explains the reason why more corresponding points were found on the face rather than in any other region.

Figure 5.8a

Figure 5.8b

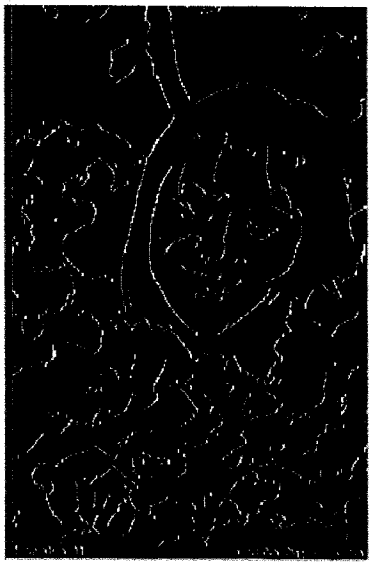

Figure 5.8c

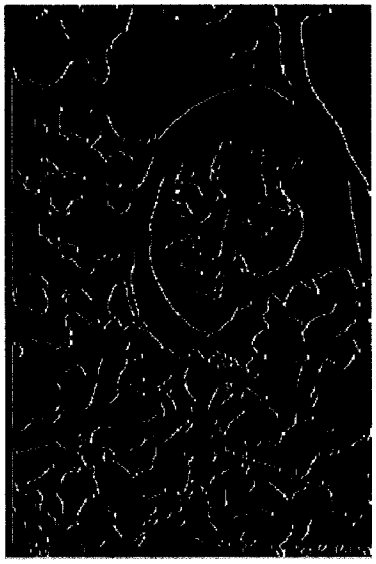

Figure 5.8d

Figures 5.8a. and 5.8b. Stereo images of a human face,

Figures 5.8c. and 5.8d. Edge detected images

Figure 5.9a

Figure 5.9b

Figure 5.9c

Figure 5.9d

Figures 5.9a. and 5.9b. Hue detected images,

Figures 5.9c. and 5.9d. Corresponding points on the stereo images

Figure 5.10a

Figure 5.10b

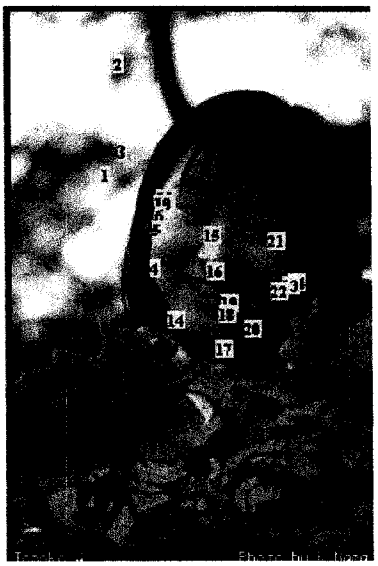

Figure 5.10c

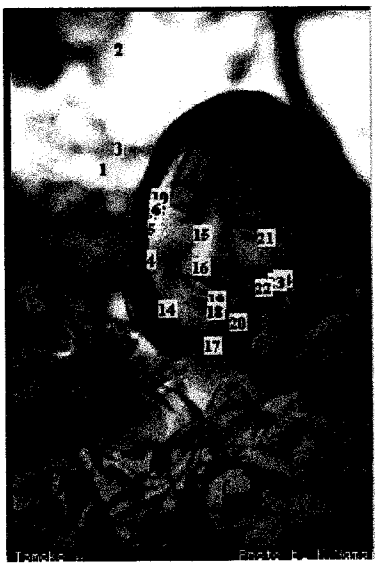

Figure 5.10d

Figures 5.10a. and 5.10b. Saturation detected images,

Figures 5.10c. and 5.10d. Corresponding points on the stereo images

Figure 5.11a

Figure 5.11b

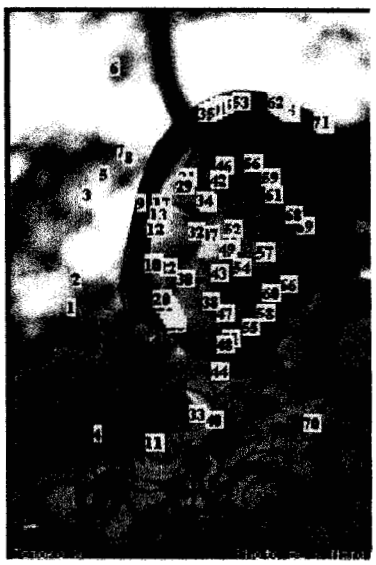

Figure 5.11c

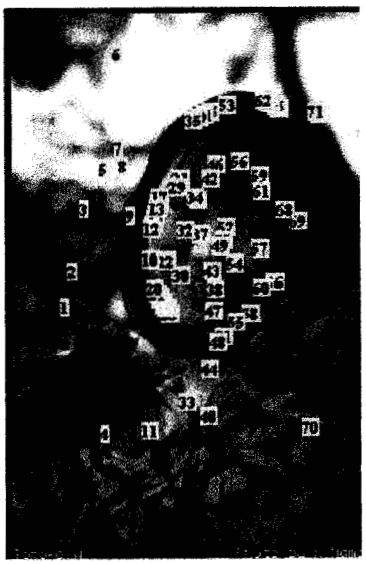

Figure 5.11d

Figures 5.10a. and 5.10b. Intensity detected images,

Figures 5.10c. and 5.10d. Corresponding points on the stereo images

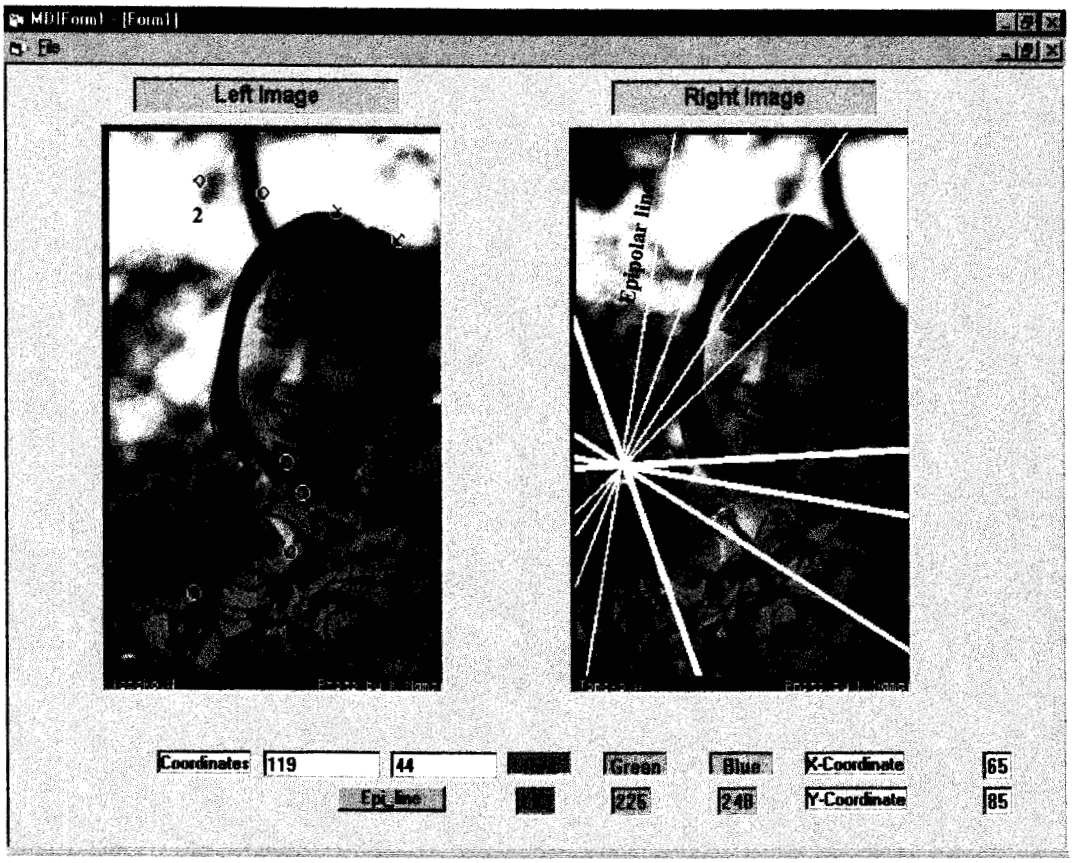

Figure 5.11. Analysis of epipolar lines for few points for the stereo image

5.2.3. Case 3

The stereo images of a shoulder bone are shown in Figures 5.12a and 5.12b (http://www.cs.cmu.edu/afs/cs/project/sensor_9/ftp/images/stills/stereo/). The corresponding edge detected images are shown in Figures 5.12c and 5.12d while Figures 5.12e and 5.12f show the matching points. Since there is not much deformation in these images the number of false matches were few (points 26 and 27). Manual selection of control points was used to find the corresponding points in these images.

