University of Alberta

Novel 3D Back Reconstruction using Stereo Digital Cameras

by

Anish Kumar

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Biomedical Engineering

Electrical and Computer Engineering

Fall 2011
Edmonton, Alberta

Dedication

To my Mom and Dad, Veena and Dr. Ajit Sinha,

to my girlfriend, Vaidehi Seth,

for your support and love

Abstract

This thesis presents the research and development of a novel procedure for creating a 3D image of the back using stereo digital 2D images. The procedure requires minimal user input and is intuitive. The procedure is comprised of 3 stages – image data acquisition, image registration and image reconstruction. Back Image data was acquired using automated template matching of a pegboard in a unique stereo camera configuration that improves speed over existing processes. To improve the registration process, a new approach combining image segmentation and differential geometry with belief propagation was developed. Stray data points were removed in the 3D Back image reconstruction process and missing data points were interpolated using a unique method of Moving Least Squares – Bezier Curve. The procedure was tested on human subjects. The results demonstrate that the procedure can be used clinically to obtain 3D back images for the evaluation of scoliosis.

Acknowledgements

I would like to thank my supervisor, Dr. Nelson G. Durdle for his guidance, support and encouragement throughout my studies.

I would like to express my gratitude towards Dr. Peter Ajemba for his valuable advice and suggestions during my research work.

I would like to thank Ahmed Elsafi for being a volunteer throughout my research experimentation.

Last but not the least, I would like to thank my parents, Veena Sinha and Dr. Ajit Kumar Sinha for their unconditional love and support, my girlfriend Vaidehi Seth for her support and prayers and to many friends for their help throughout my studies and research work.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Purpose

This thesis describes the research and development of a procedure for reconstruction of 3D images of the torso from 2D stereo back images. The aim of this procedure is to obtain 3D images of the torso that can be used in the assessment of scoliosis. The procedure expands on the existing techniques of computer vision. It improves the image registration stage of the procedure by applying a novel approach of Mean Shift Segmentation, Max Product Belief Propagation and differential geometry. 3D back image reconstruction stage is improved by using Moving Least Squares – Bezier Curve interpolation to fill in missing data points.

Using a stereo camera setup consisting of digital cameras to acquire images makes the procedure more cost effective than using commonly used techniques like laser scanning. We increase speed and accuracy of stereo image data acquisition by replacing the commonly used manual calibration process by an automated Template Matching Image Rectification. The proposed procedure provides clinicians with a cost effective and mobile method of acquiring 3D torso images to study Scoliosis.

1.2 Motivation

The assessment of severity of scoliosis is traditionally done using radiographs of the spine. However, radiographs do not describe the visible torso deformity associated with scoliosis [1]. Many three dimensional data acquisition techniques such as rasterstereography [14], laser scanning systems [15] etc. have been investigated for developing a system to assist clinicians in the evaluation of external scoliotic deformities. This is because most scoliosis patients and their families are more concerned with the shape of the torso than the internal alignment of the spine [2, 28].

Traditional procedures for assessment of torso shape are based on landmarks. Since the back surface is smooth and featureless, it becomes very difficult to locate these landmarks in real time. Techniques such as difference mapping, Moire topography [11] and laser scanning have been developed over the years for assessment of torso shape. Disadvantages in these methods range from poor resolution images, long acquisition times to expensive systems.

In light of these problems and due to advancements in stereo computer vision, there is a need to develop a cost effective, accurate technique that can be clinically used for assessing scoliosis.

1.3 Objectives

The objective of this thesis is to create a procedure to reconstruct a 3D torso image from 2D stereo images of the back using a stereo camera setup. The 3D torso image can be used for the assessment of scoliosis clinically. To be successful in a clinical environment, the procedure needs to be fast, error free and relatively inexpensive [35]. This means that we need a procedure with minimal user input to prevent errors; to remove artifacts and to fill in missing data points during reconstruction; to account for curvature and smoothness of the human torso; to be cost effective and more accurate than existing methods. The aim of this thesis is three-fold.

1. To automate and simplify stereo camera calibration with respect to processing time and ease of use.
2. To create a novel procedure of stereo image correspondence matching in a pair of digital stereo images by investigating and improving on existing methods in stereo image registration.
3. To preprocess the registered image leading to a reconstructed 3D image of the torso.

1.4 Thesis Outline

An outline of each chapter of the thesis is provided below. Chapter 1 provides an introduction to the thesis. It presents the purpose and motivation behind the research work undertaken. An introduction to each chapter is also provided below.

Chapter 2 presents a review of the literature. It describes scoliosis and the torso deformity resulting due to it. Spinal deformity in scoliosis is assessed using radiographs, however external deformities in torso shape are evaluated using range images of the torso. Operation and application of range imaging systems to the study of scoliosis is also presented. The need for a better method for imaging torso is also discussed due to disadvantages of range imaging systems. Finally, a literature survey of the latest techniques in stereo computer vision is presented. We look at state of the art procedures for registration of stereo digital images to reconstruct 3D objects.

Chapter 3 presents a novel procedure of fundamental matrix estimation for smooth and curved surfaces like the torso. The stereo images are rectified with respect to each other by using the fundamental matrix. Image rectification establishes an epipolar constraint between the images and makes the search for corresponding points easier. Epipolar constraint makes the search for corresponding points in stereo images easier. It is explained in detail in section 2.3.1.1. We also present the stereo camera positioning and placement used to acquire the images. Experimentation performed to find the most suitable stereo camera placement for image acquisition is also described. Benefits of image rectification over manual camera calibration are also discussed.

Chapter 4 describes the essential steps required in the image registration process in detail. The reasons for selection of these methods over other existing ones are discussed. Finally, improvements made to point-to-point correspondence matching (image registration) so that it can be used on smooth and curved objects like the torso are described.

The last stage in the procedure (Chapter 5) is to obtain the 3D object from the stereo digital images. Since, the registration process leaves stray points due to errors, these are removed. Triangulation to connect the 3D points (obtained through registration) into polygons is described. Occlusions are found in the triangulated 3D image especially in regions of high curvature due to stray data removal. So, lastly, we describe a method for filling occlusions.

Chapter 6 deals with tests conducted on known models and shape to evaluate the system performance against existing stereo vision procedures and range scanning systems. The methodology behind testing, accuracy of the system, computational time, sources of error in reconstruction is also described in this chapter.

Chapter 7 presents conclusion and suggestion for future work.

2. Literature Review

2.1 Scoliosis

2.1.1 What is scoliosis?

Scoliosis is a condition characterized by lateral deviation of the spine coupled with rotation of individual vertebra resulting in visible torso asymmetries [1]. Rotation of the vertebrae causes the ribcage to distort and a hump to be produced in the back of the torso [2]. Radiological assessment of the lateral curvature of scoliosis includes measurement of the Cobb angle [3] (figure 2.1).

To determine the Cobb's angle, the vertebrae that is most displaced and rotated (apical vertebrae), with the least tilted end plates is identified. Next, end vertebrae that have minimum displacement and rotation, but are most tilted from its original position are identified. A line is drawn parallel to the plates of the end vertebrae that are furthest away from the apical vertebrae. Next, two lines perpendicular to the first pair of lines are drawn towards each other until they intersect. This resulting angle is known as Cobb's angle [3]. Two to four percent of the population has scoliosis of at least 10 degree Cobb angle and ten to thirty percent of them have curvature greater than 20 degrees. Most people with scoliosis are girls and most people develop scoliosis during adolescent years [4]. In adolescents, scoliosis can progress at an alarming rate and have severe effect on cosmesis of an individual [2, 15].

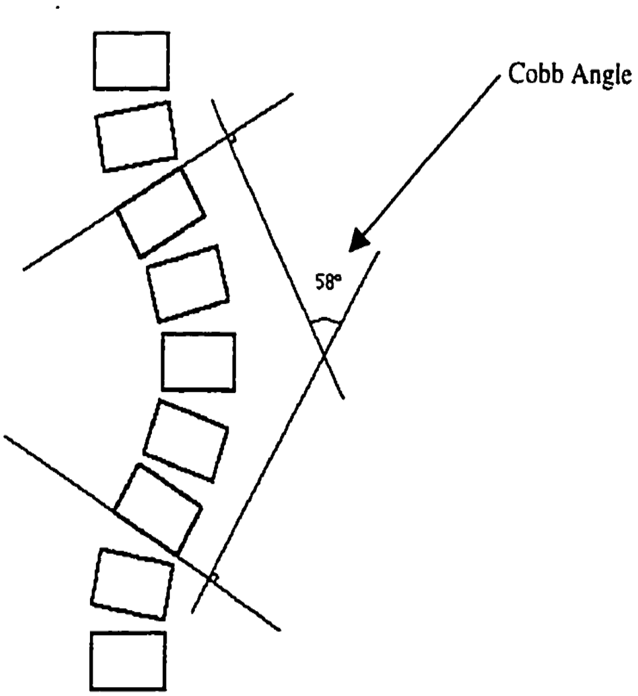

Figure 2.1 Cobb Angle Measurement

Some of the visible characteristics of scoliosis are 1) one shoulder is higher than the other; 2) one scapula (shoulder blade) may be higher or more prominent than the other; 3) the trunk is shifted over the pelvis; 4) one hip may appear to be higher or more prominent than the other is; 5) the head is not centered over the pelvis; and 6) when the patient is examined from behind and asked to bend forward until the spine is horizontal, one side of the back appears higher than the other side (figure 2.2) [1, 2, 22].

Most cases of scoliosis are of unknown etiology and termed idiopathic [1, 2]. Consequently, there are no acceptable preventive measures for scoliosis. In rare cases of severe scoliosis, organ development and pulmonary function can be affected [5]. However, most cases of late onset scoliosis will have no health risks associated with them [6].

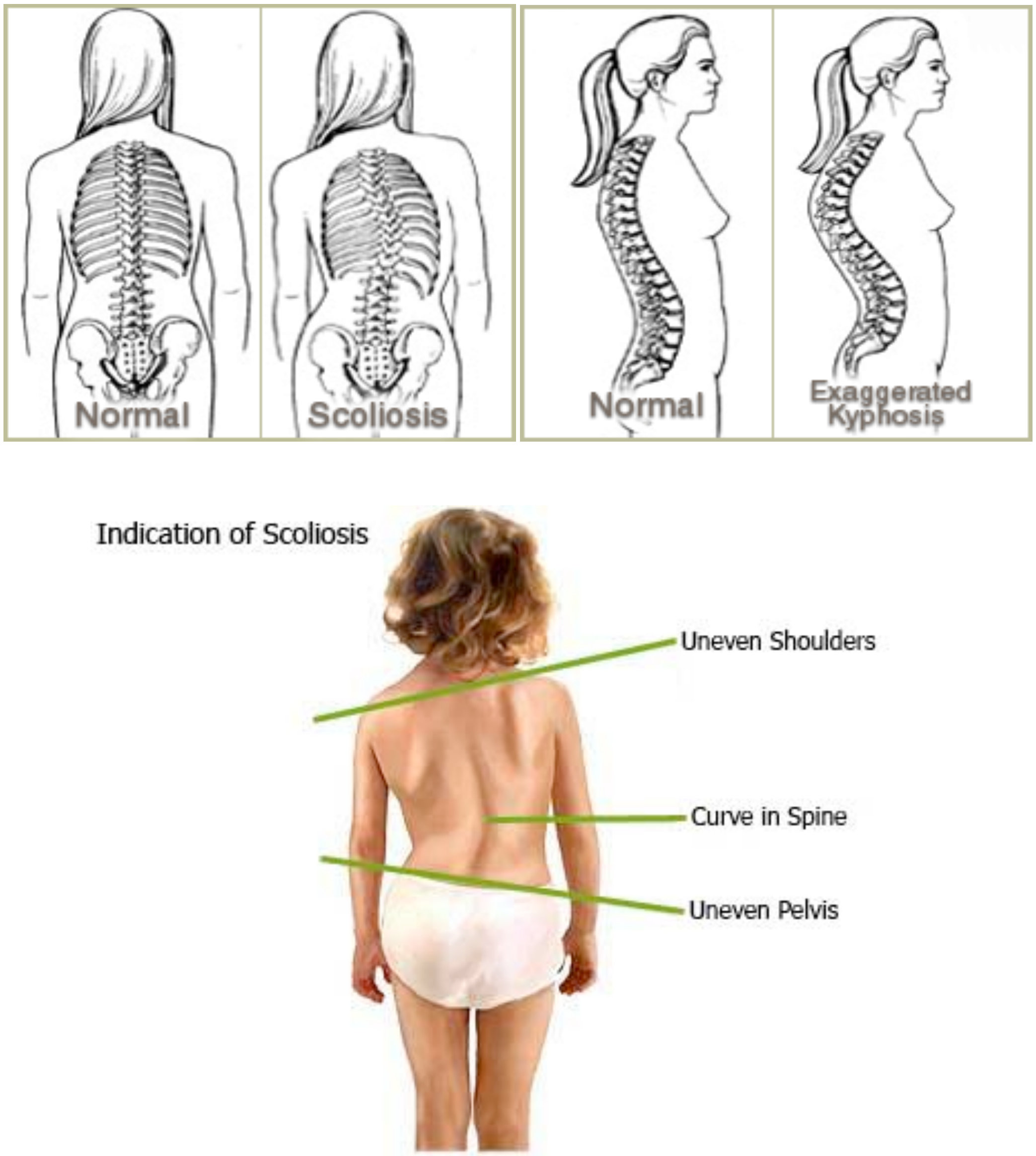

Figure 2.2 Visible characteristics of scoliosis

(source: http://chiropracticalliance.com.au/images/scoliosis.jpg)

2.1.2 Relevance of torso surface images in scoliosis

Scoliosis treatment is influenced by the extent of the spinal curvature. The aim of the treatment is three fold 1) improve internal alignment of trunk; 2) improve the external appearance of the torso; and 3) halt progression of the deformity. Quantification of the spinal curvature is done by the measurement of Cobb angle on radiographic images [7]. There are generally two sets of radiographs 1) posterior-anterior and 2) lateral; both taken in an upright position to assess scoliotic patients. However, there are several disadvantages associated to the use of radiography. Firstly, due to effect of ionization, some authors have linked radiography to increased risk of cancer [8]. Secondly, radiographs do not describe visible torso deformity associated with scoliosis [9]. This has led to increased efforts to find alternative methods for assessing the internal and external effects of scoliosis. Moreover, many scoliosis patients and their families are more concerned with the shape of the torso than the internal alignment of the spine [2, 28]. A reliable 3D model of the entire trunk surface will allow the development of predictive tools to help clinicians and their patients with decisions, not only based on spinal correction, but also on aesthetic improvement [10]. Over the years, imaging systems that produce torso surface images and assist clinicians in assessment of external torso deformities have been proposed.

Examples of back imaging techniques include Moiré Topography [11], Integrated Shape Imaging System (ISIS) scanning [12], Quantec System scanning [13], rasterstereography [14] and range scanning [15] (figure 2.3). Most of these systems are slow, have poor resolution, are not very portable or are expensive [16]. For instance, the ISIS scanner captures just the back surface in 2 seconds, with a resolution of 1.5 mm and an accuracy of 3 mm in 3D over a volume of 400x500x300 mm [10]. The effects of posture variations; sway and breathing due to the slow acquisition process and the poor resolution are major disadvantages of ISIS scanning.

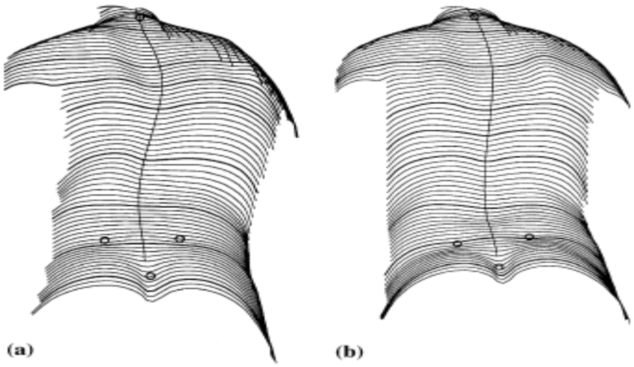

Figure 2.3(a) Rasterstereographic surface reconstruction result

(source: Hackenberg et. al. 2003 [14])

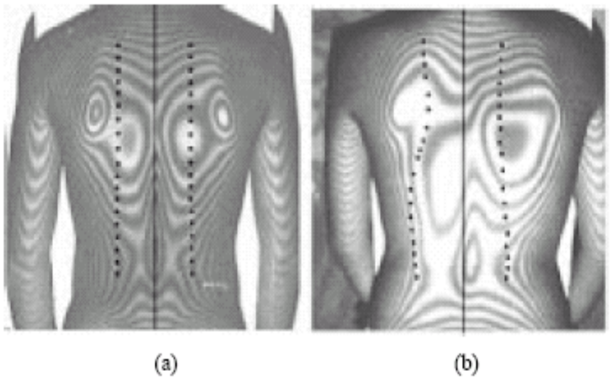

Figure 2.3(b) Moire Topography resultant output (source: Daruwalla et. al. 1985 [11])

Structured light approaches such as Quantec scanning and rasterstereography rely on projecting one or more special light patterns onto a scene, usually in order to directly acquire a range map of the scene [17]. These systems typically use at least one camera and a projector under known geometry and controlled lighting. The measurement principle of such a system is based on triangulation where the projector generates light patterns and the cameras detect the illuminated scene. Before a structured light system is used for 3D torso surface imaging, it is essential that the system be carefully calibrated to obtain its intrinsic parameters (such as focal lengths, scale factors and distortion coefficients) and extrinsic parameters (positions and orientation between the two components) [17, 18]. Though the specific geometry and controlled lighting reduces computational complexity, it adds in calibration time and restricts to system to use in a hospital or clinic. The difficulties associated with the above systems call for the development of a cost effective and accurate imaging technique for the assessment of scoliosis. Next, we discuss one the prominent imaging techniques called Range Scanning. Range scanning is used in many torso imaging clinics including the Glenrose Rehabilitation Hospital. We describe some of the problems particular to using range scanning in torso imaging.

2.2 3D Range Scanning Systems

The range scanning we describe is based on the Minolta Vivid 700 laser digitizer.

2.2.1 Operation

Range scanning digitizers use optical triangulation. A laser light beam shines on a spot on the surface of an object. This beam is scattered in many directions, and a camera records an image of the lighted spot. The center pixel of this spot is found and a line of sight traced through the pixel until it intersects the illumination beam at the point on the surface. This yields a single range point. To obtain the coordinates of an entire surface, the laser beam is systematically swept all over the surface of the object using mirrors. In the Minolta VIVID 700 digitizer ® (figure 2.4), the beam is fanned into a sheet of laser light. This casts a stripe onto the surface of the object which is then captured using a charge-coupled device (CCD) camera. CCD camera is described in detail in section 3.1. For each camera scan line of each stripe, the centre pixel is computed and a line of sight traced to intersect the corresponding portion of the laser scan. This yields a range profile of the object. The shape of the object can be obtained by sweeping the laser beam over its surface (figure 2.5).

The Minolta Vivid 700 digitizer uses an eye-safe laser beam and a galvanometer-mirror to sweep the beam over the object at a resolution of 200 by 200 and 256 levels per point. Its digital camera has a resolution of 400 by 400 pixels. It takes about 1 second to sweep laser beam over the object and about 2 seconds for data transfer to the computer using a SCSI interface. The cloud of range points obtained is triangulated and converted into a 3D surface. The digital camera captures the texture map independent of the range point cloud.