Figure 5.12a

Figure 5.12b

Figure 5.12c

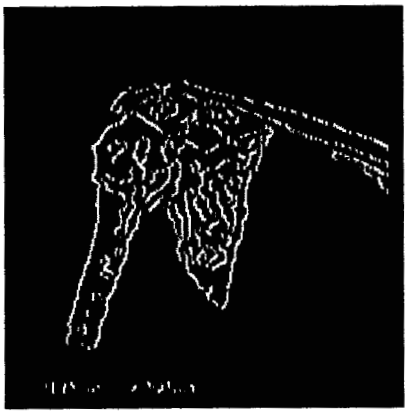

Figure 5.12d

Figures 5.12a. and 5.12b. Stereo images of shoulder bone

Figures 5.12c. and 5.12d. Edge detected stereo images

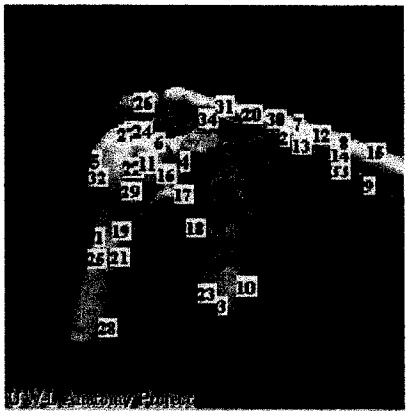

Figure 5.12e

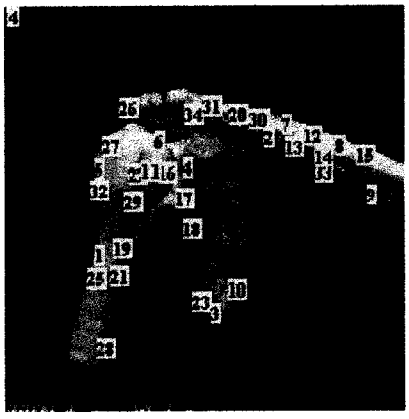

Figure 5.12f

Figures 5.12e. and 5.12f. Corresponding points in the intensity image

5.2.4. Case 4

A set of artificial stereo images are shown in Figures 5.13a and 5.13b (http://www.cs.cmu.edu/afs/cs/project/sensor_9/ftp/images/stills/stereo/). The edges were detected using the Marr-Hildreth edge detector (Figures 5.13c and 5.13d). The results are quite fair with only a few mismatches [points 49 and 52 (Figures 5.13e and 5.13f)]. For these images, only the Saturation value is taken into consideration. The Hue value tends to be zero for these images which may due to not much color difference in the images.

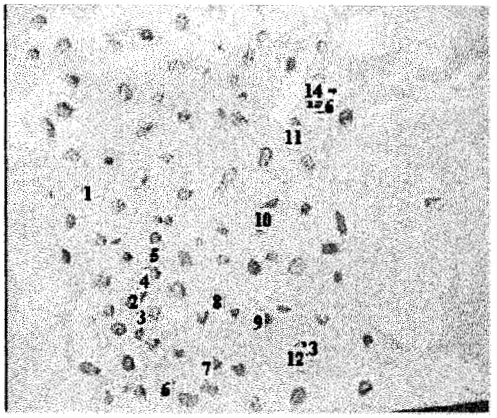

Figure 5.15c

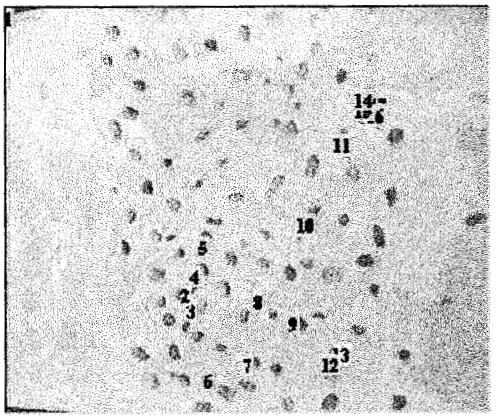

Figure 5.15d

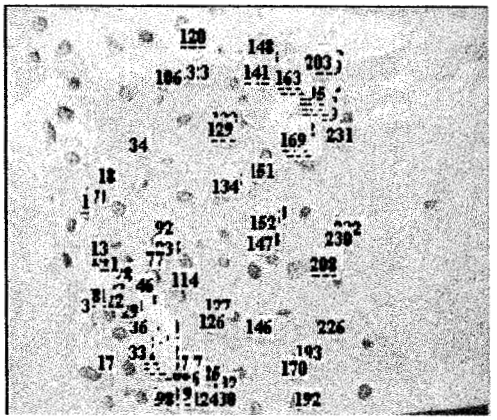

Figure 5.15e

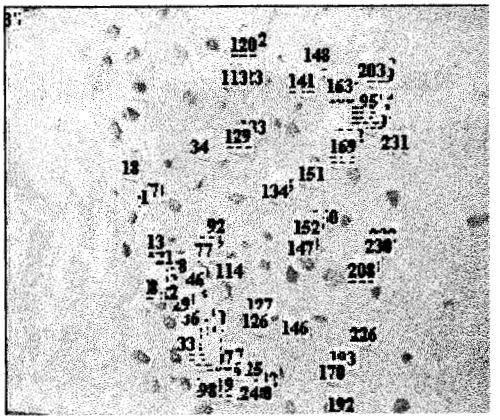

Figure 5.15f

Figures 5.15c. and 5.15d. Corresponding points in the saturation image with a correlation coefficient value of 0.9, **Figure 5.15e. and 5.15f.** Corresponding points in the intensity image with a correlation coefficient value of 0.8

Figure 5.10a

Figure 5.10b

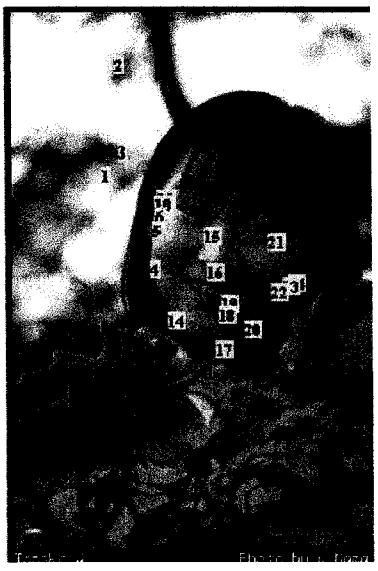

Figure 5.10c

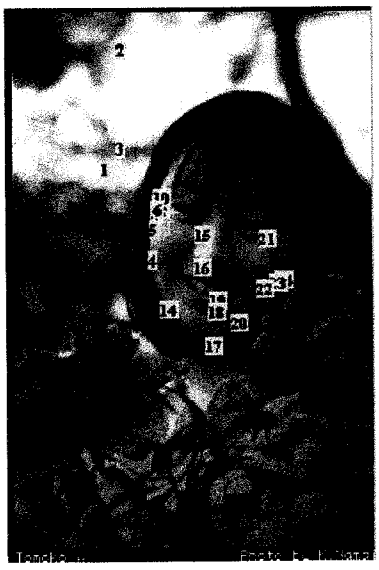

Figure 5.10d

Figures 5.10a. and 5.10b. Saturation detected images,

Figures 5.10c. and 5.10d. Corresponding points on the stereo images

Figure 5.11a

Figure 5.11b

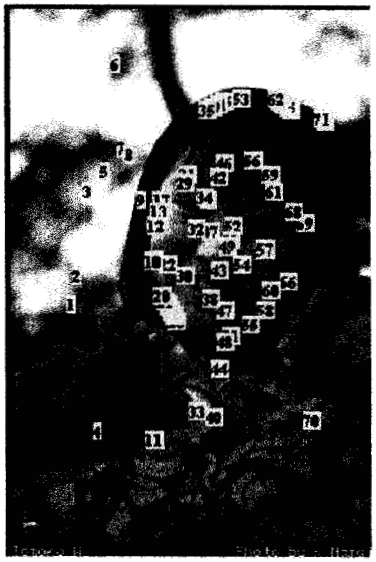

Figure 5.11c

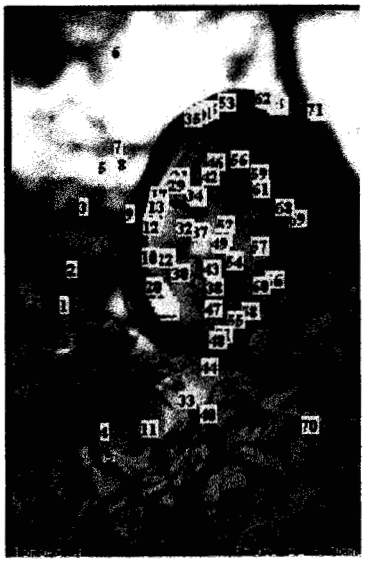

Figure 5.11d

Figures 5.10a. and 5.10b. Intensity detected images,

Figures 5.10c. and 5.10d. Corresponding points on the stereo images

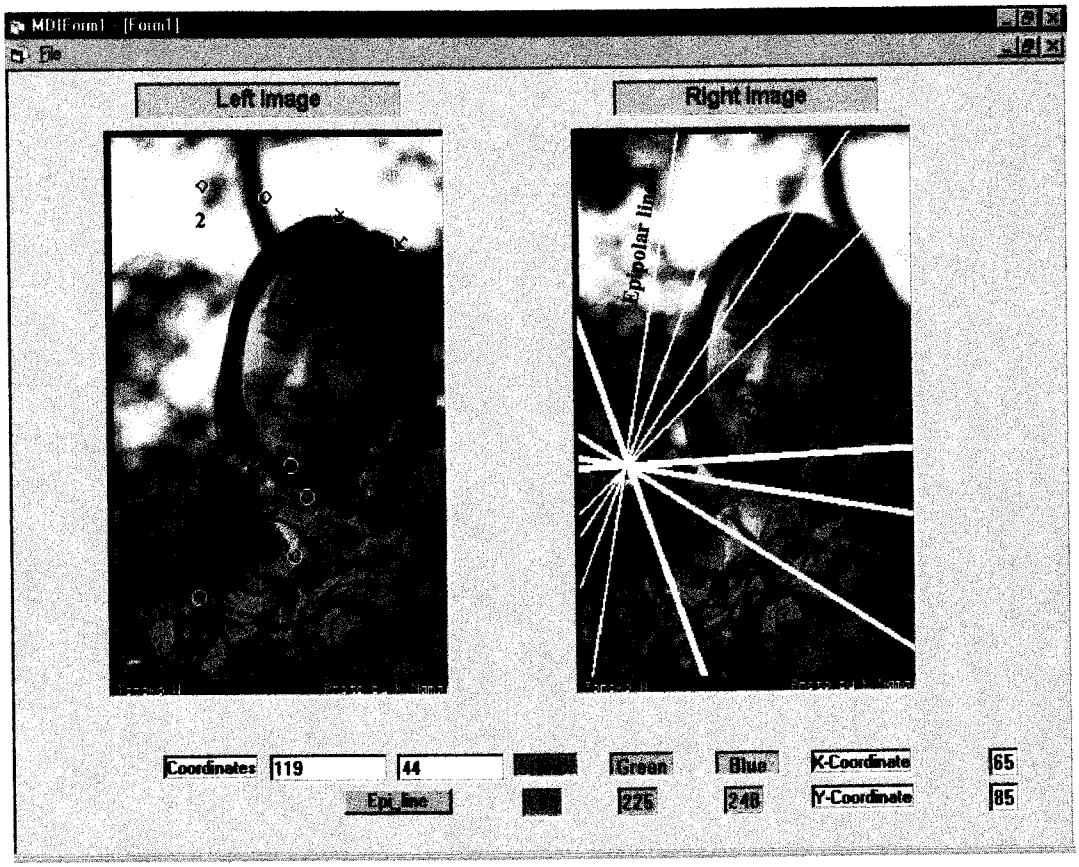

Figure 5.11. Analysis of epipolar lines for few points for the stereo image

5.2.3. Case 3

The stereo images of a shoulder bone are shown in Figures 5.12a and 5.12b (http://www.cs.cmu.edu/afs/cs/project/sensor_9/ftp/images/stills/stereo/). The corresponding edge detected images are shown in Figures 5.12c and 5.12d while Figures 5.12e and 5.12f show the matching points. Since there is not much deformation in these images the number of false matches were few (points 26 and 27). Manual selection of control points was used to find the corresponding points in these images.

Figure 5.12a

Figure 5.12b

Figure 5.12c

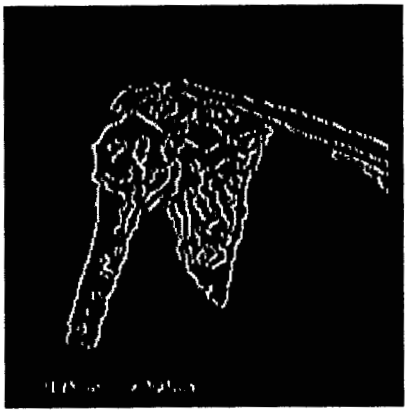

Figure 5.12d

Figures 5.12a. and 5.12b. Stereo images of shoulder bone

Figures 5.12c. and 5.12d. Edge detected stereo images

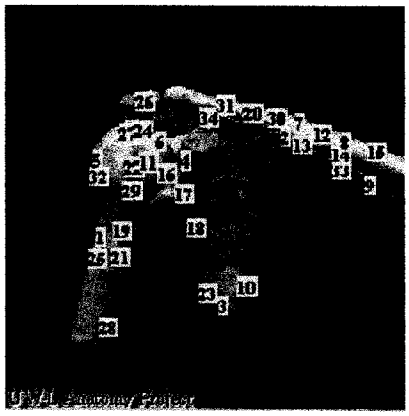

Figure 5.12e

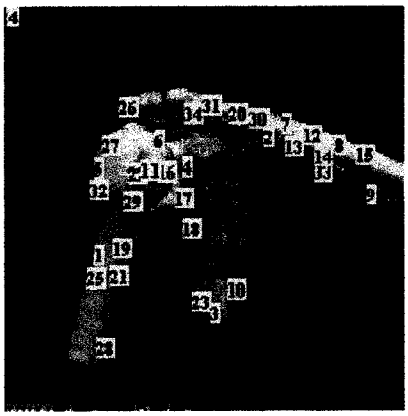

Figure 5.12f

Figures 5.12e. and 5.12f. Corresponding points in the intensity image

5.2.4. Case 4

A set of artificial stereo images are shown in Figures 5.13a and 5.13b (http://www.cs.cmu.edu/afs/cs/project/sensor_9/ftp/images/stills/stereo/). The edges were detected using the Marr-Hildreth edge detector (Figures 5.13c and 5.13d). The results are quite fair with only a few mismatches [points 49 and 52 (Figures 5.13e and 5.13f)]. For these images, only the Saturation value is taken into consideration. The Hue value tends to be zero for these images which may due to not much color difference in the images.

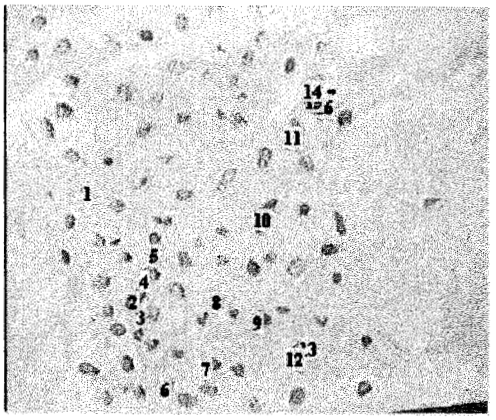

Figure 5.15c

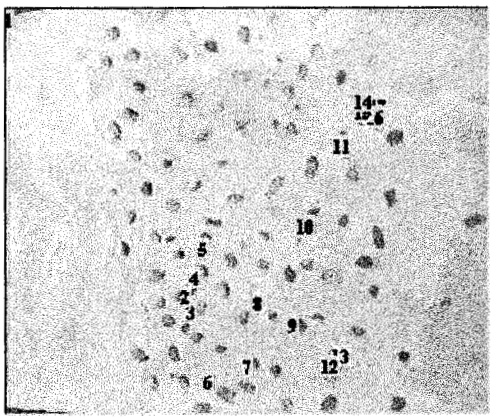

Figure 5.15d

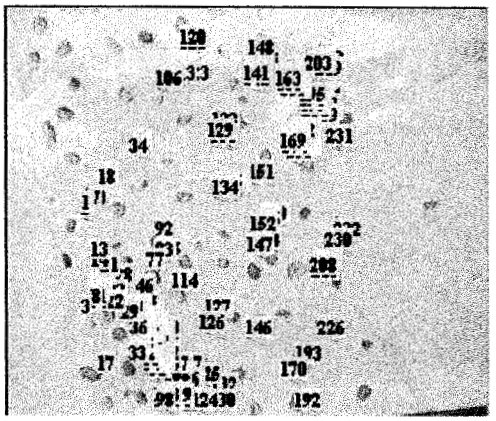

Figure 5.15e

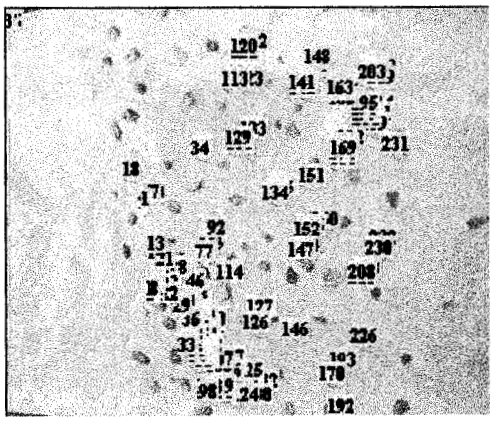

Figure 5.15f

Figures 5.15c. and 5.15d. Corresponding points in the saturation image with a correlation coefficient value of 0.9, **Figure 5.15e. and 5.15f.** Corresponding points in the intensity image with a correlation coefficient value of 0.8