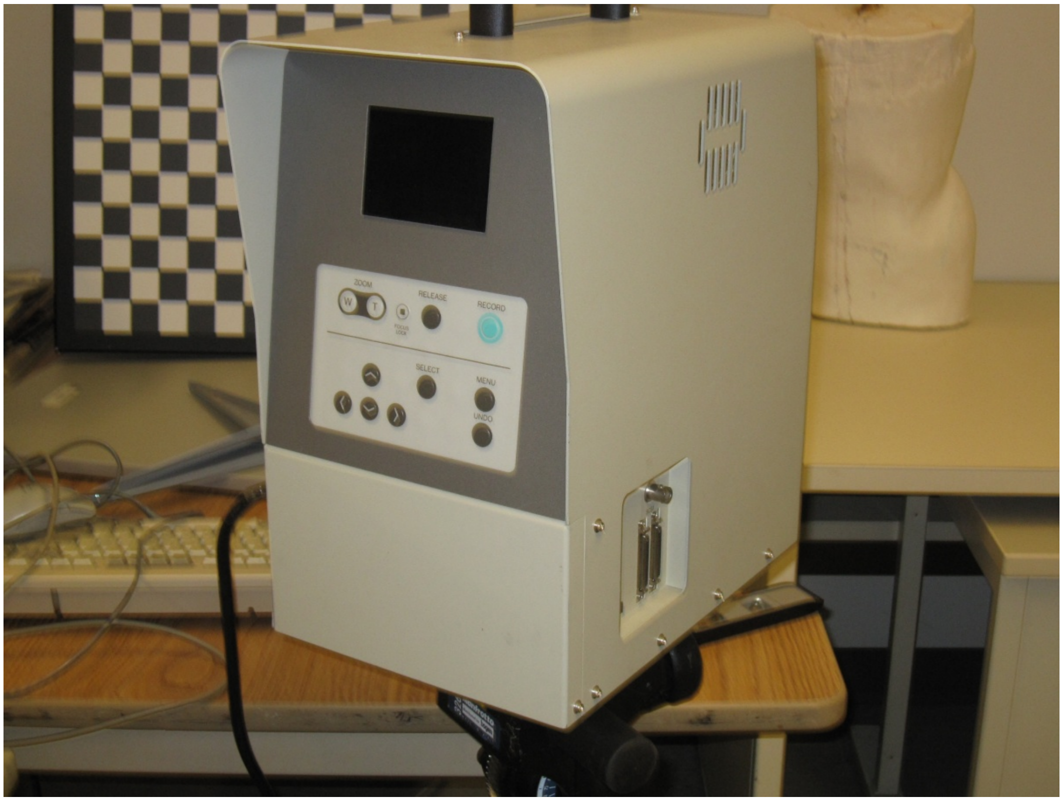

Figure 2.4 Minolta Vivid 700 digitizer

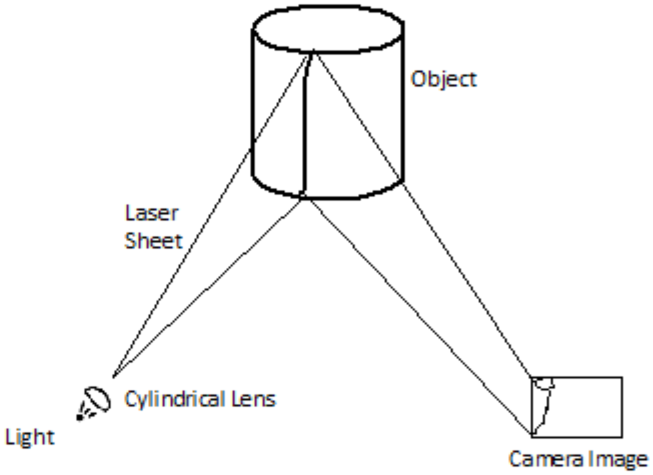

Figure 2.5 Triangulation on Range Scanning System

2.2.2 Disadvantages particular to torso imaging

Range scanning of human torsos has many disadvantages. First, they are expensive. Second, stray data points from surrounding artifacts have to be manually removed. Third, holes due to grazing angles of incidence of the laser beam are often located in obscure regions. Fourth, they are time consuming thereby increasing the risk of additional artifacts – It takes about 1 second to sweep laser beam over the object and about 2 seconds for data transfer to the computer using a SCSI interface [28]. Fifth, triangulation and smoothing algorithms applied by the digitizer during initial processing can cause additional errors. In the light of these problems, there is a need to develop a cost effective, accurate technique that can be clinically used for assessing scoliosis.

2.3 Stereo Computer Vision

Reconstructing accurate shape from images is a long-standing and challenging problem in stereo computer vision. In the past few years, there has been a fast proliferation of methods for the 3D reconstruction of objects from the analysis of camera images [19]. Stereo Computer Vision infers 3D scene geometry. It consists of a pair of digital cameras under a known geometry to acquire two images of an object and a computer to process these images (figure 2.6). There are various applications of stereo computer vision ranging from computer graphics, facial expression recognition, surgical planning, architectural structure design etc [20]. Progress in digital camera technology and availability of faster computer processors has lead to development of stereo vision algorithms for simultaneous stereo camera capture, calibration and reconstruction [16]. Stereo imaging systems are used to reconstruct 3D surface image of the torso from a pair of 2D digital stereo images under epipolar constraint.

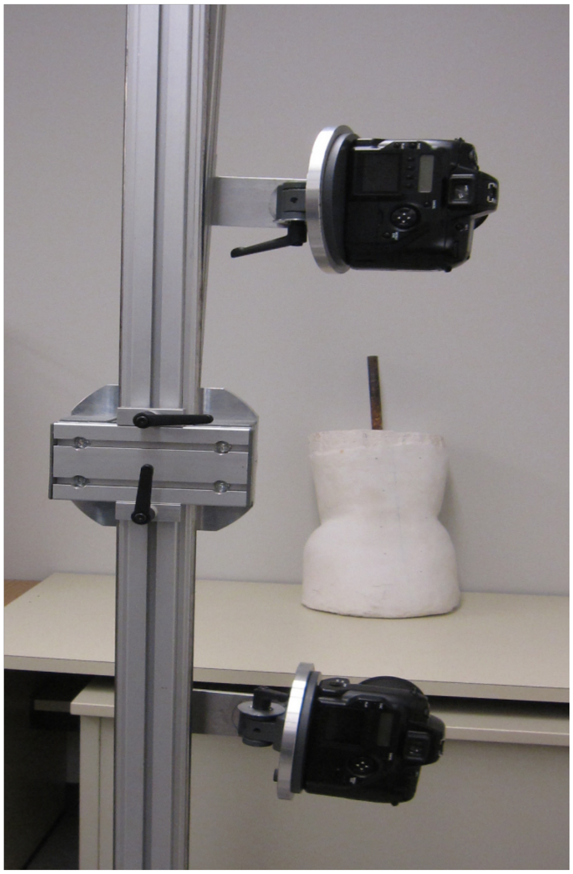

Figure 2.6 Stereo Imaging System Setup

2.3.1 Stereo Camera Setup

There have been many stereo acquisition techniques developed over the years due to progress in stereo computer vision. The most common stereo cameras used for high resolution 3D reconstruction aimed at special effects in movies and TV; restoration of 3D works of art; architectural structure design etc are digital cameras [3].

The setup of stereo digital cameras following a known geometry under the epipolar constraint is crucial for obtaining accurate 3D reconstruction. When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image's points [29]. The epipolar geometry between two views is the geometry of the intersection of the image planes between the two views having the baseline as axis. Baseline represents the distance from the optical centre of one camera to the optical centre of the other camera in the stereo setup.

2.3.1.1 Epipolar Geometry and Fundamental Matrix

Epipolar geometry makes the search for corresponding points in stereo images easier, and we will start from that objective here. Suppose a point X in 3-space is imaged in two views, at x in the first, and x′ in the second. What is the relation between the corresponding image points x and x′? As shown in figure 2.7 the image points x and x′, space point X, and camera centres are coplanar. Denote this plane as π. Clearly, the light rays back-projected from x and x′ intersect at space point X, and the rays are coplanar, lying in π. It is this latter property that is of most significance in searching for a correspondence. Supposing now that we know only x, we may ask how the corresponding point x′ is constrained. The plane π is determined by the baseline and the light ray defined by x. From above we know that the light ray corresponding to the (unknown) point x′ lies in π, hence the point x′ lies on the line of intersection l′ of π with the second image plane. This line l′ is a line in the second image plane formed by the light ray back-projected from x. In terms of a stereo correspondence algorithm the benefit is that the search for the point corresponding to x need not cover the entire image plane but can be restricted to the line l′ (figure 2.6). This restriction is called epipolar constraint. The process of establishing epipolar constraint is known as image rectification.

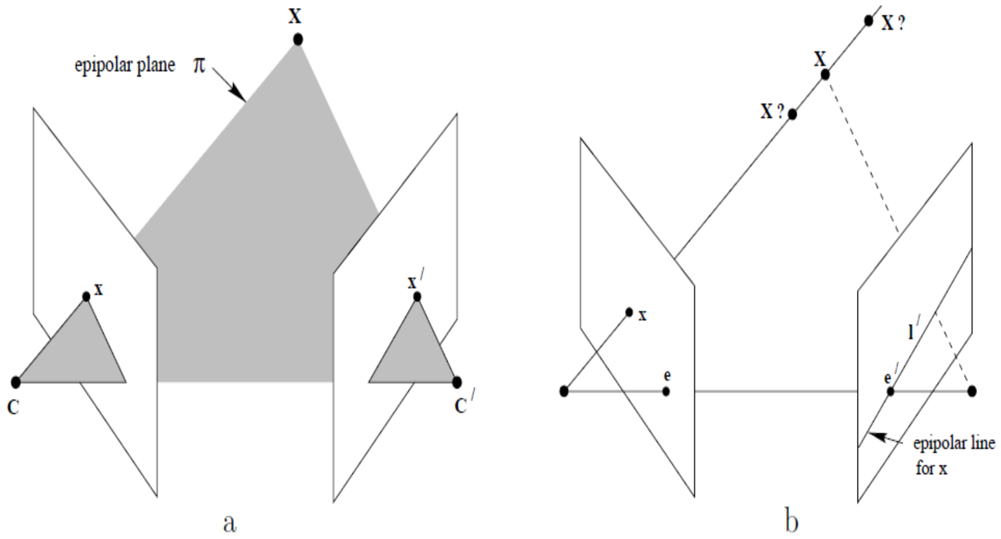

Figure 2.7 Principle of Epipolar Geometry (source: Hartley et. al., 2004)

Traditionally, digital camera calibration is used to establish the epipolar constraint. One of the most commonly used camera calibration methods is Direct Linear Transform (DLT) [30]. DLT is used to determine the internal parameters of each of the pair of digital cameras. The internal parameters of each of the cameras relate the image coordinates of the camera's image to the world coordinates of the object. Determination of internal parameters of each camera is done by manually selecting at least 6 points on the camera image. The world coordinate values of the aforementioned selected points are already known. A matrix called the essential matrix is calculated to relate the internal parameters of the two cameras. The essential matrix is expressed in terms of the intrinsic parametric matrix, rotation and translation matrices of each camera. This matrix can be used for determining both the relative position and orientation between the two cameras. Image rectification is performed using this essential matrix in order to establish the epipolar constraint. The manual selection of at least 12 points, combined with the necessity of knowing the world coordinates of these points makes this method time consuming and tedious. Therefore a method of image rectification that is automated and doesn't require knowledge of position of the imaged object is needed.

This is done by creating an accurate point to point correspondence searching method between the two images to find the required number of matches. The point to point correspondences are used to create a 3x3 fundamental matrix (similar to the essential matrix). The fundamental matrix is a matrix consisting of rotation and translation components such that when a point in one of the images is multiplied by the fundamental matrix it produces a point on the epipolar line of the other image. Image rectification is the process of finding the fundamental matrix and applying the transformation to one image in the stereo image pair to satisfy the epipolar constraint with respect to the other image. Image rectification is expressed in such a way that there is a map x → l′ from a point in one image to its corresponding epipolar line in the other image. Expressing the epipolar constraint algebraically (to infer a map between x and I’), the following equation needs to be satisfied in order for x and x’ to be matched

$$x'^{T}Fx = 0 \qquad [2\text{-}1]$$

where F is a 3x3 fundamental matrix. The role of the images can be reversed and then

$$x^{T}F^{T}x' = 0 \qquad [2\text{-}2]$$

shows the fundamental matrix as its transpose. In section 3.2, the image rectification process is explained. First, a novel template matching method of finding a collection of matching x, x’ points is shown. Then the matched set of points is used to determine the fundamental matrix.

2.3.1.2 Stereo Camera Placement

Positioning and placement of digital cameras is important part of stereo image data acquisition. The digital camera configuration must be taken into account in terms of focal length, field of view and image resolution to come up with the stereo camera setup. The placement of stereo digital cameras must be done in such a way that it facilitates point to point correspondence matching between the stereo back images. Durdle et. al. [31] investigated a structured light approach for 3D reconstruction using a stereo camera system. Even though our approach doesn't involve a projector like this structured light approach, the camera placement was constructed for torso imaging. We will use this setup as a starting point in order to arrive at our final stereo camera positioning and placement. The cameras in this setup were mounted symmetrically on a vertical bar. The base line distance between the two cameras was 1048 millimeters and the horizontal distance between the cameras and subject was about 1400 millimeters – 1500 millimeters. Convergence angles of approximately 20 degrees were set for both cameras (figure 2.8). It was determined that with these convergence angles and distances, both cameras have the same field of view in the object space and are able to capture a full image of the back [31].

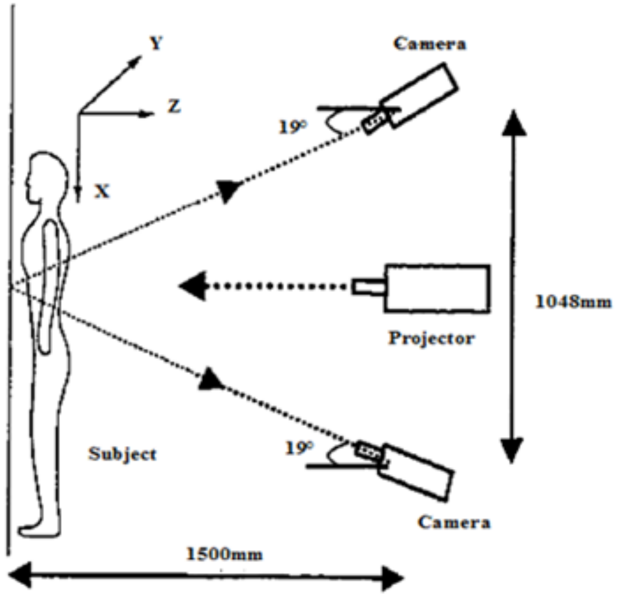

Figure 2.8 Stereo Imaging System setup (source: Durdle et. al., 1998)

2.3.2 Image Registration / Correspondence Matching

Stereo registration or correspondence matching is the process by which the closest (least error) or best point-to-point correspondence between two images is determined [16]. The result of stereo registration is pairing of points in the stereo images such that each point in the pair of points is the image of the same point in space [4]. Stereo registration / stereo correspondence is one of the most active research areas in stereo computer vision and it serves as an important step in many applications (example:-view synthesis, image based rendering etc.) [5]. Moreover, the last few years has seen a resurgence of interest in the development of highly accurate stereo registration algorithms [6]. The goal of these stereo registration algorithms is to determine disparity map from the pair of images taken with known stereo camera geometry [4, 5, 14, 16 and 20]. In the next section, disparity is defined.

There have also been great advances made in stereo registration algorithms' classification and testing as a result of publicly available Middlebury dataset [16]. A few years ago, Scharstein and Szeliski (makers of the Middlebury dataset) have provided a taxonomy and evolution of dense two-frame stereo correspondence algorithms [3]. A breakdown the basic steps common to all stereo registration algorithms is given below. .

2.3.2.1 Disparity Space Image

Disparity is often treated as synonymous to inverse depth. When first introduced to human literature, disparity was used to describe the difference in position of corresponding features as seen by left and right eye [3]. More recently, several researchers have defined disparity as a 3D projective transformation (collineation or homography) of a 3D space (X, Y, Z). If you take one of the images in the pair of images as reference image and the other image as matching image, the correspondence between pixel (x, y) in reference image r and a pixel (x`, y`) is given by

$$x` = x + s\, d(x, y);\ y` = y \qquad [2\text{-}3]$$

where s = +/- 1 is a sign chosen so that disparities are always positive.

In stereo computer vision, one of the methods used to analyze and assess stereo registration is by using Disparity Space Image (DSI). A pair of stereo images can be related to each other using a uni-valued disparity function d(x, y) of one image with respect to the other image as seen in equation above. The x, y spatial coordinates of the disparity space are taken to be coincident with pixel coordinates of the reference image. The disparity function obtained over the 2D image is known as DSI. DSI usually represents the confidence of a particular match implies by d(x, y).

2.3.2.2 Classification

There are three broad classes of stereo registration / stereo correspondence algorithms – local (window-based) algorithms, global algorithms, and segment based algorithms. For first class local (window-based) algorithms, the disparity at a given point depends only on intensity values within a finite neighboring window [4]. Local methods can easily capture accurate disparity in highly textured regions, however they often tend to produce noisy disparities in textureless regions and lead to occluded areas. The second class, global algorithms, which make explicit smoothness assumptions of the disparity map and solve it through various minimization techniques such as graph cuts and belief propagation. The third class of algorithms is the newest addition called segment-based algorithms. Based on the performance testing on the Middlebury dataset [3], this class of algorithms gives the most accurate results. Stereo based algorithms are based on the assumption that the scene structure can be approximated by a set of non-overlapping planes in the disparity space and that each plane is coincident with at least one homogenous color segment in reference image [7]. This class of algorithms distinguishes itself from global algorithms by performing color segmentation as the first step. After the color segmentation, segment based algorithms use various global energy minimizing techniques, similar to global algorithms. Though segment based algorithms have shown strong performance in conventionally difficult areas such as textureless regions, disparity discontinuous boundaries and occluded portions, they rely heavily on color changes in the image and are generally not able to handle the situation if there are disparity boundary appearing inside the color segments [10].

The four steps generally performed by the three classes of stereo registration to arrive at a disparity space image (DSI) are 1) Image segmentation, 2) Matching cost computation; 3) Cost (support) aggregation and 4) Disparity computation / optimization. The four steps and the different approaches used by each class of stereo registration algorithms are described below.

2.3.2.3 Image segmentation

Image segmentation is a broad area of research and comprises of a wide range of applications. It is used in the fields of medical imaging to locate tumors, computer aided surgery, study of anatomical structure etc., to locate objects in satellite images, for facial recognition, at traffic control systems, in stereo vision systems etc. Segmentation breaks image into groups over space and/or time. The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. Segment based stereo registration class of algorithms use image segmentation as the first step. Local (window-based) and global algorithms do not perform this step.