Additional images were acquired with fewer stones (Figures 5.16a and 5.16b). The edge detected images (Figures 5.16c and 5.16d) showed no visible correlation between the two images. The relaxation and correlation techniques were applied on these images with a correlation coefficient value of 0.9. The results found for these types of images (Figures 5.17a and 5.17b, Figures 5.17c and 5.17d) tends to be very poor. The number of matches found were few and with mostly false matches

Figure 5.16a

Figure 5.16b

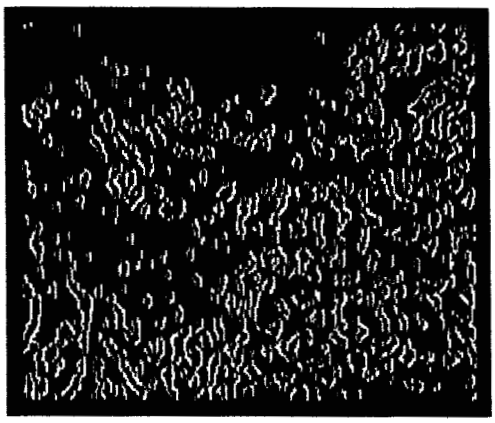

Figure 5.16c)

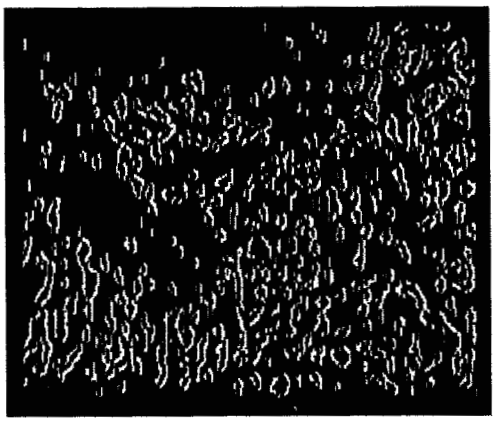

Figure 5.16d

Figures 5.16a. and 5.16b. Stereo images with only very few stones, **Figures 5.16c. and 5.16d.** Edge detected images

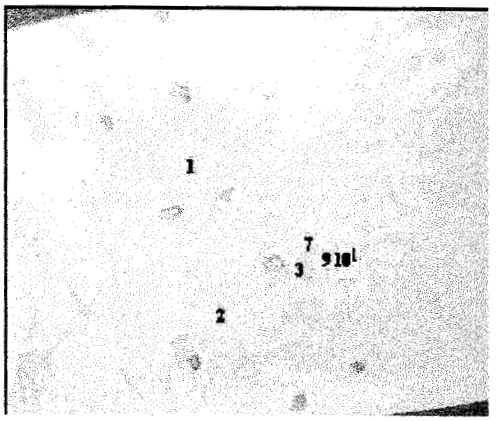

Figure 5.17a

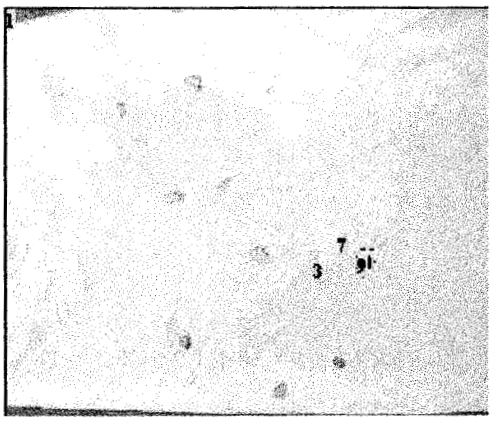

Figure 5.17b

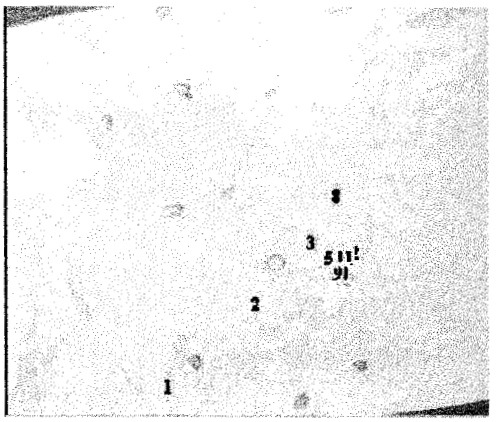

Figure 5.17c

Figure 5.17d

Figures 5.17a. and 5.17b. Corresponding points in the intensity image,

Figures 5.17c. and 5.17d. Corresponding points in the saturation image

5.3. Depth determination

The depth was determined using equation (4.12) for the stereo images. For calibration purpose two sets of simple stereo images with different heights were captured in the Hydrotechnical laboratory and depth is determined. Then the depth is determined for three sets of stereo images.

5.3.1. Calibration of depth

Two sets of images were taken with varying heights of concrete blocks. For the concrete block (Figure 5.18a and 5.18b) the height determined was 20.5 cm. and the actual height was 19.5 cm. For the concrete block (Figures 5.19a and 5.19b) the estimated height was 41.6 cm and the actual height was 40 cm. The accuracy is in the range of 95 – 98 %.

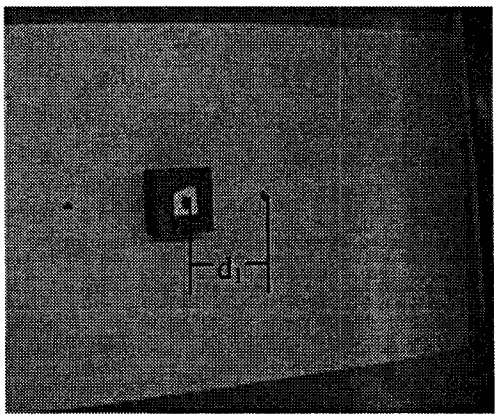

Figure 5.18a

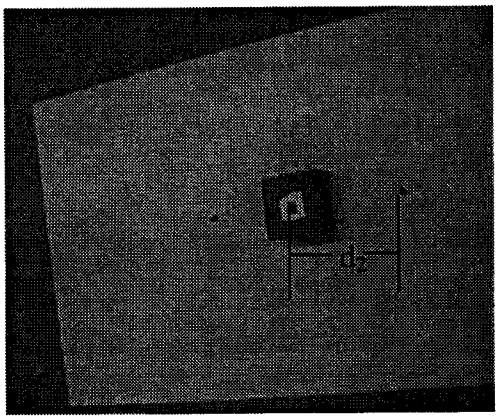

Figure 5.18b

Figure 5.18a. and 5.18b. Concrete block with a height of 19.5 cm from the ground

The corresponding points were not determined using the above-mentioned techniques. The reason is that because we need only two specific points on both the images and also there are not sufficient edges detected on those areas. So, the points were found by mouse clicking on the point on left and right image. The points selected by using this method may not be accurate to pixel position when compared to computer estimation of corresponding points.

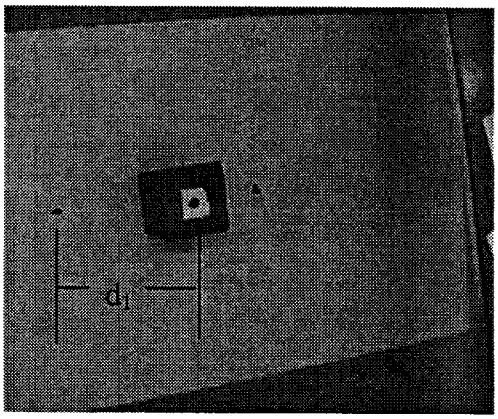

Figure 5.19a

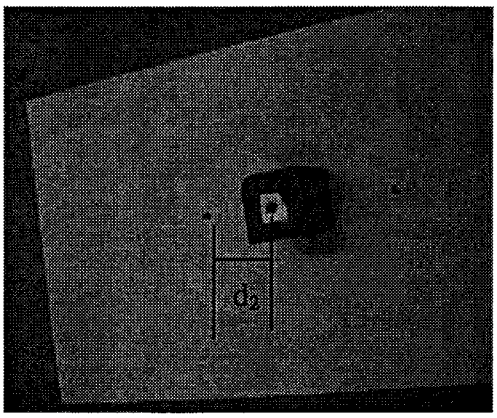

Figure 5.19b

Figures 5.19a. and 5.19b. Concrete block with a height of 40 cm from the ground

The image (Figure 5.20) was taken in the Hydrotechnical laboratory with objects of various heights and the depth is reconstructed using these images.