Image segmentation can be broadly divided into two categories: 1) top-down segmentation and 2) bottom-up segmentation. In top-down image segmentation, pixels are grouped together because they lie on the same object. In bottom-up segmentation, pixels are grouped together because of a local affinity measure based on color, intensity etc. The human back comprises of just one object and this is why top-down segmentation is ineffective. However, Bottom-up segmentation can be used to segment regions of the back based on changes in pixel intensity. The mean shift segmentation algorithm is described in detail later in section 4.1.

2.3.2.4 Matching cost computation

The most common pixel-based matching costs include squared intensity differences (SD) and absolute intensity differences (AD). Other traditional matching costs include normalized cross-correlation, which behaves similar to sum-of-squared-differences (SSD), and binary matching costs (match / no match), based on binary features such as edges. Some costs are insensitive to differences in camera gain, for example gradient based measures and non-parametric measures such as rank and census transforms. Birchfield and Tomasi [3] have proposed a matching cost that is insensitive to image sampling. Rather than just comparing pixel values shifted by integral amounts (which may miss a valid match), they compare each pixel in the reference image against a linearly interpolated function of the other image. The matching cost values over all the pixels and all the disparities form the initial disparity space image $C_0(x,y,d)$ [3]. Local, global and segment based algorithms all perform matching cost computation using one of the above approaches. However segment based algorithms perform matching cost computation as the second step, color segmentation being their first step. In local and global based algorithms, this step calculates the initial disparity values.

2.3.2.5 Cost (support) aggregation

Local, window-based and segment-based methods aggregate the matching cost by summing or averaging over a support region in the DSI C(x,y,d). A support region is a two dimensional (2D) window at a fixed disparity or 3D window in x-y-d space. Global algorithms do not perform an aggregation step, but rather seek a disparity assignment that minimizes a global cost function that combines data and smoothness terms. A 2D support region favors frontal parallel assumption while 3D support region supports slanted surface.

Frontal parallel plane assumption means that position disparity is constant over the support region. However, real world objects possess surfaces rich in shape, for example the torso, which violates this assumption. 2D cost aggregation has been implemented using square windows or Gaussian convolution, windows with adaptive sizes and windows based on connected components of constant disparity. 3D support functions that have been proposed include limited disparity difference, limited disparity gradient and Prazdny's coherence principle [3]. These functions address frontal parallel plane assumption problem by using a parameterized planar or quadratic patch fit to the images as a local model for the disparity surface. Disparity derivatives are used to deform the matching window in a refined correlation algorithm [25]. A different method of aggregation is iterative diffusion. It works well for curved surfaces. In this method, the aggregation operation is implemented by repeatedly adding to each pixel's cost the weighted values of its neighboring pixels' costs.

The formula for aggregation performed using 2D or 3D convolution is given by

$$C(x, y, d) = w(x, y, d) * C_0(x, y, d) \qquad [\ 2\text{-}4]$$

where w(x, y, d) represents the window function and $C_0(x, y, d)$ represents the initial DSI.

2.3.2.6 Disparity computation / optimization

In local (window-based) algorithms, the emphasis is on matching cost computation and on cost aggregation steps. Computing the final disparities is trivial: simply choose at each pixel the disparity associated with the minimum cost value. Thus, these methods perform a local "winner-take-all" (WTA) optimization at the each pixel. A limitation of this approach (and many other correspondence algorithms) is that uniqueness of matches is only enforced for one image (the reference image), while points in the other image might get matched to multiple points.

In global and segment-based algorithms, most of the work is performed in the disparity computation phase. Many global methods are formulated in an energy-minimization framework. The objective is to find a disparity function d(x, y) that minimizes a global energy,

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d). \qquad [2\text{-}5]$$

It involves minimizing two separate energy functions that are summed together to calculate the final energy minimization term as given by the equation (2-5) where d represents disparity. The symbol Edata(d) in equation (2-5) measures how well the disparity function agrees with the input image pair and is given by the summation of matching score over the spatial coordinates. The formulation of Edata(d) follows in equation (2-6) where C is the matching score.

$$E_{data}(d) = \Sigma\, C(x, y, d(x, y)) . \qquad [2\text{-}6]$$

The symbol $E_{smooth}(d)$ in equation (2-5) encodes smoothness in the image by measuring the differences between the neighboring pixels' disparities [3]. $E_{smooth}(d)$ can be described by equation (2-7) where ρ is some monotonically increasing function of disparity difference.

$$E_{smooth(}d) = \Sigma\, \rho(d(x, y) - d(x+1, y)) + \rho(d(x, y) - d(x, y+1)) . \qquad [2\text{-}7]$$

Geman and Geman's seminal paper [26] gave a Bayesian interpretation of a discontinuity-preserving robust ρ function based on Markov Random Fields (MRFs) and additional line processes. Based on the Middlebury dataset, Belief Propagation energy minimization framework is considered to be the best performing and is based on MRFs. MRFs are described in the following section. $E_{smooth(}d)$ can also be made to depend on intensity differences based on the equation below

$$\rho_d(d(x, y) - d(x+1, y)) * \rho_I(|| I(x,y) - I(x+1,y) ||) \qquad [2\text{-}8]$$

where ρ_I is some monotonically decreasing function of intensity differences that lowers smoothness costs at high intensity gradients. This idea encourages disparity discontinuities to coincide with intensity/color edges and appears to account for some good performance.

2.3.2.6.1 Markov Random Fields

As defined by Freeman et. al., for a given image data, y, the underlying scene, x is estimated. First, the posterior probability, P(x | y) = c * P(x, y) is calculated. The Markov network topology implies that knowing the scene at position j :- 1) provides all the information about the rendered image there, because x_j has the only link to y_j and 2) gives information about nearby scene values, by links from x_j to nearby scene neighbors. Under two common loss functions, the best scene estimate is the mean (minimum mean squared error, MMSE) or the mode (maximum a posteriori, MAP) of the posterior probability. For a Markov random field (MRF), the joint probability over the scenes x and the image y can be written as

$$P\left(x_1, x_2, \ldots., x_{N,}\, y_1, y_2, \ldots., y_N\right) = \prod_{(i,j)} \left(\psi\left(x_i, x_j\right)\right) * \prod_k \left(\Phi(x_k, y_k)\right) \qquad [2\text{-}9]$$

where ψ and Φ are pairwise compatibility functions which are learned from training data. (i, j) indicates neighboring nodes i, j and N is the number of images and scene nodes. x_n is the variable at location n, and y_n is the variable representing differences. MMSE estimate of each x_i is the mean of the marginal distribution of x_i. The MAP estimate is the labeling of x_1…. x_N that maximizes equation (2-9). [17] This can be indicated as

$$\widehat{x_j}\ MAP = \arg x_j \max max_{[all\ x_{i,i\neq j}]} *\ P\left(x_1, x_2, \ldots., x_{N,}\, y_1, y_2, \ldots., y_N\right)$$

[2-10]

Equation (2-10) above is infeasible to evaluate directly because of high dimensionality of scene variables over which $P\left(x_1, x_2, \ldots., x_{N,}\, y_1, y_2, \ldots., y_N\right)$ must be maximized. Described in section 4.4 is an approximate MAP-MRF inference algorithm called Belief Propagation to solve equation (2-10).

Now, if the log of equation (2-10) is taken, finding the MAP estimate is equivalent to minimizing a function of the form

$$E\left(x_1, x_2, \ldots., x_{N,}\, y_1, y_2, \ldots., y_N\right) = \sum_{(i,j)} -\log\psi\left(x_i, x_j\right) + \sum_{(k)} -\log\Phi(x_k, y_k)$$

[2-11]

The above equation is the same as equation (2-5). Therefore, it is concluded that maximizing the probability of pixel correspondence in the image pair is equivalent to minimizing the global energy in an energy minimization framework. So, Belief propagation can be used to minimize global energy.

However, the data term E_{data}(d) / $\Phi(x_k, y_k)$ and the smoothness term $E_{smooth(}$d) / $\psi(x_i, x_j)$ have to be altered for curved, slanted or smooth surfaces since it implicitly makes frontal parallel assumption. Once the global energy has been defined, a variety of algorithms can be used to find a local minimum. A novel approach using differential geometry, Belief propagation and local minimum finding algorithm is described in Chapter 4. Once a local minimum is computed for the entire image based on a reference image, the minimized global energy E(d) equation (2-5) is obtained. The resulting disparity function d(x,y) is used in 3D image reconstruction.

2.3.3 3D Image reconstruction

This is the last step of the image reconstruction process. z spatial coordinate values using the above-calculated DSI at a known x, y spatial coordinates is calculated. Baseline and focal length of the cameras are also used in this equation to obtain z coordinates as described in Chapter 5. The x, y, z coordinates are triangulated to obtain a wireframe of the 3D image. There are stray data points as well as holes and occlusions in the image. Also, there are several portions of the 3D image that do not represent the torso that need to be removed. Pre-processing the 3D image is done to remove stray data points and fill in occlusions. Sutherland Hodgman Clipping algorithm is used to manually remove regions that do not represent the torso. Now, to fill in the occlusions a novel approach combining Bezier Curve and Moving Least Squares is developed. This described in detail in Chapter 5.

2.3.3.1 Bezier Curve theory

The Bezier Curve (BC) theory is extensively used in model building and computer graphics [28]. The classical BC is a recursive linear weighted subdivision of the edges of the generated polygon starting with a set of points that form the *control (initial) polygon* (CP) and ending when the final point is generated for a particular weight t. The set of $N+1$ starting points is referred to as the *control* points which determine the shape of the BC of degree N. Therefore, for an ordered set of points $P = \{p_0, p_1, ..., p_N\}$, the matrix form of the classical BC is defined as

$$p(t) = Pow^N(t) * Bez^N * P^T \qquad [2\text{-}12]$$

where $p(t)$ is the BC point for a particular t, $Pow^N(t)$ represents the power basis $(1, t, t^2, ..., t^N)$ and the ij^{th} term of matrix Bez^N is found from $m_{ij} = (-1)^{j-i}\binom{N}{i}\binom{i}{j}$. t is the *parametric operator* which defines the location of the curve point, with the number of curve points depending upon the number of t values [29]. Simply put from the ordered set of points P, let the rectangular coordinates of p_i be (p_{ix}, p_{iy}, p_{iz}) where $i = 0,1,....,N$. Then the parametric equation of such a BC is given as

$$p(t) = \sum_{i=0}^{N} \binom{N}{i} (1-t)^{N-i} t^i p_i, 0 \le t \le 1 \qquad [2\text{-}13]$$

where $p(t) = (x(t), y(t), z(t))$ [28]. This parametric equation is used to construct intermediate Bezier curve points represented by x in figure 2.9 using four control points represented by o.

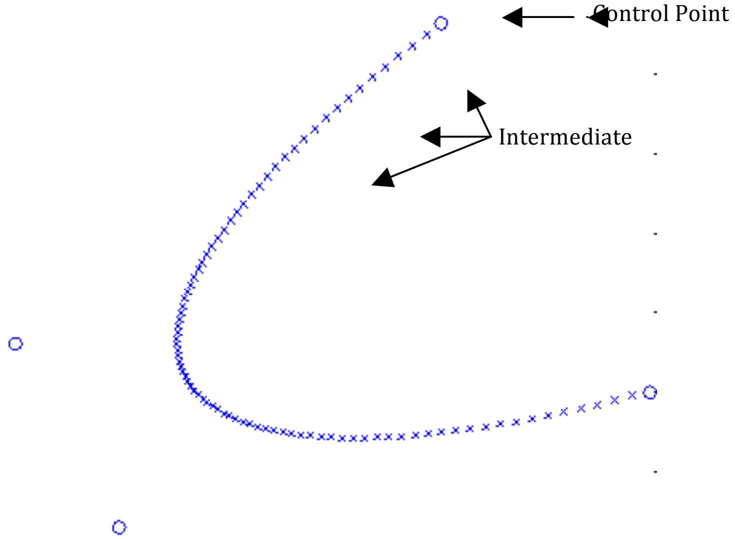

Figure 2.9: Illustration of the Bezier Curve Theory

It can be noted from the above equation that a point is generated by blending all the control points thereby implying that the BC considers the global information of a shape and yields a gap between the curve and its CP [29]. The number and location of the generated points completely depend on the values of$t$. t ranges from 0 to 1, therefore the procedure might either lead to redundant, overlapping points or insufficient points necessary for describing the shape, unless the values of t are very carefully chosen. These problems lead to a significant shape distortion [29] when only BC is used to interpolate a shape. In order to reduce this error, a procedure for using an appropriate number of points to represent shape as well as local approximation is implemented in the form of MLS projection procedure [30].

2.3.3.2 Moving Least Squares Approximation

According to the Moving Least Squares (MLS) projection procedure, given a data set of points $P = \{p_i\}$, a smooth *MLS surface* S_P based on the input points is defined. Usually, the points P defining S_P are replaced by a reduced set $R = \{r_i\}$ defining an *MLS surface* S_R that approximates S_P. The typically lighter point set R are called *representation points* [31]. This technique provides an important property of a *smooth manifold surface* (surface defined by the point set is guaranteed to be 2-manifold and C^∞ smooth, given that points are sufficiently close to the surface representation) as shown in figure 2.10 [30].

Let points $p_i \in R^3, i \in \{1,...,N\}$, be sampled from a surface S. The goal is to project a point $r \in R^3$ near S onto a two-dimensional surface S_p that approximates the p_i.

The first step involves computing a local reference plane H for r where

$$H = \{x | \langle n, x \rangle - D = 0, x \in R^3 \}, n \in R^3, \|n\| = 1 \qquad [2\text{-}14]$$

This is computed to minimize a local weighted sum of square Euclidean distances of points p_i to H. Assume q is the projection of r onto H and let $q = r + tn$ for some $t \in R$, then H is found by minimizing

$$\sum_{i=1}^{N} \langle n, p_i - r - tn \rangle^2 \theta(\|p_i - r - tn\|) \qquad [2\text{-}15]$$

where θ is a smooth, radial, monotonic decreasing function, which is positive on the whole space. This equation to find local reference plane H is an implementation of the MLS approximation theory described above. The approximation of single points is dictated by the radial weight function θ which as suggested by Levin is a Gaussian function such that $\theta(d) = e^{-\frac{d^2}{h^2}}$ [32] where h is a fixed parameter reflecting the anticipated spacing between neighbouring points [30].

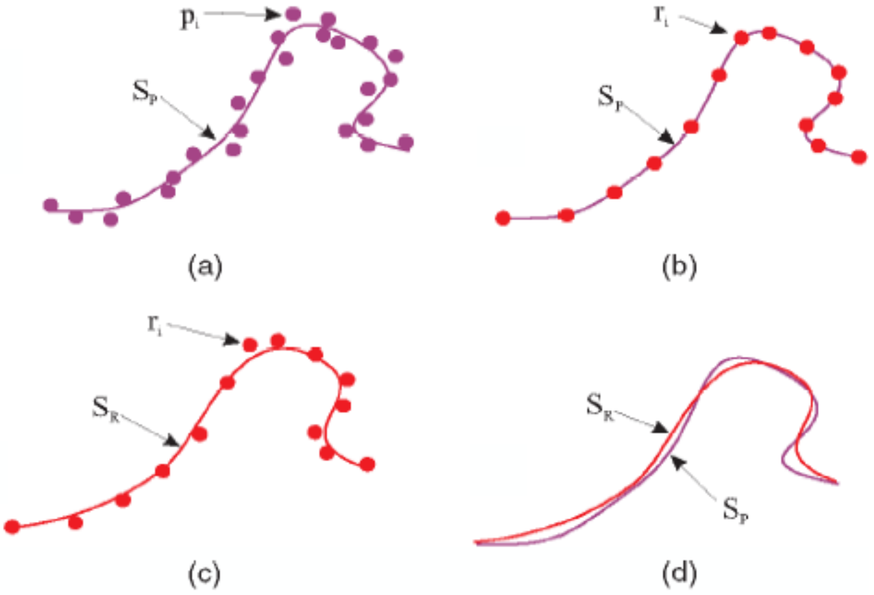

Figure 2.10: Illustration of the MLS projection procedure. (Source: ALEXA, et al., 2003)

The minimization function to compute H usually has more than one local minimum. Since H should be close to r, the local minimum with the smallest t is chosen· Past work [30] have employed a standard iterative solver to ensure that the minimization function converges to a local minimum with a small t. The initial value for n is computed by setting t in the minimization function to zero, and equating the gradient of this new quadratic function in n to zero. Thus, when $t = 0$; the minimization function becomes

$$\sum_{i=1}^{N} \langle n, p_i - r \rangle^2 \theta(\|p_i - r\|); \text{ and } \sum_{i=1}^{N} 2\langle n, p_i - r \rangle \, \theta(\|p_i - r\|)(p_i - r) = 0 \qquad [2\text{-}16]$$

The computed initial value for n can be optionally refined using Powell's iteration [33]. Now, n is substituted in the minimization function and using an iterative procedure of increasing t from 0, is used to establish a local minimum after which the subsequent t is selected. The global minimum of the minimization function is reached for $t \to \infty$, to avoid this, the function is normalized using the sum of weights θ.

The second step is used to compute a local bivariate polynomial approximation g to the surface S_p in a neighborhood of r from the computed local reference plane H and the radial weights $\theta(\|p_i - q\|)$. Let q_i be the projection of p_i onto H, and f_i be the height of p_i over H, i.e. $f_i = n.(p_i - q)$. Another minimization function is constructed based on MLS approximation theory to compute the coefficients of g thereby minimizing the weighted least squares error

$$\sum_{i=1}^{N} \left(g(x_i, y_i) - f_i\right)^2 \theta(\|p_i - q\|) \qquad [2\text{-}17]$$

where (x_i, y_i) is a representation of q_i in the local coordinate system in H. The gradient of the above equation is calculated in a way similar to the first step leading to a system of k equations, where k is the number of coefficients.

Finally, the projection of r onto S_p which is the result of the MLS projection procedure is given by the polynomial value at the origin, that is $q + g(0,0)n$ [30].

2.3.4 Summary of Literature Review

The process of creating 3D image from a pair of digital stereo images using computer vision has been broken down into 3 stages: 1) Stereo Image Acquisition, 2) Image registration and 3) 3D Image reconstruction. A number of existing methods in each of the stages has been reviewed. The purpose of this thesis is to construct a unified and improved approach for 3D torso surface reconstruction from a pair of digital stereo back images.

Existing methods of stereo camera setup and stereo image data acquisition were investigated. Manual calibration of cameras using Direct Linear Transform was found to be popular. Calibration is time consuming and prone to errors. Therefore, an automated method for image rectification needs to be developed.

Segment-based registration algorithms yield the best results for image registration based on the review. However a problem associated with the existing segment-based stereo registration algorithms such as that developed by Klauss, Sormann and Karner is that it assumes frontal parallel plane geometry. This means that it assumes depth is constant (with respect to rectified stereo pair) over a region under consideration [3]. Reconstruction of smooth and curved surfaces where depth is constantly changing violates this assumption. Li and Zucker [8] have tried to solve this problem using differential geometry but they have not adapted their algorithm to be used with that of Klauss, Sormann and Karner. We therefore, need to investigate a method of applying diffential geometry to each stage of segment based algorithms and test it against smooth surfaces like the back.