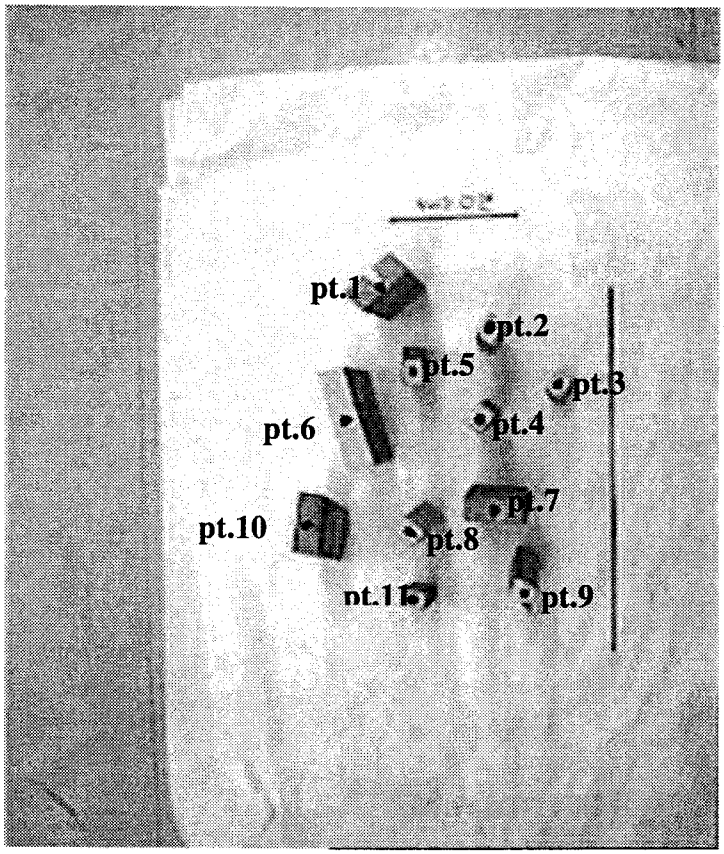

Figure 5.20. Laboratory image for objects with varying heights

5.3.2. Three-dimensional depth plot

The actual height of the objects is shown in Figure 5.21 and reconstructed depth plot is shown in Figure 5.22. The respective point numbers and the height in centimeters for each object is shown on each image for reference. The average accuracy for these images is 97.33%.

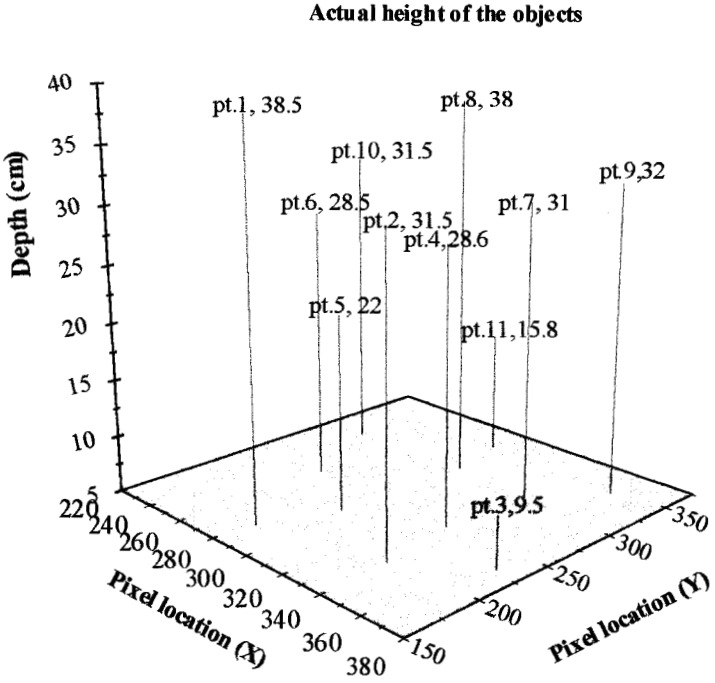

Figure 5.21. Depth plot for actual heights

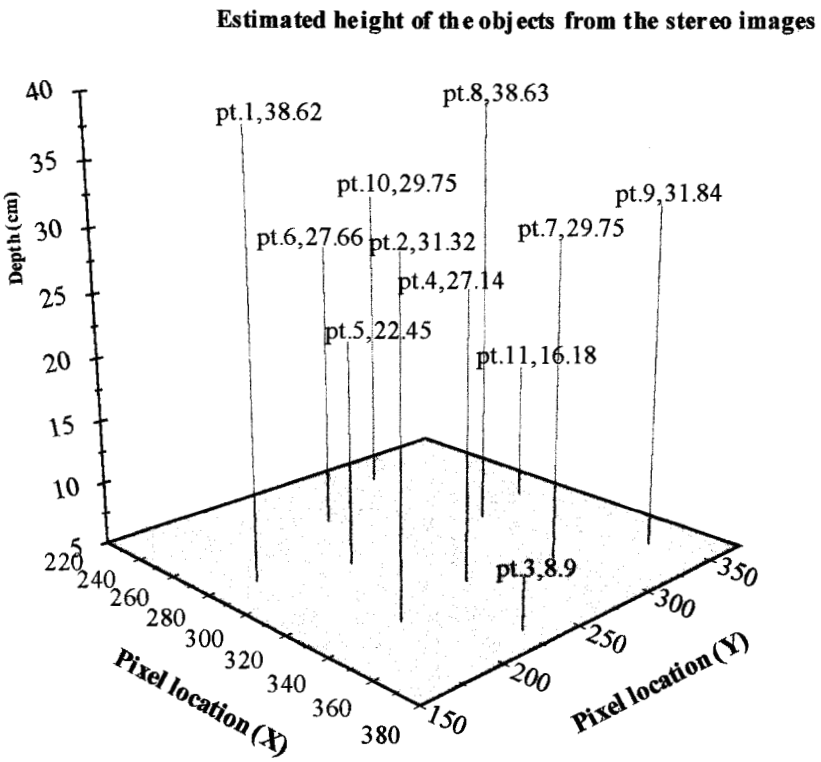

Figure 5.22. Depth plot for height estimated from stereo images

The depth reconstruction for the White House stereo image (Figure 5.2) is shown in Figure 5.23. Since the height from the ground and distance between the viewpoints is not known the depth is given in terms of pixels. The point number 12 and 13 is the roof of the White House in the stereo images and hence it shows significant height from the ground which can be observed from the depth plot.

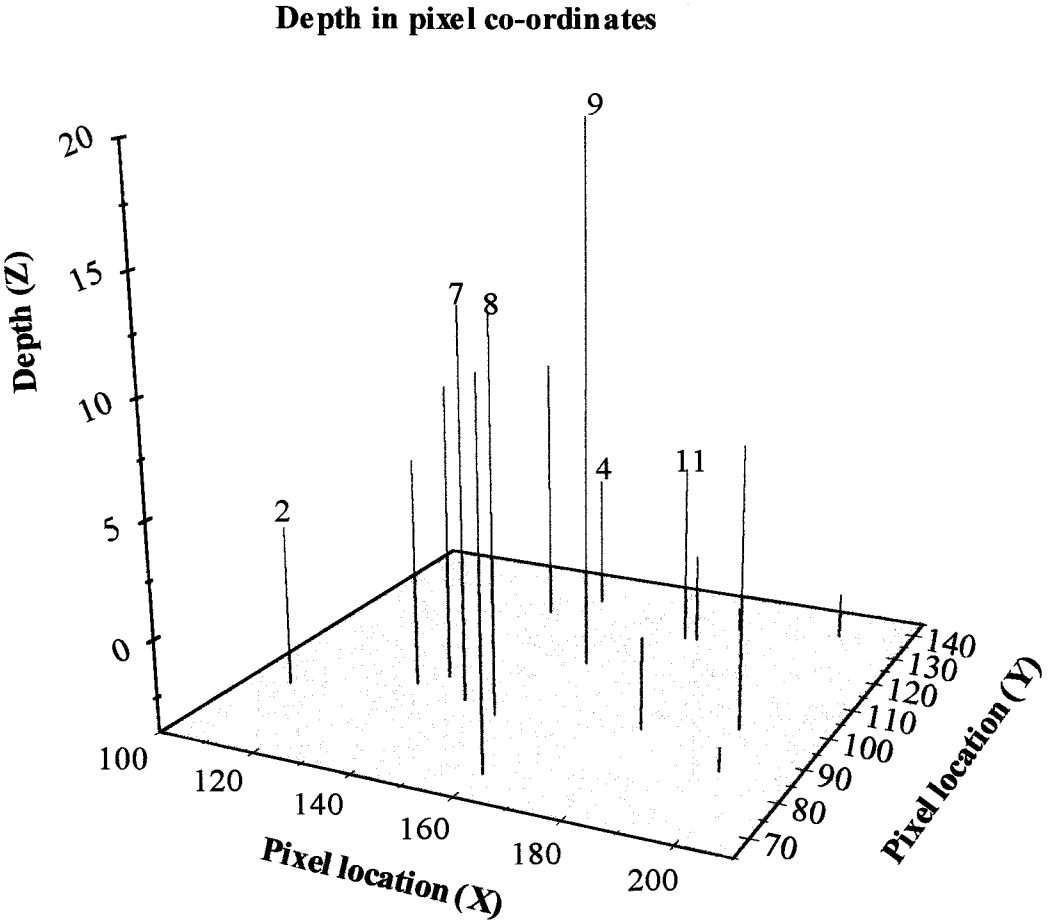

Figure 5.23. Depth plot for White House image

The depth plot for the images acquired in the Hydrotechnical laboratory (Figures 5.15c. and 5.15d.) is shown in the Figure 5.24. The depth plot is only a sparse plot because the number of points determined were not too many but the plot is an accurate plot because there are no false matches.