Finally, stereo reconstruction methods are explored. Since, the registration process leaves stray points due to registration errors, these need to be removed. Methods for filling occluded regions using various interpolation methods were also explored. The existing hole filling methods need improvement because their interpolated points did not match the shape of the torso. Therefore, we need to experiment with methods that can solve this problem. Triangulation to connect the 3D points into polygons and texture application is the last step to obtain a reconstructed 3D torso.

3. Stereo Image Data Acquisition

An accurate, quantitative, reproducible and repeatable 3D torso reconstruction is required, so that it can be used clinically. A robust technique for acquiring a pair of digital camera images is therefore required. In the digital camera configuration section, the general scene structure is discussed; then the focal length of the cameras are calculated and the choice of image resolution discussed; finally, the distance between points in the resultant image is calculated. Next, image rectification is described in detail. Image rectification is the process of establishing epipolar constraint between two images. The creation of an automated process of image rectification is introduced. This process is superior to the slow and tedious process of manual camera calibration (which is generally used to rectify images). The process of image rectification significantly reduces the complexity and the time required for stereo registration. Next, various stereo digital camera positioning and placements for acquiring stereo images are described. Finally, the choice of a particular digital camera setup is discussed. The 3 stages of stereo image data acquisition are shown in figure 3.1.

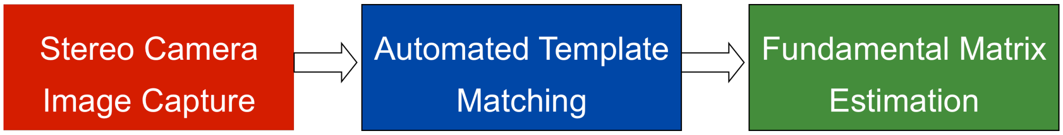
Stereo Camera Image Capture
Automated Template Matching
Fundamental Matrix Estimation

Figure 3.1 Stages of Stereo Image Data Acquisition

3.1 Digital Camera Configuration

Figure 3.2 shows a typical scene describing image capture by a single camera. There is a light source, an object and the camera.

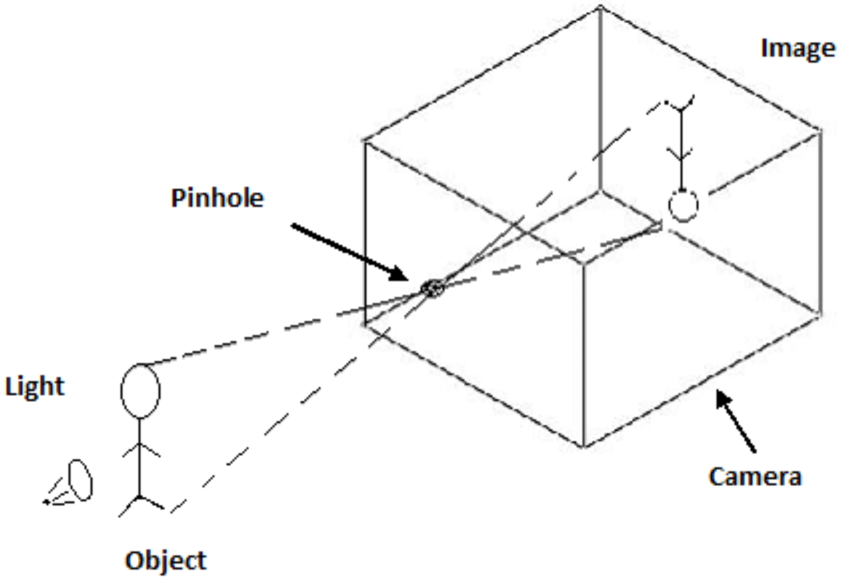

Figure 3.2 Digital Camera Image Capture

The sensor element of the digital camera (shown in the figure 3.3) is a 2D electromagnetic sensor array responsible for capturing the image and transmitting it as electrical signals to the frame grabber. In modern digital cameras, the sensor element is either a charge-coupled device (CCD) or Complementary metal–oxide–semiconductor (CMOS). The frame grabber digitizes the signals and sends it to the digital processor where it is converted to a 2D image of the object in the scene. This 2D image is formed of a 2D array of pixels. The pixels contain color and intensity which describes the image. The number of pixels on the width of the image multiplied by the number of pixels on the length of the image is called image resolution.

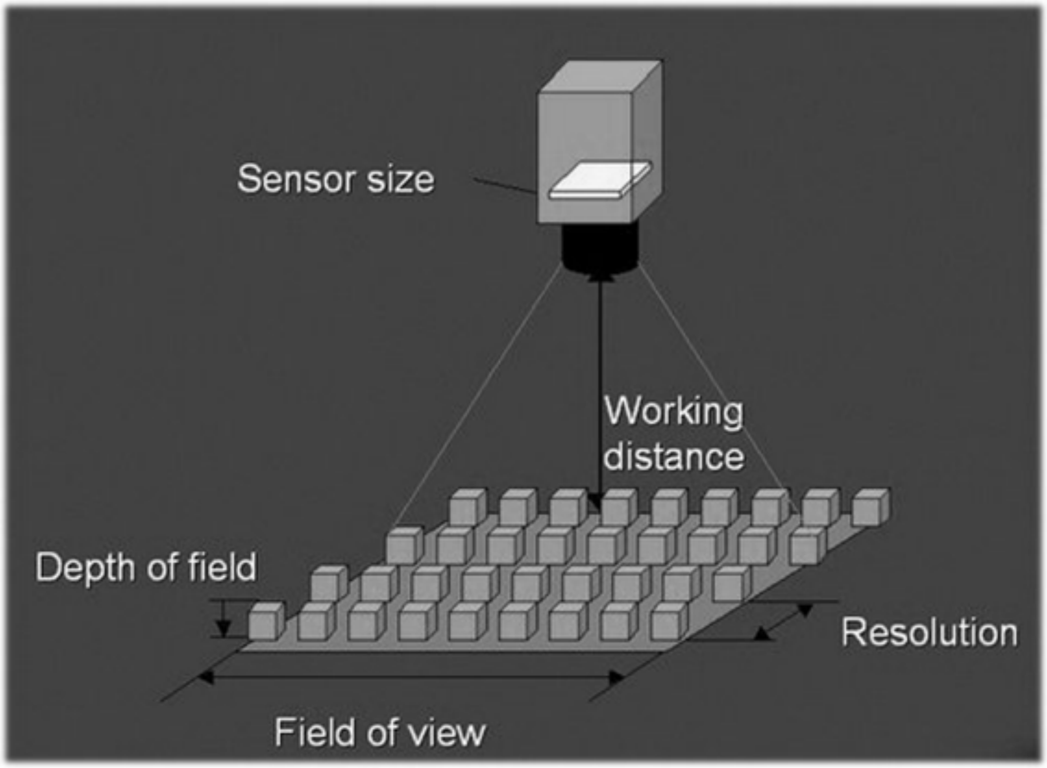

Figure 3.3 Sensor Array (magnified) of Digital Camera

The simplest camera model used to describe the process by which the lens of the camera focuses the rays of light from an object onto the sensor element of the camera is the pinhole model (figure 3.4). No digital cameras conforms perfectly to the pinhole model as the lenses distort the rays of light passing through them. But, for the purposes of focal length calculations, the distortions were ignored and the rays were assumed to converge perfectly on the sensor element of the camera. Using a geometric theorem called similar triangles; the equation for focal length of the camera is given below.

$$\text{Focal length of height} = \frac{Object\ distance\ from\ camera * Sensor\ element\ height}{Object\ height + Sensor\ element\ height}$$

and

$$\text{Focal length of width} = \frac{Object\ distance\ from\ camera * Sensor\ element\ width}{Object\ width + Sensor\ element\ width}$$

[3-1]

Focal length is the distance from lens centre to image plane. A Canon ® SLR digital camera with aspect ratio 3:2 was used.

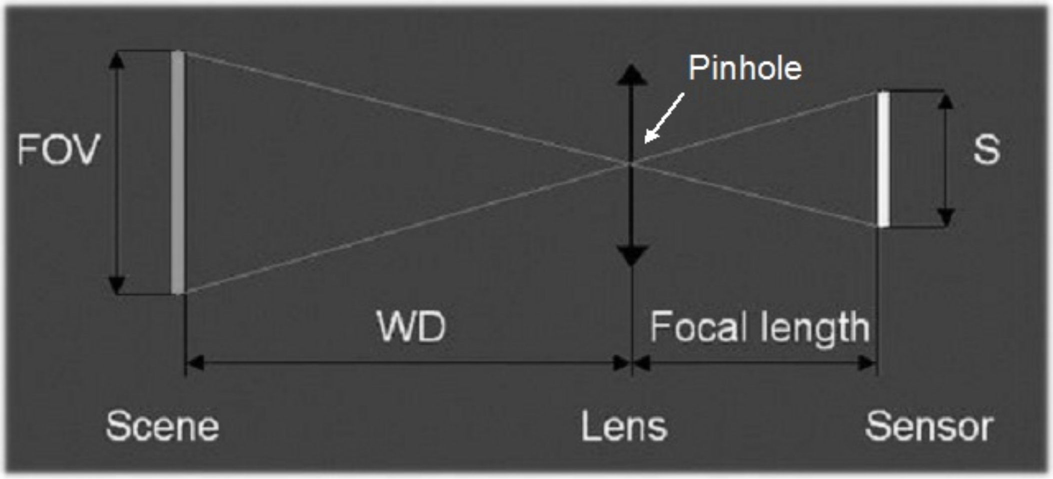

Figure 3.4 Representation of Pinhole Camera Model

(FOV: Field of View, WD: Working Distance, S: Subject Image)

The object height for torso reconstruction is the region above the waistline in a human subject. The object height was determined to be 750 mm on an average. The object width for torso reconstruction in our case is generally the highest when measured between two shoulders in the human subject. The object width is approximately 500 mm on an average. The sensor element width and height can be found from the documentation of the digital camera. A Canon ® SLR with a sensor element height of 15.1 mm and width of 22.7 mm was used. The distance of object from the lens of the camera used for experimentation were 500 mm, 1000 mm and 1250 mm. 500 mm was chosen as the shortest distance because if the cameras were placed any closer to the torso, the images do not contain the complete torso. If the object was placed further than 1250 mm away from the cameras, the cameras have to be inclined such that the angles of incidence of rays from the camera hitting are very high. This causes major errors in image transformation when image rectification is performed. The choice of using 1000 mm as the distance of object from camera is described in detail in section 3.4. The focal length of the height and the width was calculated by using object distance from camera, sensor element height and width on equation (3-1). Using Pythagoras theorem, the focal length of the stereo cameras was determined. The focal length when rounded off to the focal length settings available on the Canon ® SLR is 18 mm. This constant focal length was used during all torso stereo image data acquisition. A camera with 1.4 million pixels (1440x960) resolution for acquiring torso images was used. Through experimentation, it was determined that using any camera setup about 40% of the pixels are lost due to errors during image acquisition, registration and reconstruction or because they are not part of the torso. This leaves us with around 825000 which can be converted into 3D points. The maximum resolution on our cameras is 6.1 million pixels (3040x2016). 1.4 million pixels resolution was found to be the best tradeoff between producing high quality 3D torso images and image processing time. This is because most range scanning systems don't contain more than 500,000 pixels and produce industry standard 3D reconstruction. So, a 1440x960 resolution contains enough 3D points to reconstruct 3D back images.

By convention, the index values of the pixels start from (0,0) in the top left corner of the image, and increase in left to right and top to bottom direction respectively. i represents the indexes in the left to right direction and j in the top to bottom direction in this thesis. As mentioned previously, the maximum value for i is 1440 and j is 960. The pixel distance in metric measurements also start from (0,0) at the top left of the image and increase in the same direction as the index values. They are represented by x, y in this thesis. This distance between any two pixels was calculated in an image by using sensor dimension of 22.7 mm X 15.1 mm, image resolution of 1440 X 960 and Pythagoras theorem. The diagonal sensor dimension is $\sqrt[2]{(22.7)^2 + (15.1)^2}$ = 27.26 and similarly the number of pixels on the diagonal of the image is 1730.7. So, the distance between each pixel of the image is 0.01575 mm. Therefore x, y increment by 0.01575 when moving from left to right and top to bottom respectively. When the pixel is converted to 3D point after image registration, the x.y increments between corresponding points are still 0.01575 mm

The geometry between the object and the camera lens, the image resolution of the camera, distance between points in the image and geometry between the cameras are key aspects of the stereo camera setup. We have determined the distance between points in the image and the image resolution of the camera. The geometry between the two cameras was determined by calculating a 3x3 fundamental matrix. The fundamental matrix and epipolar geometry have been discussed in the literature review. The best placement of stereo cameras was determined with respect to the object (torso).

3.2 Image Rectification

Image rectification is the process of applying the fundamental matrix to the top image (reference image) in order to establish the epipolar constraint with respect to the bottom image. Image rectification can be divided into two stages: (1) calculate the 3x3 fundamental matrix (F in equation) and (2) multiply all the points in the top image with the fundamental matrix to establish the epipolar constraint between top and bottom images. The second stage is trivial. The first stage is described in detail below.

The aim of first stage is to solve the equation ($x'^{T}Fx = 0$). x' is a 2D point in the top image (reference image) and can be written as (x', y', 1). x is a 2D point in the bottom image and can be written as (x, y, 1). F is a 3x3 fundamental matrix. Specifically, the equation corresponding to a pair of points is

$$x'xf_{11} + xy'f_{21} + xf_{31} + yx'f_{12} + yy'f_{22} + yf_{32} + x'f_{13} + y'f_{23} + f_{33} = 0 \qquad [3\text{-}2]$$

where f is an element in the fundamental matrix and the subscript on f represents the corresponding row and column index of each element in the fundamental matrix. To calculate the fundamental matrix, the above equation for f was solved. An important fact about the fundamental matrix is that it is singular, in fact of rank 2. In order to find a linear solution for f; at least 8 equations need to be solved. This implies that at least 8 pairs of matching points have to be found in the top and bottom images. The Hartley's 8-Point algorithm [32] to used calculate the fundamental matrix [36].

The equation is written of the form

$$A^{1x9} F^{9x1} = 0 \qquad [3\text{-}2]$$

where $A = (x'x \quad xy' \quad x \quad yx' \quad yy' \quad y \quad x' \quad y \quad 1)$ and

$F = (f_{11} \quad f_{21} \quad f_{31} \quad f_{12} \quad f_{22} \quad f_{32} \quad f_{13} \quad f_{23} \quad f_{33})^T$. A linear set of equations of the form A^{8x9} $F^{9x1} = 0$ to represent the 8 linear equations was built. Finding 8 corresponding matching 2D points determined all the values in matrix A

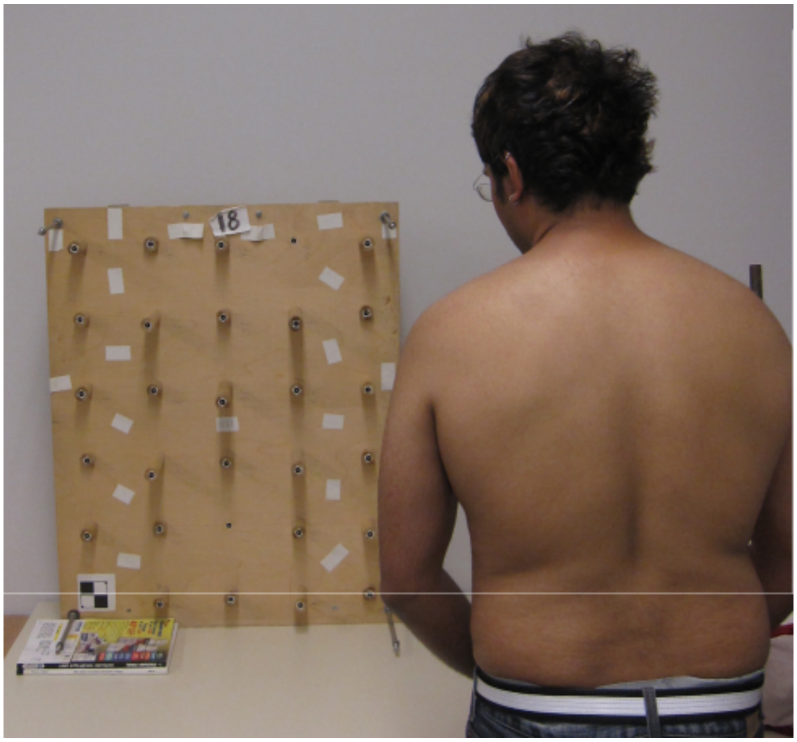

Figure 3.5(a) Image from the top camera in the stereo imaging system

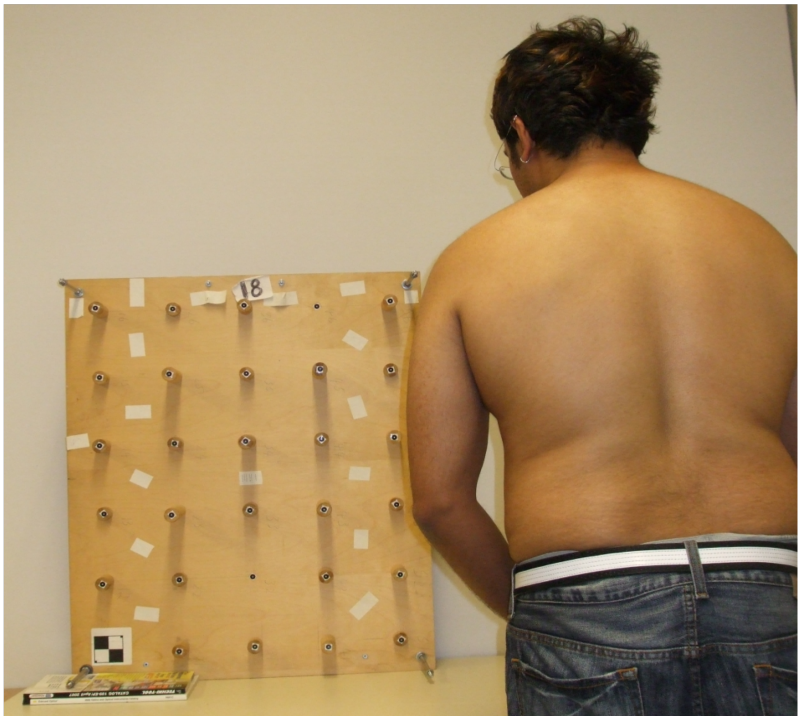

Figure 3.5(b) Image from the bottom camera of stereo imaging system

3.3 Automated Template Matching

A peg board (figure 3.5) in front of the pair of stereo cameras along with the subject to be imaged was placed. This technique of stereo camera setup is used because it increases efficiency by eliminating the need to acquire two or more images (one or more for calibration and one for subject image capture) and reduces error that may be caused due to apparatus movement between image rectification and image capture. A novel system of template matching is developed as follows. Each peg consists of a pattern of concentric circles (figure 3.6).