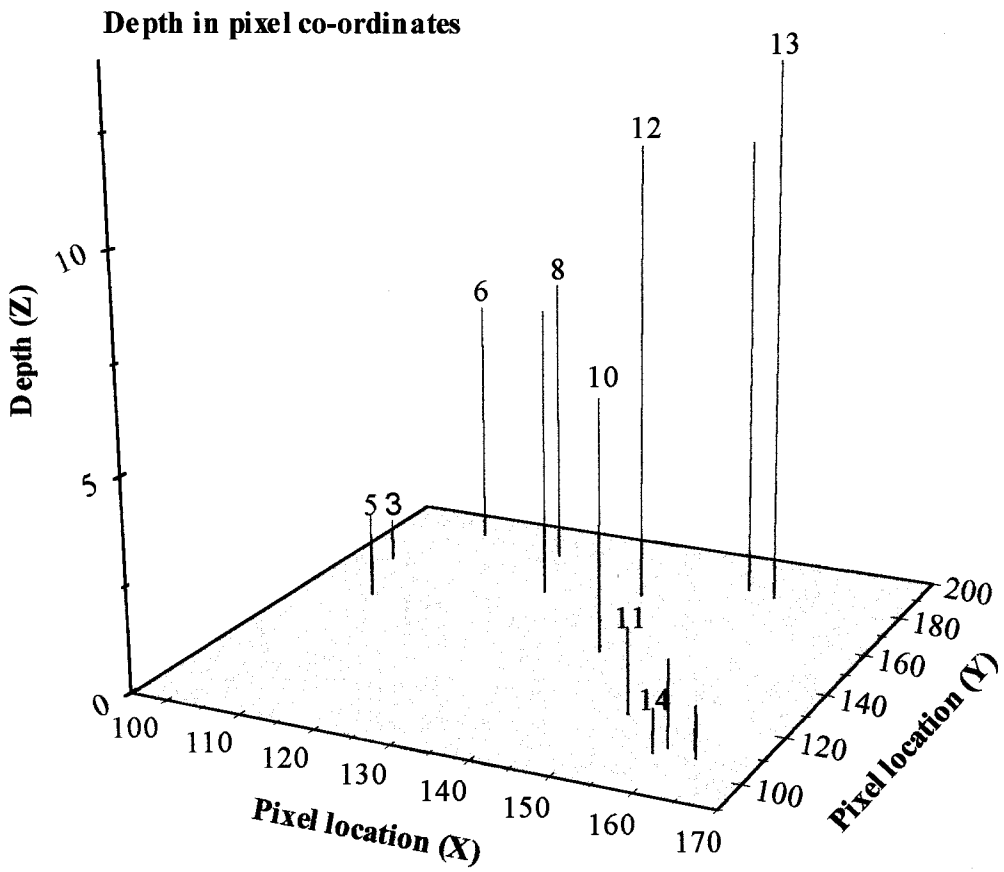

Figure 5.24. Depth plot for eroded sand bed

CONCLUSIONS AND FUTURE WORK

Conclusions

A stereoscopy scheme has been developed to measure the depth of scour in three-dimensional flow fields. The developed scheme was tested on various images. The stereoscopy scheme makes use of an epipolar constraint and a relaxation technique to match corresponding points in two images. A correlation technique has been developed to eliminate false matches. The use of color components plays a major role in determining the corresponding points on the two images. The type of color component (HSI) used depends on the image characteristics.

Several examples are shown to illustrate the usefulness of the algorithm developed. The algorithm is applied on grayscale images (e.g., White House image and surface of Mars image). The algorithm is also tested on various color images (image of a human being, shoulder bone image, artificial ball image and various laboratory images). In all cases, an appropriate edge detector (Marr-Hildreth or Shen-Caston edge detector) was used to find the edges. In most cases, the corresponding points were determined accurately.

After finding the corresponding points using the relaxation and correlation techniques, the depth is determined from the parallax between the points. The calibration of depth was also carried out with the image acquired in the laboratory.

In this work, a disparity gradient value of less than 0.5 was mostly used. A correlation coefficient value of 0.9 was observed to give good quality results for most types of images. The correspondence process fails to produce good matches if there are insufficient distinguishable features in the images. Since the computation of the essential matrix depends on the eight-point algorithm, its determination is influenced by noise and if there is too much deformation between the images. A number of features or changes in the intensity level are needed to compute the corresponding points between

the two images. In the laboratory images, the reason for poor results with fewer stones can be due to the quality of the images. There were not many features in these images. Due to poor lighting conditions there were many false edges. The results were good for the same scene with many stones as there was scope for more reliable edges. The depth of scour is determined using the parallax between the matched points.

Future Work

The task of identifying corresponding points can be made much simpler if rays of light are projected onto the scene, which eliminates the process of finding the edges. If the HSI values can be directly acquired without any calculations, the computational time can be reduced enormously.

The laboratory images were acquired with no water flowing in the sand bed. It is of practical importance to extend the algorithm for use with images having water layer on top. Furthermore, it is essential to determine the depth in real time. Determining depth using correspondence imaging may be difficult to do in real time because of the preprocessing involved (edge detection and calculating the HSI values). This merits further research.

REFERENCES

[1] *TR NEWS*, Nov – Dec 1997, No. 193, Critical Issues in Transportation, FAA's Security Research and Development Program, TRB Annual Report 1997.

[2] R. Balachandar, J. A. Kells and A. Sathish, "Instantaneous water surface and bed scour profiles using video image analysis", *CSCE Annual Conference*, Vol. 3, May 1997, pp. 31-40.

[3] P. R. Wolf, *Elements of Photogrammetry*, Second Edition, McGraw-Hill Book Company, 1983

[4] D. B. Diner and D. H. Fender, *Advances in Computer Vision and Machine Intelligence: Human Engineering in Stereoscopic Viewing Devices*, Plenum Press, New York, 1993.

[5] W. R. Hendee and P. N. T. Wells, *The Perception of Visual Information*, Second Edition, Springer-Verlag New York, Inc., 1997.

[6] J. C. Russ, *The IMAGE PROCESSING Handbook*, CRC Press Inc., Florida, 1992.

[7] M. J. Hannah, "A System for Digital Stereo Image Matching", *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 12, December 1989, pp. 1765-1770.

[8] A. Goshtasby, S. H. Gage and J. R. Bartholic, "A two-stage cross correlation approach to template matching", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 6, 1984, pp. 374-378.

[9] B. J. Super and W. N. Klarquist, "Patch-Based Stereo in a General Binocular Viewing Geometry", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 19, No. 3, March 1997, pp. 247-252.

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

[11] S. B. Pollard, J. E. W. Mayhew and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit", *Perception*, Vol. 14, 1985, pp. 449-470.

[12] P. Burt and B. Julesz, "Modifications of the classical notion of Panum's fusional area", *Perception*, Vol. 9, 1980, pp. 671-682.

[13] R. Deriche and O. Faugeras, "2-D Curve Matching Using High Curvature Points: Application to Stereo Vision", *10th International Conference on Pattern Recognition*, Vol. 1, 1990, pp. 240-242.

[14] G. Medioni and R. Nevatia, "Segment-Based Stereo Matching", *Computer Vision, Graphics and Image Understanding*, Vol. 31, 1985, pp. 2-18.

[15] Y. Chang and J. K. Aggarwal, "Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images", *Computer Vision, Graphics and Image Understanding*, Vol. 67, No. 2, August 1997, pp. 186-201.

[16] J. Canny, "A Computational Approach to Edge Detection", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-8, No. 6, November 1986, pp. 679-698.

[17] R. Horaud and T. Skordas, "Stereo Correspondence Through Feature Grouping and Maximal Cliques", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, No. 11, November 1989, pp. 1168-1180.

[18] J. S. Greenfield and A. R. Schenk, "Experiments with Edge-Based Stereo Matching", *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No. 12, December 1989, pp. 1771-1777.

[19] Y. C. Hsieh, F. Perlant and D. M. McKeown, "Recovering 3D Information from Complex Aerial Imagery", *10th International Conference on Pattern Recognition*, Vol. 1, 1990, pp. 136-146.

[20] A. Woods, T. Docherty and R. Koch, "Image Distortions in Stereoscopic Video Systems", *Stereoscopic Displays and Applications IV, Proceedings of the SPIE*, Vol. 1915, Feb 1993. (http://info.curtin.edu.au/-iwoodsa/spie93pa.html)

[21] Q. Yu, D. Zhang and T. Quan, "Accurate Measurement of 3D Coordinate of an Object with Subpixel Technique", *IEEE*, 1996, pp. 484-450.

[22] R. Y. Tsai and T. S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-6, No. 1, January 1984, pp. 13-26.

[23] Z. Zhang, R. Deriche, O. Faugeras and Q. Luong, "A robust technique for matching two uncalibrated images through the recovery of unknown epipolar geometry", *Artificial Intelligence J.*, Vol. 78, October 1995, pp. 87-119.

[24] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections", *Nature*, Vol. 293, 1981, pp. 133-135.

[25] R. I. Hartley, "In Defense of Eight-Point Algorithm", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 19, No. 6, June 1997, pp. 580-593.