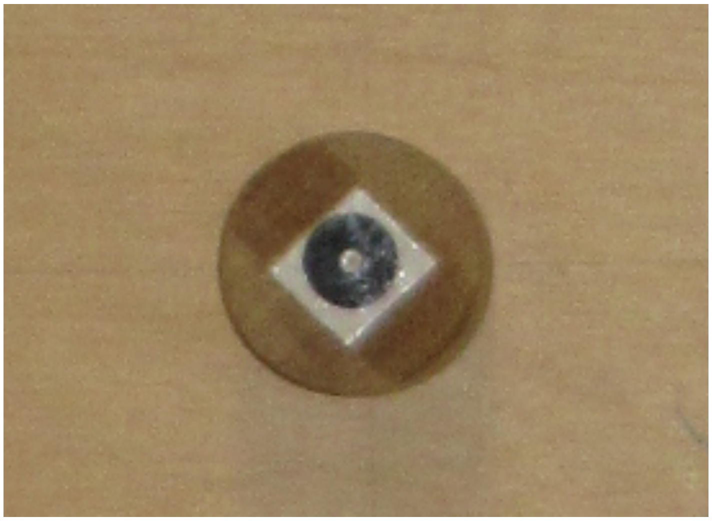

Figure 3.6 Image of a single peg from pegboard

The image of any such peg is called a template. Our aim is to find the position of 8 pegs in the top and bottom image of the peg board in either the horizontal or the vertical fashion. The position of the best match between our template and a peg on the peg board image is a 2D point in the A matrix. To find the best match, the differential localized pixel matching cost of the template (reference image) with respect to the peg board image was calculated. When the matching cost dropped to a minimum and started increasing again, the best match was found (figure 3.7) [36]. This process for the peg board image from the bottom camera in the stereo imaging system was repeated. The process of differential localized pixel matching has been described in detail later in section 4.2. After all values of A are populated, we look for a least-squares solution to equation 3-2 inorder to calculate fundamental matrix F.

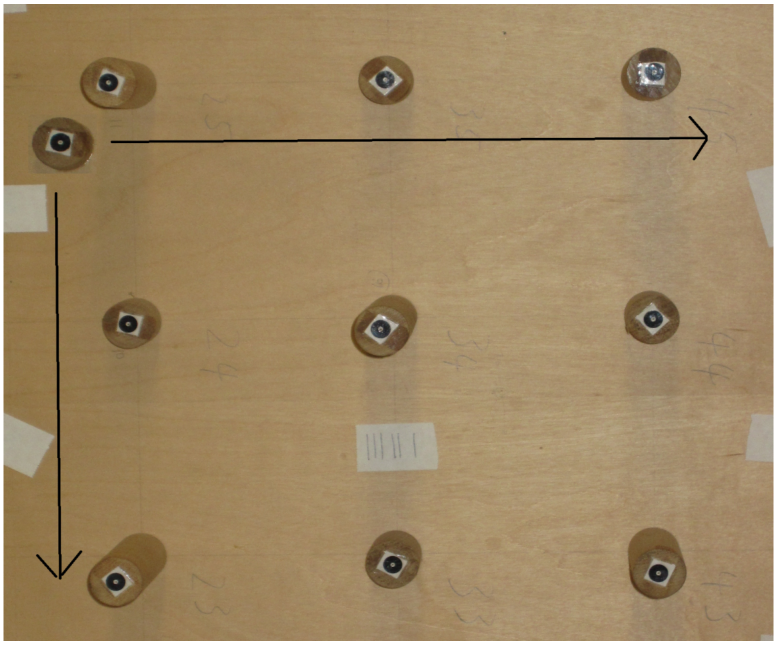

Figure 3.7 Magnified pegboard image showing peg template matching against the pegs

3.4 Stereo Digital Camera Setup

The stereo digital camera system comprises of a pair of digital cameras placed at a known distance from the object under a known geometry. The distance between the digital cameras, the angle of convergence of each digital camera and the distance between the digital cameras and the object (figure 3.8) form the geometry of the stereo digital camera setup.

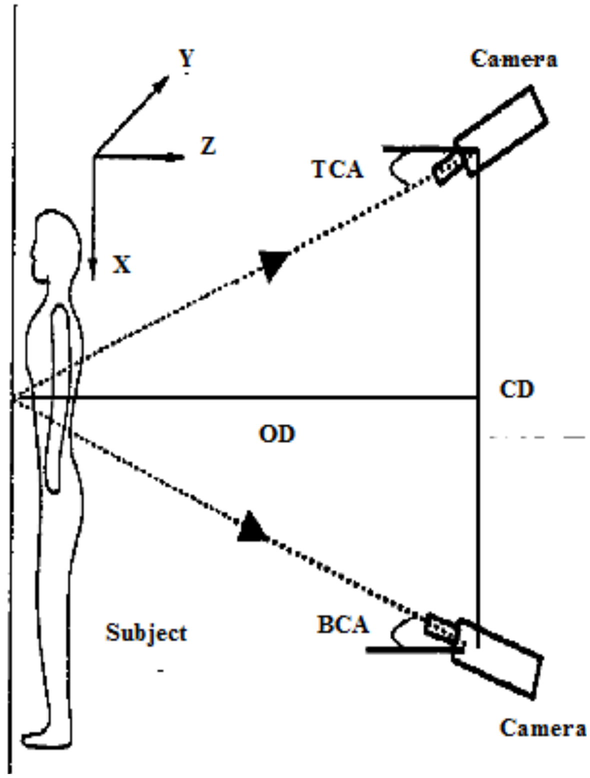

Figure 3.8. Stereo Digital Camera Geometry

(TCA: Top Convergence Angle, BCA: Bottom Convergence Angle, CD: Distance between digital cameras, OD: Distance between digital cameras and object)

The most optimal stereo digital geometry for image acquisition was determined by calculating the accuracy of 8 template matched peg points established by image rectification against known values of the respective peg points. Using the top left corner of the pegboard as origin (0,0), the x, y values of the each point formed by the concentric circles on the peg were measured. The geometry defined by Durdle et. al. [31] (section 2.3.1.2) was used as a reference and we tried 10 varying geometric setups to find the most accurate template matched 8 peg points with respect to the 8 measured values of the same peg points.

3.4.1 Demonstration of camera setups

The experiment used 10 geometric setups that are established by using different values of top convergence angle (TCA), bottom convergence angle (BCA), distance between digital cameras (CD) and distance between object and camera (OD). Figure 3.9 shows the top and bottom camera image from one of the geometric setups.

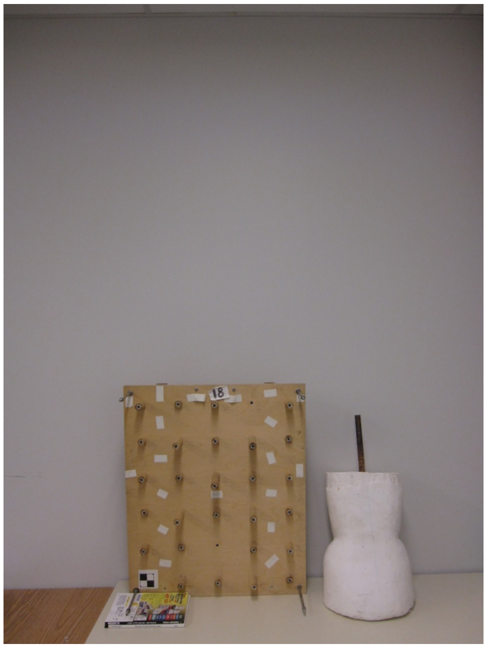

Figure 3.9 (a) Pegboard image from top camera using TCA 35° CD 100cm and OD 235cm

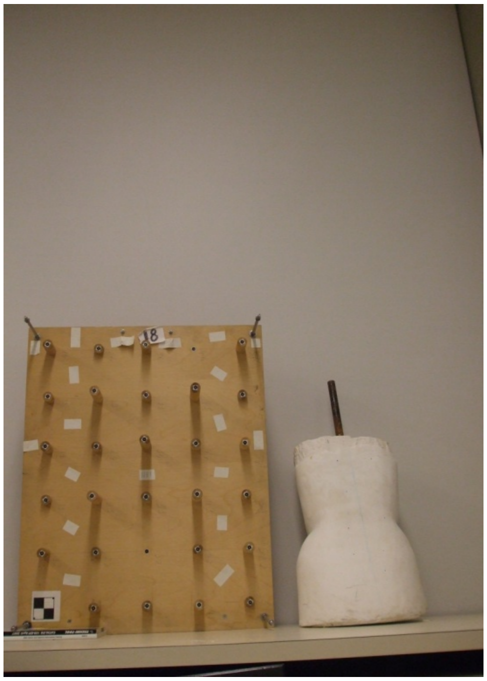

Figure 3.9 (b) Pegboard image from bottom camera using BCA 35° CD 100cm and OD 235cm

The x', y' values of the template matched peg points were compared against the known x, y values of the pegs on the pegboard, with modulus of x-x' and modulus of y-y' divided by 2 being the accuracy measurement of each point. 0 is the most accurate and sum of x and y divided by 2 being the least accurate. The least accurate template matched peg point amongst the 8 points for each geometric setup is used as the overall accuracy (in percentage) for the setup. Table 3.1 shows the various setups and their respective accuracy measurements.

	TCA (°)	BCA (°)	CD (cm)	OD(cm)	Accuracy (%)
1	0	20	19	188	65
2	0	5	19	235	74
3	5	0	19	235	74
4	5	0	44	280	72
5	19	19	100	140	95
6	35	35	100	235	81
7	16	16	100	235	98.7
8	16	16	100	280	87
9	16	16	65	280	83
10	18	18	65	150	79

Table 3.1 Peg point coordinates calculation accuracy for stereo geometries. Accuracy (%) denotes how close the calculated x', y' peg values are to the known x, y peg values.

3.4.2 Justification for using particular setup

The experimentally developed optimal stereo camera setup is shown in figure 3.10. The Field of View is higher vertically than horizontally in this setup which follows the shape of the torso. The angle of convergence in this setup is kept as minimal as possible to reduce the errors during registration.

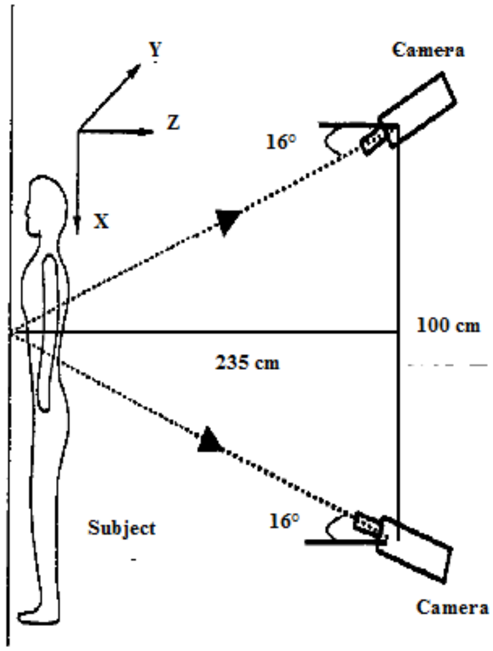

Figure 3.10 Optimal Stereo Digital Camera Setup

3.5 Achievements

In this section, an automated system for stereo image rectification was developed. We also experimented with various stereo camera setups to arrive at the most optimal setup for stereo image data acquisition.

4. Stereo Image Registration

Registration of the two digital images acquired using the top-down camera configuration following the epipolar constraint [30] is described in this chapter. The top camera image was taken as the reference image and the Disparity Space Image (DSI) on the bottom camera image was calculated. There is a major problem with existing stereo image registration algorithms which prevents us from using it in 3D Torso surface reconstruction. Many stereo algorithms either implicitly or explicitly exploit the frontal parallel assumption. This assumes disparity is constant with respect to rectified stereo pair over a region under consideration [3]. Since the back surface is smooth and curved, this assumption leads to errors. The existing "state of the art " segment based class of algorithms (as mentioned in the literature review) was used and modifications were made so that they can be used on stereo back images. As described in the following sections, a novel approach was developed combining Mean Shift segmentation with differential geometry constraints into existing segment based algorithms. The process of stereo image registration is divided into four stages: (1) Mean Shift Color Segmentation, (2) Differential Localized Pixel Matching, (3) Disparity plane definition and (4) Differential Max Product Belief Propagation (figure 4.1).

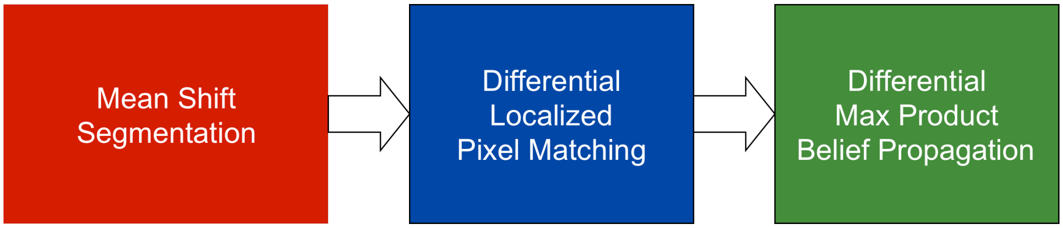

Figure 4.1 Stages of Stereo Image Registration

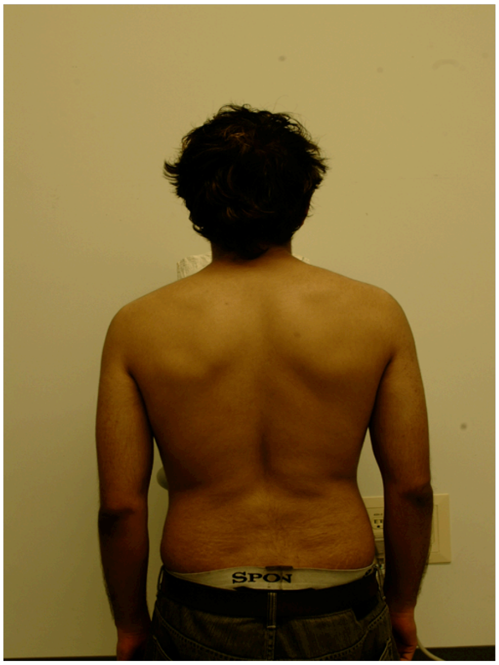

Figure 4.2(a): Top Camera Image

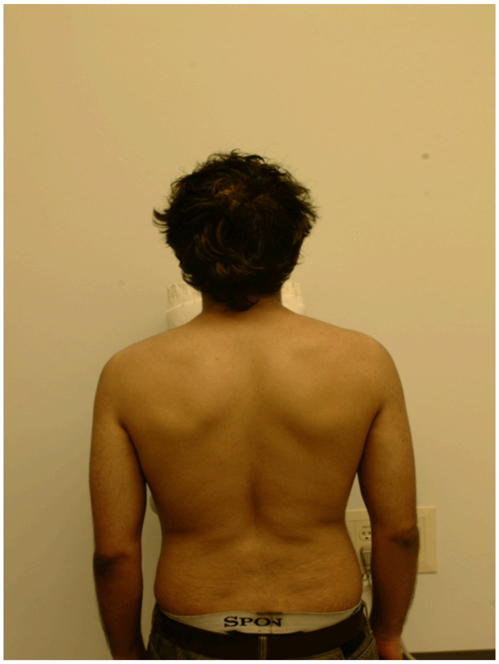

Figure 4.2(b): Bottom Camera Image

4.1 Mean Shift Segmentation

Segment-based registration methods have attracted attention due to their good performance [7]. The main objective is to reduce the high-resolution space and enforce disparity smoothness in homogenous intensity and color regions. Regions of homogenous intensity and color are located by applying a segmentation method. The process of mean shift segmentation is to decompose the image into regions of color or grayscale. The mean shift segmentation was applied to the top and bottom camera images. Comaniciu and Meer's mean shift segmentation procedure is insensitive to differences in camera gain and is therefore used [9]. Mean shift segmentation is divided into two steps: (1) Mean shift filtering of the original image data (in feature space) and (2) Mean shift clustering of the filtered data points

Mean shift filtering is defined as a gradient ascent search for maxima in a probability density function (pdf) defined over a high dimensional feature space. This step consists of analyzing a pdf underlying the image data in feature space to find the local maxima (modes of the pdf). Consider the feature space consisting of the original image data represented as the (x, y) location of each pixel, plus its colour in L*u*v* space (L*, u*, v*). The modes of the pdf underlying the data in this space will correspond to the locations with highest data density. In terms of segmentation, it is intuitive that the data points close to these high probability density points (modes) should be clustered together.

The pdf for mean shift filtering can be parametric or non-parametric. A parametric pdf can express data distribution in few parameters like mean and variance. But, this makes the pdf limited in flexibility and does not express a complex data set accurately. A Parzen window method (named after Emanuel Parzen), a non-parametric way of estimating the pdf [4] was therefore used. For data points x_1, x_2,…., x_N the kernel density approximation of pdf is given by

$$f(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right) \qquad [4\text{-}1]$$

where K is a kernel function, N represents the number of pixels/points in the image and h is the window size or bandwidth. A Gaussian kernel is used, so K is given by

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2} \; and \; K\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{h^2}} \qquad [4\text{-}2]$$

also K(x) profile is given by $k(x) = e^{-x}$ [4-3]

The bandwidth (kernel window size), represented by h in equation [4-2], is split into two components one representing the spatial (x,y) domain (h_s) and the other representing range(L*, u*, v*) domain (h_r). Using equation [4-2] and [4-3],

$$K(x) = c_k k(\|x\|^2) \qquad [4\text{-}4]$$

where c_k is a normalization constant equal to $\frac{1}{\sqrt{2\pi}}$. In order to split the kernel into spatial-range domain K(x) is defined in terms of its profile as

$$K(x) = \frac{c}{h_s h_r} \, k\left(\left\|\frac{x_s}{h_s}\right\|^2\right) k\left(\left\|\frac{x_r}{h_r}\right\|^2\right) \qquad [4\text{-}5]$$

The process of mean shift filtering can be divided into the following steps. First, choose a search window size (bandwidth) in the spatial-range domain. Second, choose the initial location of the search window. Third, compute the mean location (centroid of the data) in the search window i.e. find the local maxima of the search window. Fourth, center the search window at the mean location computed in step three. Finally, repeat steps 3 and 4 until convergence (figure 4.3).

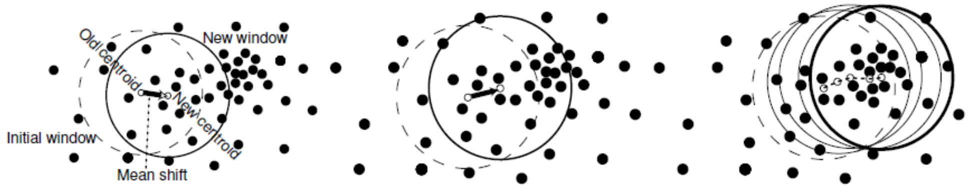

Figure 4.3 Illustration of 4 steps of mean shift filtering

Over segmentation was preferred because several clustered segments can be later combined to deduce a set of disparity planes to obtain DSI. ($h_{s,}$ h_r) = (75, 0.3) was found to be a suitable comprise between over segmentation and processing time through several trials. To determine the local maxima for the window, the gradient of the kernel density approximation (equation [4-1]) should be zero. The gradient is estimated as being proportional to the mean shift vector:

$$f(x)\ \alpha\ \frac{\sum_{i=1}^{N} x_i\, g_i}{\sum_{i=1}^{N} g_i} - x \qquad [4\text{-}6]$$

where x_i are the data points, x is a point in the feature space, $g_i = g\left(\left\|\frac{x_s - x_{is}}{h_s}\right\|^2\right) g\left(\left\|\frac{x_r - x_{ir}}{h_r}\right\|^2\right)$ and g(x) = k'(x). k'(x) is the derivative of the kernel profile k(x) given in equation 4-3. For every data point/pixel x in the image, the mean shift vector was calculated (equation 4-6) in the spatial-range domain and move x in the direction until convergence is reached.

Using this method of finding the mode (local maxima) associated with each data point helps to smooth the image while preserving discontinuities. Intuitively, if two points x_i and x_j are far from each other, in feature space, then x_i will not be in same domain as x_i, x_j doesn't contribute to the mean shift vector gradient estimate and the trajectory of x_i will move it away from x_j. Therefore, pixels on either side of a strong discontinuity will not attract each other. The final result is a mean shift segmented image (figure 4.4).

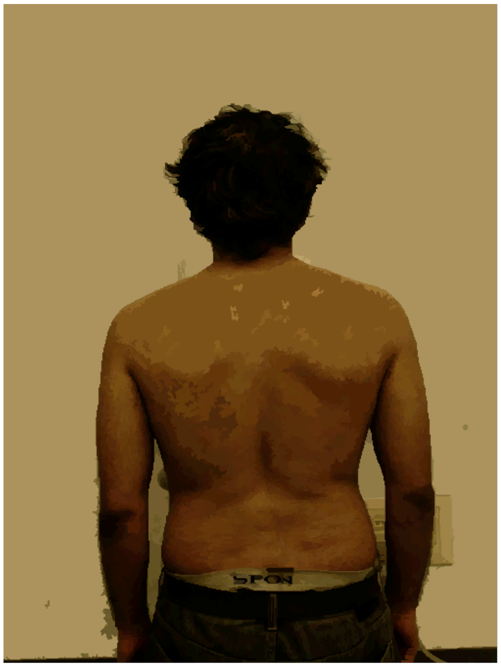

Figure 4.4(a): Segmented Top Image

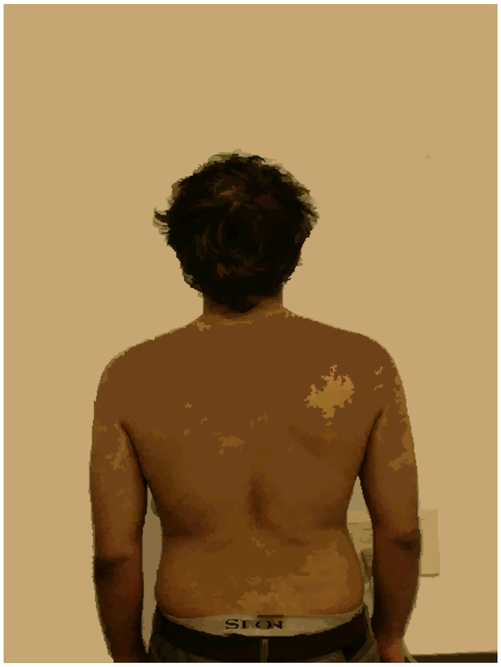

Figure 4.4(b): Segmented Bottom Image

4.2 Differential Localized Pixel Matching

The objective of differential localized pixel matching is to compute matching costs of pixel in the bottom camera image to pixel in the top camera image (reference image).

As explained in section (2.3.2.4), local pixel matching calculates a matching score defined over an aggregation window. The latest matching scores described in the literature review such as gradient-based and non-parametric measures are more robust to changes in camera gain and bias but they violate the frontal parallel assumption. A differential localized dissimilarity measure was developed that combines sum of absolute intensity differences (SAD) and deformed differential SSD (sum of squared intensity differences) based measure defined as follows:

$$C_{SAD}(\mathrm{x, y, d}) = \sum_{(i,j)\in N(x,y)} \left(I_1(i,j) - I_2(i+d,j) \right)$$

and

$$C_{DSSD}(\mathrm{x,\ y,\ d}) = \mathrm{arg} \quad min_{\left\{d, \frac{\partial d}{\partial i}, \frac{\partial d}{\partial j}\right\}} \sum_{(i+\partial i, j+\partial j)\in N(x,y)} \left(I_1(i+\partial i, j+\partial j) - I_2\left(i + \partial i - d - \frac{\partial d}{\partial i}\partial i - \frac{\partial d}{\partial j}\partial j, j + \partial j\right) \right)^2 \qquad [4\text{-}7]$$

where N(x, y) is a 3x3 surrounding at position (x, y) and i and j are array indexes of the two dimensional array representing the camera images. (i,j) are (0,0) on the top left corner of the image as mentioned in section 3 [3]. C$_{DSSD}$ calculates matching cost for every (i,j) in the bottom image in a deformed window SSD using a direction set method as defined in [8]. The direction set method is a multidimensional minimization method, initialized with disparity d (obtained using traditional SSD) and zeroes for first order disparities. If correspondence of (i, j) in top (reference) image is (i-d, j) in the bottom image, then to a first order approximation the correspondence of $(i + \partial i, j + \partial j)$ is $\left(i + \partial i - d - \frac{\partial d}{\partial i}\partial i - \frac{\partial d}{\partial j}\partial j, j + \partial j\right)$. $\frac{\partial d}{\partial i}$ and $\frac{\partial d}{\partial j}$ are the partial derivatives of disparity d with respect to i and j respectively. ∂i and ∂j are small step size in each direction. I_I is intensity of the top camera image and I_2 is the linearly interpolated intensity of the two nearest integer index positions in the bottom camera image.

An optimal weighting ω between C_{SAD} and C_{DSSD} is determined by maximizing the number of reliable correspondences that are filtered out by cross-checking top to down and down to top disparities and applying winner take all optimization (choosing the disparity with the lowest matching cost) [3]. The cross checking test is done by calculating C_{SAD} first using the top image as reference image and then using bottom image as reference image. If the C_{SAD} is not the same in both cases, then those pixels are discarded as occluded pixels and not used in any further calculations. The remaining pixels are called non-occluded pixels. After determining ω, the resultant deformable disparity matching cost is given by

$$C_O(\mathrm{x,y,d}) = (1-\omega)\mathrm{C_{SAD}}(\mathrm{x,y,d}) + \omega\mathrm{C_{DSSD}}(\mathrm{x,y,d}) \qquad [4\text{-}8]$$

Among all the possible disparities for pixel at position (x,y) the one with the minimum matching cost $C_O(\mathrm{x,y,d})$ is selected as the initial disparity.

4.3 Disparity Plane Definition

A disparity plane is specified by three parameters c_1, c_2, c_3 that determine a disparity d for each reference image pixel (x, y)

$$d = c_1x + c_2y + c_3 \qquad [4\text{-}9]$$

Due to huge number of disparity planes, the number is reduced by extracting a set of disparity planes that is sufficient to represent scene structure. This is achieved by calculating a disparity plane for each segment and refining plane fitting similar to [7]. The segmented image and the associated matching cost of pixels within each segment was used to define disparity planes. Our goal is not to find the best plane for each segment but rather extract all possible planes for the image. The reason behind this approach is that small fragmented segments should be grouped to provide more reliable pixels to form the linear equation (4-9).

The steps for disparity plane definition are: (1) calculate a disparity plane to represent each segment, (2) calculate matching cost for each segment-to-plane assignment, (3) assign each segment the plane that gives minimum matching cost, (4) group neighboring segments which have the same disparity plane and (5) Repeat steps (1) to (3) for all the grouped segments. Steps 1 and 2 are described in detail below. Steps 3 to 5 are trivial.

In step one; each of the parameters was solved (c_1, c_2, c_3) separately by applying a decomposition method. Horizontal slant is determined using the initial disparities that are lying on the same horizontal line within a segment. The derivates $\frac{\partial d}{\partial x}$ are inserted to a list and an estimate of the horizontal slant is determined by sorting the list and applying convolution with a Gaussian kernel K(x) from equation (4-2). Vertical slant was determined in the same way the horizontal slant was determined. The determined slant is used to obtain an estimate in the center of the segment. The corresponding center disparities for each point, that are calculated by considering the estimated slant, are inserted into a list and an estimate is obtained by applying convolution with a Gaussian kernel (like before). This gives us the three parameters (c_1, c_2, c_3) and this created disparity planes for all the segments in the bottom image using equation (4-9).

In the second step, the matching cost was calculated for each segment to plane assignment. Using the method defined in [3], the sum of the matching cost for each pixel inside the segment S is given by:

$$C(S,P) = \sum_{(x,y)\in S-O} C_O\,(x,y,d)\; e^{1-\frac{s}{n}} \qquad [4\text{-}10]$$

where S is a segment, P is a disparity plane, d = $c_1^P x + c_2^P y + c_3^P$ (c_1^P, c_2^P and c_3^P are parameters of the plane P), n is the number of non-occluded pixels in segment S, s is the number of pixels that are part of both the segment S and the plane P and O is the occluded portion in S represented by the occluded pixels obtained in the last section.

After steps 3 to 5 are completed, a set of disparity planes were used to represent the bottom image. In the final stage of stereo image registration, the disparity planes were optimized using a novel method of max product belief propagation with differential geometry message passing.

4.4 Belief Propagation

Belief Propagation is a disparity optimization method used by segment and global stereo class of registration algorithms. This class of algorithms performs better than local stereo registration algorithms as noted by Scharstein and Szeliski [3]. The aim of disparity optimization in segment based and global stereo registration is to minimize the global energy. As explained in section 2.3.2.6, minimizing the global energy in an image is equivalent to solving maximum a posteriori – Markov Random Field (MAP-MRF) problem. Belief propagation is a method used to solve the MAP-MRF problem. It does so by iteratively passing messages relating to differences in disparities amongst neighboring pixels in a image and updating disparities until the global energy E(d) achieves convergence.

As defined by Sun et. al. in [5], Belief Propagation is an iterative inference algorithm that passes messages in a network of nodes represented by image pixels/points. Let $m_{ij}(x_i,x_j)$ be the message that node x_i sends to x_j, $m_i(x_i,y_i)$ be the message that observed node y_i, sends to node x_i, and $b_i(x_i)$ be the belief at node x_i. Belief Propagation is formulated as the estimation of a random variable x_k for every node k in a MRF. Let $\psi(x_i, x_j)$ denote the compatibility function, which encodes the compatibility between two immediate neighboring nodes i and j and $\Phi(x_k, y_k)$ denote local evidence that variable x_k is consistent with observation y_k. $m_{ij}(x_i,x_j)$ is represented as $m_{ij}(x_j)$, $m_i(x_i,y_i)$ as $m_i(x_i)$ and $\Phi(x_k, y_k)$ as $\Phi(x_k)$. Local evidence $\Phi(x_k)$ is

$$\Phi(x_k) = (1 - \varepsilon_{data})e^{-\frac{|C(x_k,y_k,d_k)|}{\sigma_{data}}} + \varepsilon_{data} \qquad [4\text{-}11]$$

where $C(x_k, y_k, d_k)$ is the matching cost calculated at position x_k, y_k with $d_k = c_1x_k + c_2y_k + c_3$ where c_1, c_2 and c_3 are parameters of the plane, P such that k ε P [18]. By varying ε_{data} and σ_{data} the shape of $\Phi(x_k)$ was optimized. There are two kinds of Belief Propagation (BP) algorithms with different message update rules: max-product and sum-product. Max Product BP is based on the

MAP-MRP and Sum Product BP is based on MMSE-MRP as explained in section 2.3.2.6.1. Max Product BP maximizes the joint posterior P(X|Y) of the network. The Max Product BP is modified by applying differential geometry to message calculation that are not constrained by the frontal parallel assumption.

4.4.1 Differential Max Product Belief Propagation

The steps for Max Product Belief Propagation with differential message passing are as follows:

(1) Initialize all messages $m_{ij}(x_j)$ as $\psi(x_i, x_j)$ and messages $m_i(x_i)$ = $\Phi(x_i)$.

$m_i(x_i)$ messages contain the differential local matching cost component, as shown in section 4.2. The original $\psi(x_i, x_j)$ as seen in [16], is given by

$$\psi(x_i, x_j) = (1 - \varepsilon_{smooth})e^{-\frac{|d_i - d_j|}{\sigma_{smooth}}} + \varepsilon_{smooth} \qquad [4\text{-}12]$$

where d is disparity at position x.

The above $\psi(x_i, x_j)$ equation follows the frontal parallel assumption. This means that if disparities x_i and x_j lie on the same plane, P, then based on equation (4-9) parameters c_1, c_2 and c_3 are same for calculation of d_i and d_j.

So, when message $m_{ij}(x_j)$ is used, x_i, x_j and all other points that lie on P will have constant disparities. However, disparity/depth values in many smooth and curved surfaces like the torso changes more rapidly than in objects like a cardboard box. Therefore, incorporating differential geometric constraints in $\psi(x_i, x_j)$ gives

$$\psi(x_i, x_j) =$$

$$\left((1-\varepsilon_{smooth})e^{-\frac{|d_i-d_j|}{\sigma_{smooth}}} + \varepsilon_{smooth}\right)((1-\varepsilon_N)e^{-\frac{\left\|N_{d_i}-N_{d_j}\right\|^2}{\sigma_N}} + \varepsilon_N)$$

[4-13]

where N is the surface normal in the disparity map and can be computed as

$N = \frac{(-z_x, -z_y, 1)}{\sqrt{1+z_x^2+z_y^2}}$. $z_x = -\frac{\propto b}{d^2}\frac{\partial d}{\partial x}\frac{\propto}{f}$ and $z_y = -\frac{\propto b}{d^2}\frac{\partial d}{\partial y}\frac{\propto}{f}$ with b as the stereo baseline, α is the focal length in pixels and f is focal length in physical unit like mm [7, 19]. The formula for z is based on the epipolar geometric constraint described in section 3.

(2) Update the messages $m_{ij}(x_j)$ iteratively for i = 1:T

$$m_{ij}^{t+1}(x_j) = \alpha \, max_{x_i} \, \psi(x_i, x_j) \, \Phi(x_i) \prod_{k \in N(i) \backslash j} m_{ki}^t(x_i) \qquad [4\text{-}14]$$

where m_{ij}^{t+1} is the message that node at index i sends to node at index j at iteration t+1, N(i)\j is the set of nodes neighboring node i except node j itself and T is the number of iterations.

(3) Compute beliefs

$$b_i(x_i) = \alpha \, \Phi(x_i) \prod_{k \in N(i)} m_{ki}(x_i) \qquad [4\text{-}15]$$

(4) Calculate the MAP solution at node i

$$x_i^{MAP} = \arg max_{x_k \in \{d_{1,\dots,}d_N\}} b_i(x_k) \qquad [\ 4\text{-}16]$$

After each iteration, the energy for each segment is calculated as below

$$E_S^t = \sum_{i \in S} b_i\,(x_i) \qquad [4\text{-}17]$$

If the energy for iteration t and t+1 are the same, then it is assumed to have reached minimum energy and Differential Max Product Belief Propagation convergence has been achieved. Repeating this process over all the segments in the image, gives us the minimum energy for each segment and solves the energy minimization problem. The disparity levels achieved after differential max product belief propagation convergence over all the segments in the image are the final disparity levels. These disparity values were used in the next section to determine the depth (z) values for all (x, y) points/pixels in the image in order to reconstruct the 3D Torso image.

4.5 Results

The resultant Disparity Space Image (DSI) of the bottom image with top image as reference image is shown in figure 4.5.

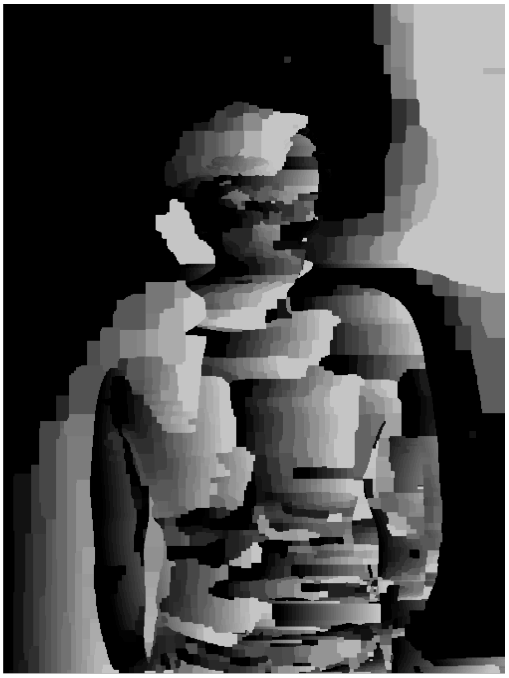

Figure 4.5: DSI of Bottom Image

The registration is compared to 2 existing registration procedures, segment-based adaptive belief propagation (adaptive BP) and color-weighted hierarchical belief propagation (hierarchical BP). These registration procedures were ranked as best performing by Scharstein et. al. in their taxonomy of stereo registration algorithms [3]. The resultant DSI obtained from the 3 registration processes was converted into x, y, z coordinate values and we obtained 360 cross sectional 3D point values of the torso. Each cross section was compared to the 3D Torso Image obtained from the Konica

Minolta 700 Scanner (ground truth of evaluation). The evaluation methodology and results are described in detail in section 6.

4.6 Achievements

Differential geometry was applied to "state of the art" stereo registration algorithms to reduce registration errors for smooth and curved surfaces like the torso. We created a novel Differential Segment based Belief Propagation for registering stereo images of smooth and curved surfaces.

5. 3D Image Reconstruction

3D image reconstruction comprises of 5 stages - converting the disparity values (generated through stereo image reconstruction) to 3D data points, triangulating the 3D point cloud to attain a torso surface, cross sectioning the torso surface, removing stray data points and finally filling in occlusions/holes to get the reconstructed 3D torso image. Known principles in 3D geometry are applied to complete the first stage of obtaining 3D data points from disparity values. In the second stage, a nearest neighbor triangulation algorithm for smooth surfaces was applied to get the 3D surface image [34]. To extract the region that represents the torso from this 3D surface image, cross sectioning was performed and stray data points were determined and then deleted. Finally, a novel occlusion filling algorithm called BC-MLS (Bezier Curve – Moving Least Squares) was developed to fill in missing data points and obtain the reconstructed 3D torso image.

5.1 Conversion of Disparity Space Image to 3D Points

The x,y values data points were obtained from pixel values in the image as shown in Chapter 3. Using equation (5-1) from 3D camera geometry, the z spatial coordinate is determined.

$$z = \frac{b \times f}{d(x, y)} \qquad [5\text{-}1]$$

In equation (5-1), b (baseline) represents the distance from the optical centre of the top camera to that of the bottom camera, f is the focal length of the cameras and d(x,y) is the disparity at that x, y location on the image. The focal length setting on the digital camera is 18 mm (manually set on both the top and bottom digital cameras) as explained in section 3.1 and the baseline is set to 100 cm as detailed in section 3.4. The x, y, z values are plotted using Kitware's Visualization Toolkit ® (VTK) software [14].

5.2 Triangulation

The *vertex* array of 3D points (p_i) was constructed. This in rectangular co-ordinate format is given by (p_{ix}, p_{iy}, p_{iz}) where $p_{ix}$is the width, p_{iy} is the height and p_{iz} is the depth of the 3D image. This array of 3D points is constructed from the x, y, z values obtained previously. The 3D surface image was generated by connecting 3D points with line segments using the nearest neighbor interpolation of distance functions for smooth surface reconstruction [34]. Now, this surface wireframe in VTK was drawn using the built-in visualization tools – vtkPoints, vtkLine and vtkPolygon.