[26] J. R. Jordan and A. C. Bovik, "Computational Stereo using Color", *Proceedings of the IEEE International Conference on System, Machine Analysis and Cybernetics*, 1987, pp. 531-535.

[27] D. C. Brockelbank and Y. H. Yang, "An Experimental Investigation in the Use of Color in Computational Stereopsis", *IEEE Transanctions on Systems, Man and Cybernetics*, Vol. 19, No. 6, November/December 1989, pp. 1365-1383.

[28] M. Dubuisson and A. K. Jain, "2D Matching of 3D Moving Objects in Color Outdoor Scenes", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 887-891.

[29] J. S. Weska, "A survey of threshold selection techniques", *Comput. Graph. Image Process*, Vol. 7, 1978, pp. 259-265.

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

[31] Z. Zhang, O. Faugeras and R. Deriche, *Videre: A Journal of Computer Vision Research (MIT Press)*, Vol. 1, No. 1, 1997, pp. 58-68.

[32] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach* , Kluwer Academic Publishers, 1996, ISBN 0-7923-4199-6.

[33] D. Marr and E. Hildreth, "Theory of Edge Detection", *Proc. Roy. Soc. London*, Vol. B 207, 1980 , pp. 187-217.

[34] J. R. Parker, *Algorithms for Image Processing*, Wiley Computer Publishing, Newyork, c1997.

[35] Z. Li and G. Hu, "Analysis of Disparity Gradient Based Cooperative Stereo", *IEEE Transactions on Image Processing*, Vol. 5, No. 11, November 1996, pp. 1493-1506.

[36] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, 1993.

[37] A. Rosenfeld, R. A. Hummel and S. W. Zucker, "Scene Labeling by Relaxation Operations", *IEEE Transantions on Systems, Man and Cybernetics*, Vol. SMC-6, No. 6, June 1986, pp. 420-433.

[38] J. C. Russ, *The IMAGE PROCESSING Handbook*, CRC Press Inc., Florida, 1996.

[39] B. Jahne, *Practical handbook on image processing for scientific applications*, CRC Press Inc., Florida, 1997.

[40] A. Boyde, "Quantitative photogrammetric analysis and qualitative stereoscopic analysis of SEM images", *Journal of Microscopy*, Vol. 98, Pt 3, August 1973, pp. 452-471.

[41] A. Koschan, V. Rodehorst and K. Spiller, "Color Stereo Vision Using Hierarchical Block Matching and Active Color Illumination", *Proceedings of ICPR 96, International Conference on Pattern Recognition*, August 25-29, 1996, pp. 835-839.