The resultant image is shown in figure 5.1.

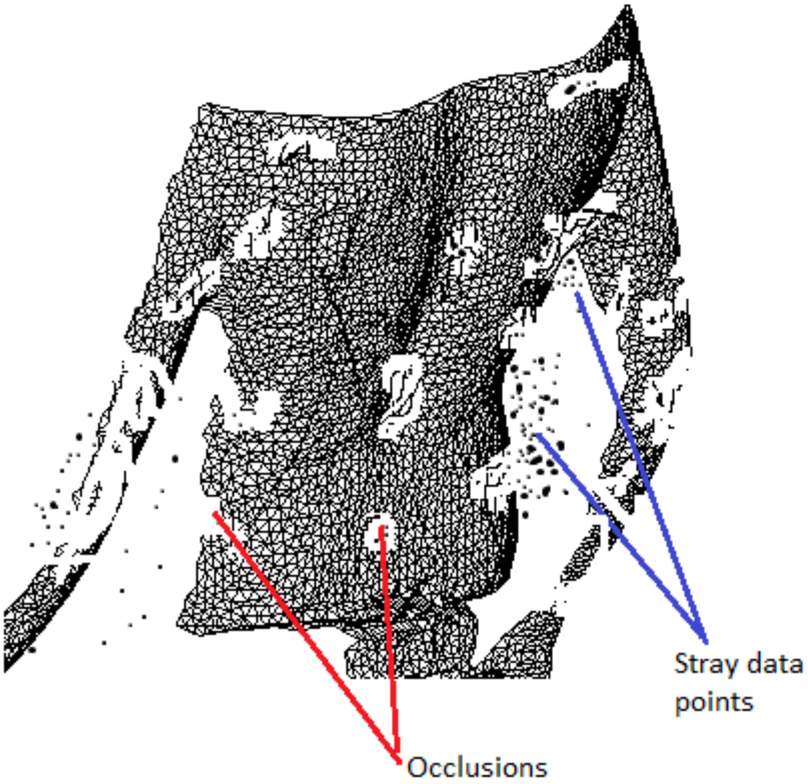

Figure 5.1: Triangulated 3D Point Map from DSI

There are two major issues as seen in the figure above - the presence of stray data line segments and points that are not part of the torso and the existence of occlusions due to missing data points. First, cross sectioning is performed to reduce a 3D problem to a 2D one. Then, the stray data points are removed in each cross section and BC-MLS interpolation is performed to generate data points in occluded regions of each cross section.

5.3 Cross sectioning

The stray data points and the occlusions in the 3D surface image are due to errors in DSI calculation in the stereo image registration stage. We divide the surface image into cross sections to remove the unconnected (points not joined by line segments) stray data points in each cross section.

The process of obtaining cross sections is carried out using horizontal planes whose origins are set at a particular height p_{oy} where p_{oy} represents the height coordinate of the *origin* $p_o = (p_{ox}, p_{oy}, p_{oz})$. The origin of the first plane is located at $p_{iy(MIN)}$ (minimum height of the 3D torso image). The number of cross sections is user defined. We used a value of 360 cross sections. The successive increments of plane origin are computed by dividing $p_{iy(MIN)}, p_{iy(MAX)}$ (the maximum and minimum height) of the torso image by the number of cross sections.

Implicit functions vtkPlane and vtkCutter in VTK are used to perform the cross sectioning. The vtkPlane function is used to define the plane. vtkCutter uses the plane to create cross sections of points at pre-defined heights (p_{oy}). It does so by using vtkPlane to intersect the torso image at p_{oy} and creating array 3D points from the points of intersection to represent the cross section (figure 5.2).

A vertex array and a connectivity map were obtained for each cross-section. The vertex array is a N x 3-dimensional array that stores the physical locations of each point in the cross-section. The vertices that were connected by line segments in the triangulation procedure form the connectivity array. They were numbered counter-clockwise starting from the left-most point on the image to form the connectivity map.

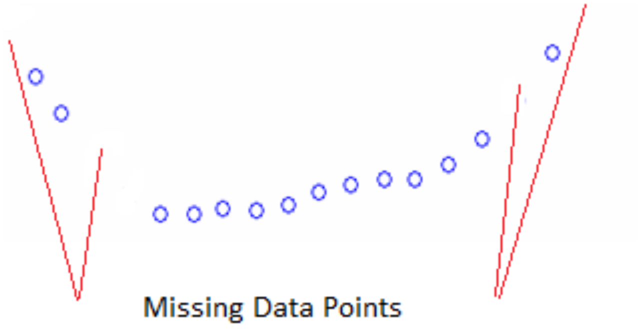

Figure 5.2: Single Cross section of 3D Torso point cloud.

The red lines show occlusions in the cross section

5.4 Stray Data Point Removal

Stray data point removal comprises of 2 passes. In the first pass, we identify stray data points as 3D points that connected to no other point and exist in the vertex array but not in the connectivity map. These are removed from the vertex array and appropriate modifications are made to the vertex array structure to maintain consistency.

In the second pass, the remaining stray data points are identified as 3D points (p_i) whose p_{iz} is more than twice the Euclidean distance from its nearest neighbor in each cross section. This is because the torso is a gradually curving surface. These stray data points are removed from the vertex array and connectivity array. The resulting surface image is generated in VTK using the vertex and connectivity array as shown in figure 5.3.

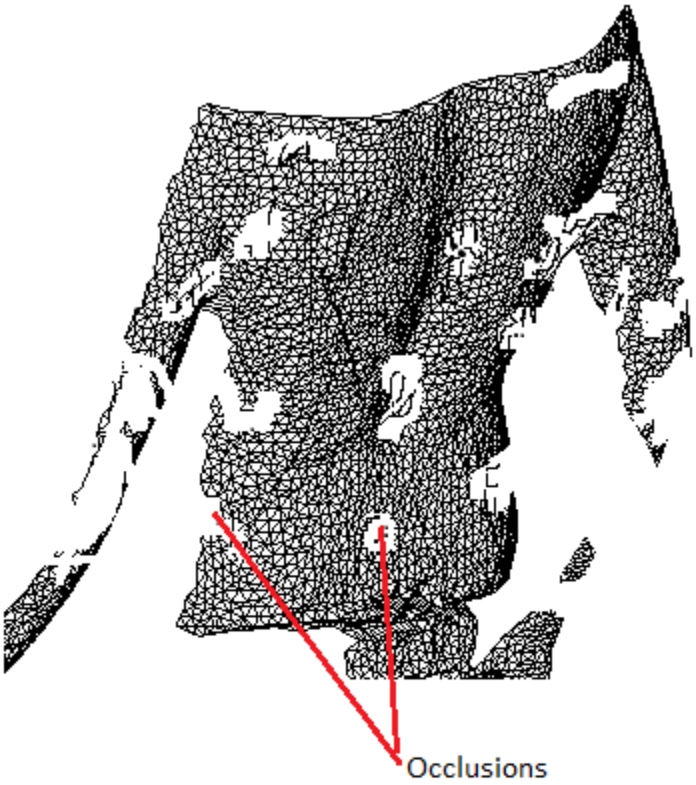

Figure 5.3: Removal of Stray Data Points from Triangulated 3D Point Map

5.5 Torso Extraction

A two step procedure for clipping the extremities of the 3D image was implemented based on existing Sutherland and Hodgman clipping algorithm [27].

Step 1 is a Plane clipping for upper and lower extremities. The upper extremities (corresponding to the regions of the images above the base of the neck) and the lower extremities (corresponding to regions of the images below the waist) were cropped using a horizontal cutting plane using the Sutherland and Hodgman plane-clipping algorithm [27]. The vertical coefficients of the cutting planes in the neck and waist were manually defined and delineated the extent of the crop. The plane was constructed in VTK using the vtkPlane function.

As the human torso is usually asymmetric, the left and right extremities rarely attach to the torso in planes parallel to the torso medial plane [2]. This made it difficult to automatically crop the left and right extremities of most torsos using plane-clipping algorithms. In step 2, the left and right extremities were clipped using an implementation of the Sutherland–Hodgman box-clipping algorithm. For every instance of the box clipper, variables that delineate the extent of the bounding box were user defined. The box clipper was constructed in VTK in the vtkBox function (figure 5.4).

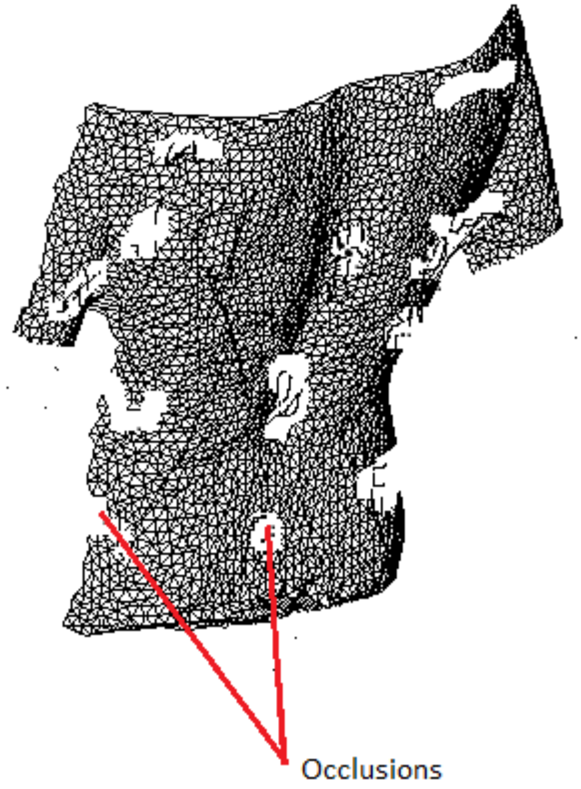

Figure 5.4 Result of Torso extraction from Triangulated 3D Point map

5.6 Occlusion Filling

The points generated from Bezier Curve (BC) and Moving Least Squares (MLS) interpolation do not match the shape of the torso and will cause significant shape distortions to the torso sections if used to fill *holes*. The BC and MLS projection procedure (described in section 2.3.3) were modified and used together, to develop an algorithm called BC-MLS for *hole filling*. The BC-MLS *hole filling* algorithm is a 2D algorithm since it is applied on each 2D cross section. It is performed in two steps. The first step is the basic setup and BC implementation. The second step is the modified MLS projection interpolation.

In the first step, assume p_1 and p_2 are two 3D points lying on a 2D cross sectional plane such that $p_1, p_2 \in R^3$ and p_1, p_2 have a *hole* in between them. Let $p_3,\ p_4$ be two 3D points such that $p_3, p_4 \in R^3$. The 3D points are joined by line segments to p_1, p_2 respectively on either side of the *hole*. Using the basic principles of geometry, the two lines joining $p_1,\ p_3$ and $p_2,\ p_4$ can be defined as

$$y - p_{1y} = \frac{p_{3y} - p_{1y}}{p_{3x} - p_{1x}}\left(x - p_{1x}\right) \text{ and } y - p_{2y} = \frac{p_{4y} - p_{2y}}{p_{4x} - p_{2x}}\left(x - p_{2x}\right) \qquad [5\text{-}2]$$

The point of intersection of the two lines is given by p_5 such that $p_5 \in R^3$ where

$$p_{5x} = \left(\frac{\left(p_{2y} - p_{1y}\right)\left(p_{3x} - p_{1x}\right)\left(p_{4x} - p_{2x}\right) - p_{1x}\left(p_{3y} - p_{1y}\right)\left(p_{4x} - p_{2x}\right) - p_{2x}\left(p_{4y} - p_{1y}\right)\left(p_{3x} - p_{1x}\right)}{\left(p_{3y} - p_{1y}\right)\left(p_{4x} - p_{2x}\right) - \left(p_{4y} - p_{2y}\right)\left(p_{2x} - p_{1x}\right)}\right)$$

$$\text{and } p_{5y} = \left(\frac{p_{3y} - p_{1y}}{p_{3x} - p_{1x}}\right)\left(p_{5x} - p_{1x}\right) + p_{1y}. \qquad [5\text{-}2]$$

Two new 3D points $p_6, p_7 \in R^3$ are constructed by averaging the initial 3D points p_1 and p_2 (which have a *hole* in between them) with p_5, such that

$$p_6 = \left(\frac{p_1 + p_5}{2}\right) \text{ and } p_7 = \left(\frac{p_2 + p_5}{2}\right). \qquad [5\text{-}3]$$

Now, the four 3D points lying on the 2D cross-sectional plane p_1, p_6, p_7, p_2 are the *control (initial) points* and together form a *control polygon* (CP) for a resultant Bezier curve. From classical BC theory, the four 3D *control points* are used to establish the following cubic BC parametric equation

$$p(t) = (1-t)^3 * p_1 + 3*(1-t)^2 * t * p_6 + 3*(1-t)*t^2 p_7 + t^3 * p_2, 0 \le t \le 1$$

where $p(t) = (x(t), y(t), z(t))$. An initial value of t is chosen to determine the location of the first intermediate point starting from p_1. t is varied according to $\{t, 2t, 3t, ...It\}$ i.e. at equal intervals where I represents the total number of intermediate points between p_1 and p_2 such that $I \in Z$. Let CP distance (CPd) be the sum of the Euclidean distances between p_1, p_6 and p_6, p_7 and p_7, p_2 such that $CPd \in R$

$$CPd = \sqrt{(p_{6x} - p_{1x})^2 + (p_{6y} - p_{1y})^2} + \sqrt{(p_{7x} - p_{6x})^2 + (p_{7y} - p_{6y})^2} + \sqrt{(p_{2x} - p_{7x})^2 + (p_{2y} - p_{7y})^2} \qquad [5\text{-}4]$$

Let the average Euclidean distance between *connected* 3D points lying on the 2D cross sectional plane be $avgCd$, such that $avgCd \in R$

$$avgCd = \sum_{i=1}^{C-1} \left(\frac{\sqrt{(p_{(i+1)x} - p_{ix})^2 + (p_{(i+1)y} - p_{iy})^2}}{N-1} \right) \qquad [5\text{-}5]$$

where C is the total number of points in the 3D torso image and i is the index or location of one such 3D point p. In the second step CPd and avgCd was used on MLS projection procedure . The MLS projection is performed under each iteration of BC procedure on every cross section. The setting of $i = i - 1, N = i + 2, h = avgCd$ is applied to the MLS procedure in section 2.3.3.2 (figure 5.5).

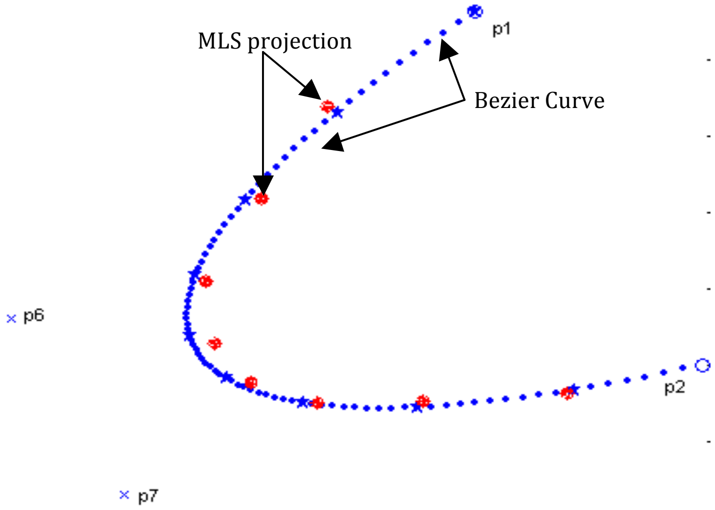

Figure 5.5: Illustration of the MLS-BC hole filling

The new set of data points in P provides a better local approximation. The BC-MLS procedure is repeated for each cross section in the 3D torso image.

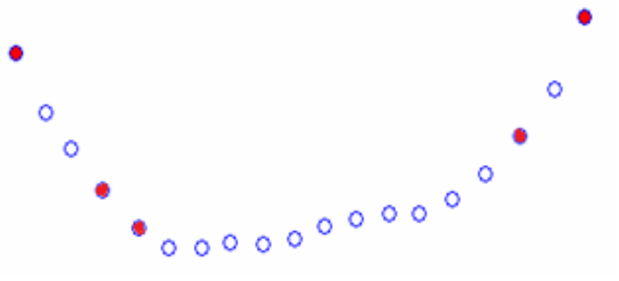

Figure 5.6 Demonstration of BC-MLS hole filling on Torso cross section

(points in red are generated by BC-MLS)

5.7 Results

The points on all the cross sections as defined in section 5.2 were re-triangulated. This results in an occlusion free 3D back surface image that can be used for assessment of scoliosis (figure 5.7).

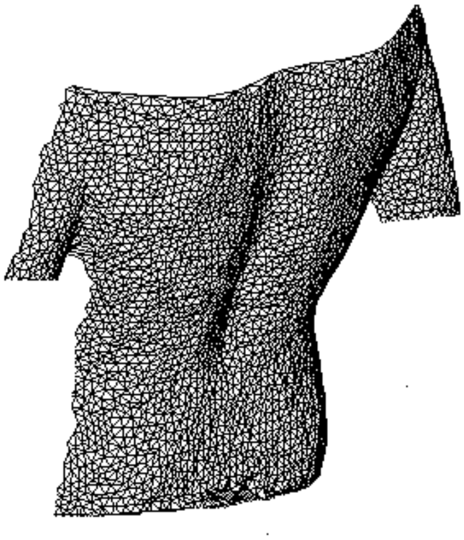

Figure 5.7: Reconstructed 3D Torso Image

5.8 Achievements

A novel process of Bezier Curve – Moving Least Squares (BC-MLS) occlusion filling was created. Finally, using known geometric principles, 3D points were converted to a 3D back surface image.

6. System Validation and Testing

The relative clinical utility of the 3D Back Reconstruction (Stereo Image Acquisition + Stereo Image Registration + 3D Image Pre-processing and Reconstruction) using Stereo Digital Cameras described in this thesis vis-à-vis existing Image Registration and 3D Image pre-processing systems was assessed by determining the variability in the computation of the Cosmetic Score index from 3D back images reconstructed using each method. Figure 6.1 illustrates the six landmarks used in the computation of the Cosmetic Score [28]. Each set of six landmarks was used to obtain a Cosmetic Score. The reconstruction accuracy of our Differential Belief Propagation (differential BP) was evaluated against registration methods of segment-based adaptive belief propagation (adaptive BP) and color-weighted hierarchical belief propagation (hierarchical BP) on 10 reconstructed 3D back images of 2 human subjects. Similarly, our (Bezier Curve – Moving Least Squares) BC-MLS image pre-processing and reconstruction was compared against Bezier Curve (BC), Moving Least Squares (MLS) and FastRBF approximation on the 10 3D back images. FastRBF is a commercially available surface reconstruction package based on radial basis functions created by FarField Technology, New Zealand [33]. The 4 validation indices used to evaluate reconstruction accuracy are detailed in Table 6.1.

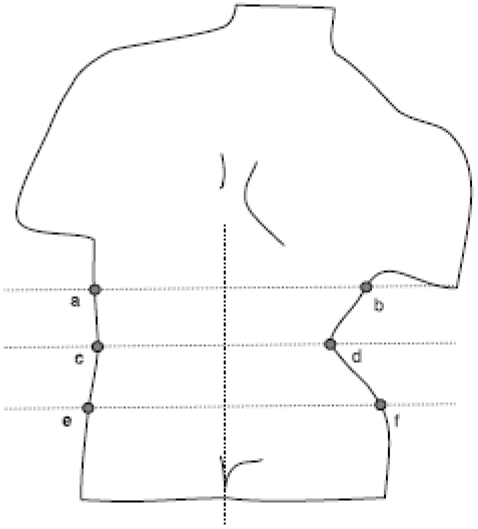

Figure 6.1: Torso landmarks used to compute the Cosmetic Score. The first step to calculating the Cosmetic Score is to obtain the positive or negative offsets of lines a–b, c–d and e–f from the centerline. The Cosmetic Score itself is the normalized ratio between the offset of c–d and the average offset of a–b and e–f [28]

6.1 Validation methods

Five back digital stereo images and five range image scans each of two human subjects were obtained using stereo image acquisition from stereo digital cameras and the Konica Minolta® Vivid 700 laser scanner respectively and simultaneously. The range images scans were processed into 3D back images using the Polygon Editing Tool Version 2.0 supplied with the Vivid 700 laser scanner. 360 horizontal cross-sections of the range reconstructed 3D back images were obtained from each of the 10 3D back images. These 360 cross sections served as a gold standard for comparison. Ten back images obtained through stereo image acquisition were registered and reconstructed into 3D back images using our Differential BP + BC-MLS hole filling (DP + BC-MLS), Adaptive BP + Bezier Curves (ABP + BC), Adaptive BP + FastRBF (ABP + FRBF), Hierarchal BP + Bezier Curves (HBP + BC) and Hierarchial BP + FastRBF (HBP + FRBF) methods. Illustration of each step of 3D Back Reconstruction from Differential BP + BC-MLS method is shown in figure 6.2. Each of the stereo reconstructed back images were divided into 360 cross sections at the same horizontal heights as the range reconstructed back images (gold standard) and evaluated using four validation indices (Table 6.1). The stereo reconstructed cross-sections were optimally aligned to the range reconstructed cross-sections using their centroids and maximal diameters. A minimum-bounding semi-circle containing the range and stereo reconstructed cross-sections was defined as the universal set for the purpose of computing indices C and D. An average value of each index was obtained over the 360 cross-sections. (A minimum-bounding ellipse did not produce a significant change in the results.)

	Formula*	Psuedo-name	Description
A	$\frac{\mathrm{Cr} \cap \mathrm{Cs}}{\mathrm{Cr}}$	Sensitivity	This is a ratio of the size of the overlap between the range and stereo reconstructed cross-section to the size of the range reconstructed cross-section.
B	$\frac{\mathrm{Cr} \cap \mathrm{Cs}}{\mathrm{Cs}}$	Positive predictive value	This is a ratio of the size of the overlap between the range and stereo reconstructed cross-section to the size of the stereo reconstructed cross-section.
C	$\frac{\Im - \mathrm{Cr} \cup \mathrm{Cs}}{\Im - \mathrm{Cr}}$	Specificity	This is a ratio of the size of the difference between the union of the two cross-sections from the universal set to the difference between the range reconstructed cross-section from the universal set.
D	$\frac{\Im - \mathrm{Cr} \cup \mathrm{Cs}}{\Im - \mathrm{Cs}}$	Negative predictive value	This is a ratio of the size of the difference between the union of the two cross-sections from the universal set to the difference between the stereo reconstructed cross-section from the universal set.

Table 6.1 Validation Indices

*Cr is the range reconstructed cross-section; Cs is the stereo reconstructed cross-section (or volume) and I is the universal set consisting of the minimum bounding circle (or cylinder) containing the original cross-section (or volume) and the reconstructed cross-section (or volume).

Figure 6.2(a) Original Top and Bottom Camera Images

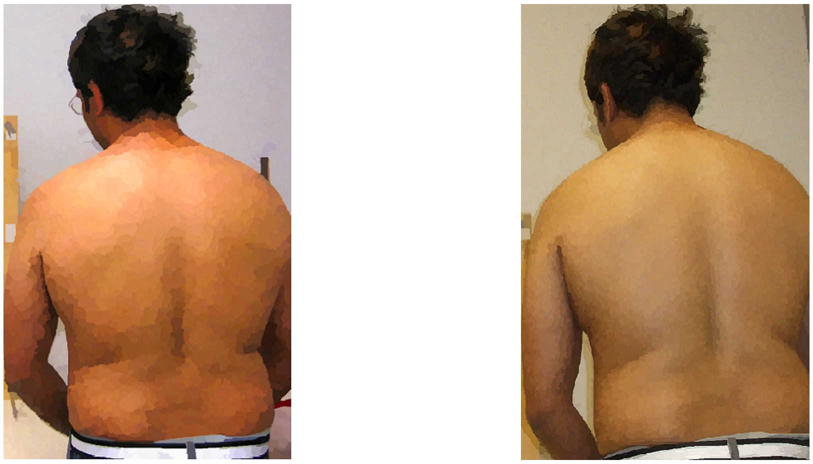

Figure 6.2(b) Segmented Top and Bottom Camera Images

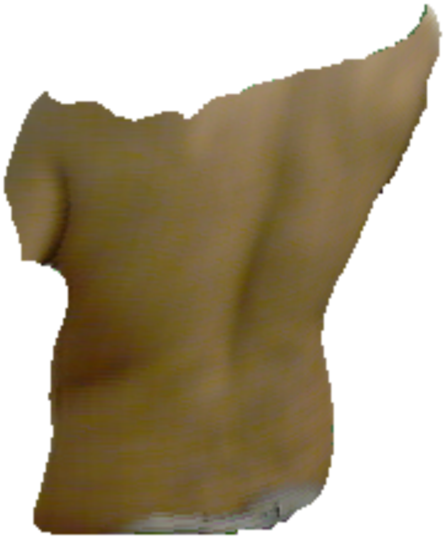

Figure 6.2(c) Reconstructed 3D Torso Image

Clinical assessments of torso deformity utilize indices that are derived from ratios of the distances and angles between anthropometric landmarks on the torso. Several indices are currently in use (these include the Cosmetic Score), the posterior trunk symmetry index score, the ISIS score and the Quantec score [2]. These landmarks are currently used in scoliosis clinics and have been shown to significantly correlate with the underlying spinal deformity

that causes scoliosis [28]. A source of error in the computation of torso deformity indices is uncertainty in determining the location of the relevant landmarks. As the accuracy of torso reconstruction increases, the relevant landmarks become easier to locate and the error associated with locating them may reduce. This in turn reduces the variability in the clinical indices of torso asymmetry for the torso reconstruction. Thus, a measure of the clinical utility of a torso image reconstruction process is the variability in cosmetic score index obtained from landmarks obtained from torsos reconstructed using the process. Cosmetic Score (the ratio of the waist to hip diameters measured in 3D) [28] is used to evaluate the clinical effect of the various reconstruction methods. The choice of a surface deformity index such as the Cosmetic Score rather than a spinal deformity index, is due to a weak correlation between the internal and external manifestations of scoliosis on the torso [1].

6.2 Results and Analysis

Table 6.2 shows the results obtained. The Differential Belief Propagation registration with Bezier Curve – Moving Least Squares occlusion filling outperformed the other stereo reconstruction methods for all the indices measured.

	Indices			
Methods	**A**	**B**	**C**	**D**
Adaptive BP+ Bezier Curves (ABP + BC)	0.89	0.91	0.93	0.73
Adaptive BP + FastRBF (ABP + FRBF)	0.90	0.95	0.91	0.75
Hierarchical BP + Bezier Curves (HBP + BC)	0.88	0.92	0.94	0.70
Hierarchical BP + FastRBF (HBP + FRBF)	0.93	0.96	0.93	0.76
Differential BP + BC-MLS (DBP + BC-MLS)	0.97	0.99	0.99	0.76

Table 6.2 Average validation indices obtained for 360 cross-sections and 10 reconstructed human back cross sections. The indices A-D correspond to the validation indices explained in Table 6.1

Adaptive BP was outperformed due to the frontal parallel assumption (which isn't valid on curved surfaces like the torso). The localized pixel matching applied in adaptive BP, however did produce good results. Hierarchical BP achieves great accuracy on the color segmentation stage but again assumes frontal parallel geometry on belief propagation. Differential Max Belief Propagation uses differential geometry on pixel matching and belief propagation and therefore outperforms other registration methods for curved surfaces.

Bezier curve approximation produced symmetrical C1 smooth interpolation arcs while BC-MLS produced somewhat-skewed smooth interpolation arcs. The degree of skew of the arcs was chiefly determined by the MLS parameters used. Cross-sections of the human torso do not exhibit local perfect symmetry, thus the BC-MLS interpolation yielded a closer fit to the original curve as compared to FastRBF.

The clinical utility of reconstructions using the registration methods was assessed from the variability in the computation of Cosmetic Score [28]. Ten scores were obtained for each of the 10 torso images for each of the five methods as shown in Table 6.3

	Cosmetic Scores				
Images	ABP + BC	ABP + FRBF	HBP + BC	HBP + FRBF	DBP + BC-MLS
One	1.2	1.2	1.3	1.0	1.1
Two	1.0	1.1	1.4	1.3	1.2
Three	1.0	0.8	1.0	0.8	0.9
Four	1.0	1.2	0.9	1.2	1.0
Five	1.5	1.2	1.1	1.3	1.2
Six	1.0	1.0	0.8	0.8	0.9
Seven	0.9	0.7	0.9	0.9	0.8
Eight	0.9	1.1	1.2	0.9	1.0
Nine	0.8	0.8	0.6	0.8	0.7
Ten	1.2	1.0	1.3	1.0	1.1

Table 6.3 Cosmetic Score indices obtained for 360 cross-sections and 10 reconstructed human back cross sections. Images 1-5 are of Subject 1 and images 6-10 of Subject 2.

The variability of each of the five methods was calculated from the standard deviations of the scores obtained for each torso image. The following average variability values were obtained Adaptive BP + Bezier Curves – 13%, Adaptive BP + FastRBF – 7.5%, Hierarchal BP + Bezier Curves – 15% and Hierarchial BP + FastRBF – 8% and Differential BP + BC-MLS hole filling – 6%.

Results demonstrate that the process of reconstructing 3D torso images from stereo 2D torso images described in this thesis out performs existing stereo reconstruction methods.

7. Conclusion and Suggestions for Future Work

7.1 Conclusion

The reconstruction of 3D torso images using stereo digital cameras is described in this thesis. The existing stereo reconstruction methods do not perform well for curved surfaces like the torso and leave behind stray data points and occlusions. The main contributions lies in solving these problems by using (1) differential localized pixel matching, (2) differential max product belief propagation registration and (3) Bezier curve moving least squares occlusion filling on rectified pair of top bottom stereo images. The most optimal stereo digital camera setup is also described in the thesis. Finally, our reconstruction method was demonstrated to be more accurate than existing stereo reconstruction methods and how it can be used clinically.

7.2 Suggestions for Future Work

There are three areas where future can be done to improve the 3D torso reconstruction system from stereo cameras.

Firstly, the system can be extended to reconstruct complete 3D torso surfaces rather than just 3D back surface image. This would require the following changes in the reconstruction method. The image acquisition process will need 3 pairs of digital cameras; their optimal setup will require calculation and each pair of image would need image rectification. The image registration step would need to be performed on each pair of images and the image reconstruction process will require stitching of three 3D torso images to obtain 1 complete 3D torso image.

Secondly, more clinical indices need to be determined and their relevance assessed so that scores other than the Cosmetic Score Index can be developed to measure external torso deformity.

Finally, the stereo reconstruction method can be further automated so that no human intervention is required. The top and bottom images of the subject were captured along with the pegboard (for image rectification) and subsequently registration was performed on these images. But, in the image reconstruction stage, the region that represents the torso for pre-processing and surface reconstruction was manually extracted. A smart edge detection algorithm that can automatically detect and extract the region that represents torso can automate this stage. This should prevent human errors and increase the clinical relevance of the system.

8. References

[1] Ajemba, P.O, Durdle, N.G, Hill, D.L & Raso, V.J., Re-positioning Effects of a Full Torso Imaging System for the Assessment of Scoliosis. Canadian Conf on Electrical and Computer Engg. 2004, Vol: 3, pp. 1483- 1486.

[2] Ajemba, P.O, Durdle, N.G, Hill, D.L & Raso, V.J., A Torso Imaging System for Quantifying the Deformity Associated with Scoliosis. Instrumentation and Measurement, IEEE Transactions on, Vol. 56, Issue 5, Oct. 2007, pp. 520 – 1526.

[3] Scharstein, D. & Szeliski, R., A Taxonomy and Evaluation of Dense Two-Frame Stereo Algorithms. International Journal of Computer Vision, 47(1/2/3), pp. 7-42, April-June 2002.

[4] Yang, Q., Wang, L., Yang, R., Stewenius, H. & Nister, D., Stereo Matching with Color-Weighted Correlation, Hierarchial Belief Propogation and Occlusion Handling. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 2, pp. 2347- 2354

[5] Sun, J., Zheng, N. & Shum, H., Stereo Matching Using Belief Propagation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol 25, Issue 7, July 2003, pp. 787 - 800

[6] Li, G. & Zucker, S.W., Stereo for Slanted Surfaces: First Order Disparities and Normal Consistency. Proc. of EMMCVPR, LNCS, 2005 - Springer

[7] Klaus, A., Sormann, M. & Karner, K., Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. Pattern Recognition, ICPR 2006, Vol. 3, pp. 15-18

[8] Li, G. and Zucker, S.W., Differential Geometric Consistency Extends Stereo to Curved Surfaces. Proc. of 9-th European Conference on Computer Vision, ECCV (3) 2006, pp. 44-57

[9] Comaniciu, D. & Meer, P., Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, May 2002 (Vol. 24, No. 5), pp. 603-619.

[10] Bleyer, M. & Gelautz, M., Graph-based surface reconstruction from stereo pairs using image segmentation. Proc. of the SPIE, Volume 5665, pp. 288-299, 2004

[11] Daruwalla, J.S. and Balasubramaniam P., Moire topography in scoliosis. Its accuracy in detecting the site and size of the curve, Journal of Bone and Joint Surgery - British Volume, Vol 67-B, 1985, Issue 2, 211-213

[12] Theologis, T.N., Fairbank, J.C. T., Turner-Smith, A.R. & Pantazopoulos, T., Early Detection of Progression in Adolescent Idiopathic Scoliosis by Measurement of Changes in Back Shape With the Integrated Shape Imaging System Scanner, Spine - 1 June 1997 - Volume 22 - Issue 11 - pp 1223-1227

[13] Assous, M., Lawson C., Douglas D.L. & Cole A.A., Reliability of Quantec Scanning in Predicting Curve Progression in Early Onset Scoliosis, Journal of

Bone and Joint Surgery - British Volume, Vol 88-B, 2006, Issue SUPP_II, 228-229.

[14] Hackenberg L, Hierholzer E, Pötzl W, Götze C, Liljenqvist U., Rasterstereographic back shape analysis in idiopathic scoliosis after posterior correction and fusion, Clinical Biomech (Bristol, Avon). 2003 Dec; 18(10):883-9.

[15] Hill DL, Berg DC, Raso VJ, Lou E, Durdle NG, Mahood JK, Moreau MJ., Evaluation of a laser scanner for surface topography, Studies in Health Technology and Informatics. 2002;88:90-4.

[16] Pazos, V., Cheriet, F., Song, L., Labelle, H. & Dansereau,J., Accuracy assessment of human trunk surface 3D reconstructions from an optical digitizing system, Medical & Biological Engineering & Computing 2005, Vol. 43, 11-15

[17] Scharstein D. & Szeliski R., High-Accuracy Stereo Depth Maps Using Structured Light, IEEE Computer Society Conference on Computer Vision and Patttern Recognition (CVPR), Vol 1, 2003, 195 – 202

[18] Chen, S.Y. & Li, Y.F., Self-recalibration of a colour-encoded light system for automated three-dimensional measurements, Measurement Science and Technology, Vol 14, 2003, 33 - 40

[19] Li, G. & Zucker, S.W., Surface Geometric Constraints for Stereo in Belief Propagation. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol 2, 2006, pp. 2355 - 2362

[20] Kolmogorov, V., Convergent Tree-Reweighted Message Passing for Energy Minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 10, October 2006.

[21] Visualization Toolkit, www.vtk.org

[22] Kumar A., Ajemba P., Durdle N. & Raso J., Pre-processing Range Data for the Analysis of Torso Shape and Symmetry of Scoliosis Patients. Studies in health technology and informatics,123, pp. 483-487, 2006.

[23] Zhang, L., Cureless B. & Seitz S. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic programming, in: International Symposium on 3D Data Processing Visualization and Transmission, Padova, Italy, 2002, pp. 24–26.

[24] Kawasaki, H. & Furukawa, R. Dense 3D Reconstruction with an Uncalibrated Stereo System using Coded Structured Light in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, Jun. 2005, pp. 107–107.

[25] Tappen, M. & Freeman, W. Comparison of Graph Cuts with Belief Propagation for Stereo, using Identical MRF Parameters in IEEE International Conference on Computer Vision, 2003.

[26] Freeman W., Pasztor, E. & Carmichael, O. Learning Low-Level Vision in International Journal of Computer Vision 40(I), 25-47, 2000.

[27] Meer, P. & Comaniciu, D. Mean Shift Analysis and Applications in The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999. Page(s): 1197 - 1203 Vol. 2

[28] Ajemba, P.O., Kumar, A., Durdle, N.G. and Raso, V.J., Range data preprocessing for the evaluation of torso shape and symmetry in scoliosis, Computer Methods in Biomechanics and Biomedical Engineering, 2009, Vol 12: Issue 6, 641 – 649

[29] Hartley, R.I. & Zisserman, A., Multiple View Geometry in Computer Vision Second Edition, Cambridge University Press, March 2004, Chapter 10.

[30] Goktepe A. & Kocaman E., Analysis of camera calibrations using direct linear transformation and bundle adjustment methods, Scientific Research and Essays, 2010, Vol. 5(9), 869-872

[31] Durdle, N.G. Thayyoor, J. Raso, V.J., An improved structured light technique for surface reconstruction of the human trunk, IEEE Canadian Conference on Electrical and Computer Engineering, 1998, Vol 2, 874 – 877

[32] Hartley, R.I., In Defense of the Eight-Point Algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19(6), 580-593, June, 1997.

[33] FastRBF Toolbox Guide –

http://www.farfieldtechnology.com/download/toolbox/FastRBF_matlab.pdf

Farfield Technology, Christchurch, New Zealand

[34] J.D. Boissonnat and F. Cazals. Smooth shape reconstruction via natural neighbor interpolation of distance functions. ACM Symposium on Computational Geometry, 2000.

[35] Kumar A. and Durdle N., A Novel 3D Torso Reconstruction Procedure using a pair of digital stereo back images. Wessex Institute of Technology (WIT) Press 2009.

[36] Kumar A. and Durdle N., Automated Calibration of Stereo Camera Systems for Imaging Scoliosis Patients, IRSSD Montreal 2010.