A. RELAXATION AND CORRELATION TECHNIQUES

The visual basic program shown in the appendix reads the image files and acquires the grayscale and color values. The epipolar line for the points of interest is determined using the essential matrix. There is an option in the program which allows for visualization of these epipolar lines. The corresponding points are then computed using relaxation technique. The disparity gradient, epiwidth value, number of iterations, image window size, correlation coefficient value and area of interest can be set by the user. After determining the corresponding points using relaxation technique, the false matches are eliminated using the correlation technique. The results are displayed in a window for easy analysis. The results are also stored in separate files for further processing. The main program is as follows:

```
Private Sub strrelax_Click()
Dim LocateX(1 To 2000), LocateY(1 To 2000) As Integer
Dim ResultX(1 To 2000), ResultY(1 To 2000) As Integer
Dim X, Y, i, j As Integer
Dim m1_x, m1_y, m1_z, Epival, High As Double
Dim R1C1, R1C2, R1C3 As Double
Dim R2C1, R2C2, R2C3 As Double
Dim R3C1, R3C2, R3C3 As Double
Dim HUE_L(1 To 11, 1 To 11), HUE_R(1 To 11, 1 To 11), Satur_L(1 To 11, 1 To 11),
Satur_R(1 To 11, 1 To 11) As Double
'ResultX(0) = 0
R1C1 = 0.000009649376
R1C2 = 0.0001975648
R1C3 = -0.01786204
R2C1 = -0.0001762583
R2C2 = 0.00001137961
R2C3 = -0.00775396
R3C1 = 0.01516555
R3C2 = -0.004582961
R3C3 = 0.9996848


Picture1.Visible = True
Picture2.Visible = True
Form1.MousePointer = 0
```

```
ScaleMode = 3 'Set it for pixel mode
Picture1.FillStyle = 0
Picture2.FillStyle = 7
CURZ = 1
i = 1
n = 1
Matchfound = 1
Form1.MousePointer = 11 'Change the pointer to hourglass
Rval = 0
Bval = 50
Gval = 250

  'FOR LEFT EDGE IMAGE

  Set Picture1.Picture = LoadPicture("C:\SATHISH\hslimage\Marrrocksl.jpg")

  For X = 1 To 317 Step 1
    For Y = 1 To 236 Step 1
      Color = Picture1.Point(X, Y)       'get the pixel value
      Color = Hex(Color)
      Red = Right(Color, 2)        'get the red value
      R = Val("&H" + Red)
      LeftImage(X, Y) = R
    Next Y
  Next X

  'FOR RIGHT EDGE IMAGE
  Set Picture2.Picture = LoadPicture("C:\SATHISH\hslimage\Marrrocksr.jpg")

  For X = 1 To 317 Step 1
    For Y = 1 To 236 Step 1
      Color = Picture2.Point(X, Y)       'get the pixel value
      Color = Hex(Color)
      Red = Right(Color, 2)        'get the red value
      R = Val("&H" + Red)
      RightImage(X, Y) = R
    Next Y
  Next X



  For CURX = 50 To 270 Step 1
  For CURY = 40 To 220 Step 3
  R = LeftImage(CURX, CURY)
```

```
If R > 240 Then
   For f = 1 To 900
      SM(1, f) = 0
   Next f
   m = 2

      LocateX(i) = CURX
      LocateY(i) = CURY
      m1_x = (R1C1 * CURX) + (R1C2 * CURY) + (CURZ * R1C3)
      m1_y = (R2C1 * CURX) + (R2C2 * CURY) + (CURZ * R2C3)
      m1_z = (R3C1 * CURX) + (R3C2 * CURY) + (CURZ * R3C3)
      Picture1.FillStyle = 0    ' Choose random FillStyle.

'To find the line which satisfies the equaion
'm2 transpose * Fundamental matrix * m1 matrix
l = 1

   For X = CURX + 10 To CURX + 40 Step 1
      For Y = CURY - 5 To CURY + 5 Step 1
         Epival = (X * m1_x) + (Y * m1_y) + (1 * m1_z)
         If Epival >= 0 And Epival <= 0.1 Then
         R = RightImage(X, Y)
            If R > 240 Then
               CorrX(i, l) = X
               CorrY(i, l) = Y
               l = l + 1
            End If

         End If
      Next Y
   Next X
   'MsgBox "L value is " + Str(l)
   l = l - 1
   cnt = l
   Change1 = 0
   Change2 = 1
   Ch_nge1 = 2
   Ch_nge2 = 2
   DirectionX1 = 2
   DirectionX2 = 2
   DirectionY = 1
   xstep1 = 1
   xstep2 = 100
   stepval = 1
   TempY1 = CURY
   TempY2 = CURY + 1
   Ywidth1 = 1
```

```
   Ywidth2 = 1
   param = 3
   m = 1

Do While (m < 15)
   Label3.Caption = "Counter value" + Str(m)
Labelskip2:
If param Mod 2 = 1 Then
   Picture1.Circle (CURX, CURY), 2, QBColor(9)
   If (DirectionX1 Mod 2 = 0) Then
      xstep1 = 1
      xstep2 = 100
      stepval = 1
      DirectionX1 = DirectionX1 + 1
   ElseIf (DirectionX1 Mod 2 = 1) Then
      xstep1 = -1
      xstep2 = -100
      stepval = -1
      param = param + 1
      'Change = Change + 3
      DirectionX = DirectionX + 1
   End If
      Picture1.SetFocus
      cnt_val = 0
      For TempX = CURX + xstep1 To CURX + xstep2 Step stepval
         R = LeftImage(X, Y)
         cnt_val = cnt_val + 1
         If R > 240 Then
            LeftX = TempX
            LeftY = TempY1
            LeftX_val = R
            GoTo Label11
         End If
      Next TempX
      If cnt_val > 99 Then
         GoTo Labelskip1
      End If
Label11:
      Call It_Y1(1, i, xstep1, xstep2, stepval, m)
      If DirectionX1 Mod 2 = 0 Then
         TempY1 = TempY1 - 1
      End If
Labelskip1:
Else
   Picture1.Circle (CURX, CURY), 2, QBColor(9)

      If (DirectionX2 Mod 2 = 0) Then
```

```
            xstep1 = 1
            xstep2 = 100
            stepval = 1
            DirectionX2 = DirectionX2 + 1
        ElseIf (DirectionX2 Mod 2 = 1) Then
            xstep1 = -1
            xstep2 = -100
            stepval = -1
            param = param + 1
            'Change = Change + 3
            DirectionX2 = DirectionX2 + 1
        End If
        cnt_val = 0
        For TempX = CURX + xstep1 To CURX + xstep2 Step stepval
            R = LeftImage(X, Y)
            cnt_val = cntval + 1
            If R > 240 Then
                LeftX = TempX
                LeftY = TempY2
                LeftX_val = R
                GoTo Label12
            End If
        Next TempX

        If cnt_val > 99 Then
            GoTo Labelskip2
        End If
Label12:
        Call It_Y2(l, i, xstep1, xstep2, stepval, m)
        If DirectionX2 Mod 2 = 0 Then
            TempY2 = TempY2 + 1
        End If

    End If

    Loop

High = SM(i, 1)
    For c = 1 To l
        If SM(1, c + 1) > High Then
            High = SM(1, c + 1)
        Else
            High = High
        End If
    Next c
    'MsgBox "High is " + Str(High)
    'If High <> 0 Then
```

```
      For c = 1 To l
      If SM(1, c) = High Then
         ResultX(i) = CorrX(i, c)
         ResultY(i) = CorrY(i, c)
         Picture1.Circle (LocateX(i), LocateY(i)), 2, QBColor(12)
         Picture2.Circle (ResultX(i), ResultY(i)), 2, QBColor(12)
         'MsgBox "Wait"
         GoTo Label230
      End If
   Next c
   'Else
      'GoTo Nextpoint
   'End If
Label230:
   Label3.Caption = "Corresponding X Point" + Str(ResultX(i))
   Label4.Caption = "Correspoinding Y Point" + Str(ResultY(i))
   i = i + 1
Nextpoint:
   End If
      Next CURY
   Next CURX
   'Next p

   Set Picture1.Picture = LoadPicture("C:\SATHISH\hslimage\Satrcocksl.jpg")
   Set Picture2.Picture = LoadPicture("C:\SATHISH\hslimage\Satrocksr.jpg")
'If Matchfound = 1 Then
   s = 1
   For j = 1 To i - 1 Step 1
      a = 1
      For TempX = LocateX(j) - 3 To LocateX(j) + 3 Step 1
         b = 1
         For TempY = LocateY(j) - 3 To LocateY(j) + 3 Step 1
            Color = Picture1.Point(TempX, TempY)
            Color = Hex(Color)
            Red = Right(Color, 2)          'get the red value
            R = Val("&H" + Red)
            Green = Mid(Color, 3, 2)       'get the green value
            GR = Val("&H" + Green)  'convert it to integer value
            Blue = Left(Color, 2)          'get the blue value
            BL = Val("&H" + Blue) 'convert it to integer value
            ResultLR(a, b) = BL
            'ResultLG(a, b) = GR
            'ResultLB(a, b) = BL
            'U = (0.666667 * R) - (0.333333 * GR) - (0.333333 * BL)
            'V = (-0.333333 * R) + (0.666667 * GR) - (0.333333 * BL)
            'HUE_L(a, b) = Atn(V / U)
            'Satur_L(a, b) = Sqr(U ^ 2 + V ^ 2)
```

```
            b = b + 1
          Next TempY
          a = a + 1
        Next TempX

        Mean_Left = 0
        For a = 1 To 7
          For b = 1 To 7
            Mean_Left = Mean_Left + ResultLR(a, b)
            'Mean_Left = Mean_Left + HUE_L(a, b)
          Next b
        Next a
        Mean_Left = Mean_Left / 49
a = 1
        For TempX = ResultX(j) - 3 To ResultX(j) + 3 Step 1
          b = 1
          For TempY = ResultY(j) - 3 To ResultY(j) + 3 Step 1
            Color = Picture2.Point(TempX, TempY)
            Color = Hex(Color)
            Red = Right(Color, 2)          'get the red value
            R = Val("&H" + Red)
            Green = Mid(Color, 3, 2)       'get the green value
            GR = Val("&H" + Green)  'convert it to integer value
            Blue = Left(Color, 2)          'get the blue value
            BL = Val("&H" + Blue) 'convert it to integer value
            ResultRR(a, b) = BL
            'ResultRG(a, b) = GR
            'ResultRB(a, b) = BL
            'U = (0.666667 * R) - (0.333333 * GR) - (0.333333 * BL)
            'V = (-0.333333 * R) + (0.666667 * GR) - (0.333333 * BL)
            'HUE_R(a, b) = Atn(V / U)
            'Satur_R(a, b) = Sqr(U ^ 2 + V ^ 2)
            b = b + 1
          Next TempY
          a = a + 1
        Next TempX
Mean_Right = 0
        For a = 1 To 7
          For b = 1 To 7
            Mean_Right = Mean_Right + ResultRR(a, b)
            'Mean_Left = Mean_Left + HUE_R(a, b)
          Next b
        Next a
        Mean_Right = Mean_Right / 49

        Strength1 = 0
        Strength2 = 0
```

```
      Strength3 = 0
      For a = 1 To 7
        For b = 1 To 7
        Strength1 = Strength1 + Abs(ResultLR(a, b) - Mean_Left) * Abs(ResultRR(a, b)
- Mean_Right)
        Strength2 = Strength2 + ((ResultLR(a, b) - Mean_Left) ^ 2)
        Strength3 = Strength3 + ((ResultRR(a, b) - Mean_Right) ^ 2)
        Next b
      Next a
      If Strength1 = 0 Then
        Coeff = 1
      Else
        Coeff = Strength1 / Sqr(Strength2 * Strength3)
      End If

      'MsgBox "For I = " + Str(j)
      'MsgBox "Correlation coeff " + Str(Coeff)
      'Coeff = InputBox("Enter the coeff value", "Coeff", 0.75)
      If Coeff >= 0.9 Then
        someLX = LocateX(j)
        someLY = LocateY(j)
        someRX = ResultX(j)
        someRY = ResultY(j)
        LocateX(s) = someLX
        LocateY(s) = someLY
        ResultX(s) = someRX
        ResultY(s) = someRY
        Picture1.Circle (LocateX(s), LocateY(s)), 3, QBColor(12)
        Picture2.Circle (ResultX(s), ResultY(s)), 3, QBColor(12)
        'MsgBox "Wait"
        s = s + 1
      End If
    Next j
'MsgBox "S is " + Str(s)
Form1.MousePointer = 0 'Back to normal
Picture1.FillStyle = 0
Picture2.FillStyle = 7
    Set Picture1.Picture = LoadPicture("C:\SATHISH\hslimage\rockl.jpg")
    Set Picture2.Picture = LoadPicture("C:\SATHISH\hslimage\rockr.jpg")
For j = 1 To s - 1
    Picture1.SetFocus
    Picture1.Circle (LocateX(j), LocateY(j)), 3, QBColor(12)
    Load Label1(j)
    Label1(j).Visible = True
    Label1(j).Left = LocateX(j)
    Label1(j).Top = LocateY(j)
    Label1(j).AutoSize = True
```

```
   Label1(j).Caption = j
   Picture2.SetFocus
   Picture2.Circle (ResultX(j), ResultY(j)), 3, QBColor(12)
   Load Label2(j)
   Label2(j).Visible = True
   Label2(j).Left = ResultX(j)
   Label2(j).Top = ResultY(j)
   Label2(j).AutoSize = True
   Label2(j).Caption = j
Next j
Lastline:
Dim FilName As String
 Message = "Enter the file name"   ' Set prompt.
 Title = "Matched points" ' Set title.
 'Default = "C:\sathish\thesis\whmatchpts.doc"   ' Set default.
Default = "c:\sathish\thesis\satrocks3dpts.doc"   ' Set default.
 ' Display message, title, and default value.
FilName = InputBox(Message, Title, Default)
Dim Today
Today = Now ' Assign current system date and time.
Open FilName For Output As #10
   Print #10, Str(Now)
   Print #10, Tab; Tab; "Matched Points"
   Print #10, "No."; Spc(3); "LeftX"; Spc(3); "LeftY"; Spc(3); "RightX"; Spc(3);
"RightY";
   For j = 1 To s - 1
      Print #10, Str(j); Spc(3); Str(LocateX(j)); Spc(3); Str(LocateY(j)); Spc(3);
Str(ResultX(j)); Spc(3); Str(ResultY(j));
   Next j
Close #10

End Sub
```

B. MATCHED POINTS

The matched points determined using the relaxation and the correlation techniques are stored in a file. Here LeftX represents the x coordinate of the left image and LeftY represents the y coordinate of the left image. Similarly, RightX and RightY represents the x and y coordinates of the right image respectively. The results for the White House image are as follows:

7/27/98 3:21:19 PM

Matched Points

No.	LeftX	LeftY	RightX	RightY
1	108	84	103	83
2	116	90	109	88
3	162	105	151	102
4	172	108	158	105
5	176	168	163	168
6	187	185	180	183
7	189	99	171	98
8	205	94	187	93
9	219	136	194	136
10	224	61	207	59
11	236	167	227	167
12	239	107	234	106
13	250	179	248	178
14	252	102	247	101
15	282	169	257	169
16	289	115	276	111
17	289	188	286	185
18	290	87	288	86