# Sirindhorn International Institute of Technology

# Thammasat University


**Thesis ICTES-MS-2009-05**

---

**STEREO VISION BASED CONTROL FOR A TELE-CARDIAC AUSCULTATION SYSTEM**

**Supawan Kumpituck**

# STEREO VISION BASED CONTROL FOR A TELE-CARDIAC AUSCULTATION SYSTEM

A Thesis Presented

by

**Supawan Kumpituck**

Master of Engineering
Information and Communication Technology for Embedded Systems
Program
Sirindhorn International Institute of Technology
Thammasat University
May 2010

# STEREO VISION BASED CONTROL FOR A TELE-CARDIAC AUSCULTATION SYSTEM

A Thesis Presented

By

Supawan Kumpituck

Submitted to

Sirindhorn International Institute of Technology

Thammasat University

In partial fulfillment of the requirement for the degree of

MASTER OF ENGINEERING

Approved as to style and content by

| | |
|---|---|
| Advisor and<br>Chairperson of Thesis Committee | ______________________<br>Itthisek Nilkhamhang, Ph.D. |
| Co-Advisor | ______________________<br>Asst. Prof. Toshiaki Kondo, Ph.D. |
| Committee Member and<br>Chairperson of Examination Committee | ______________________<br>Kanokvate Tungpimolrut, Ph.D. |
| Committee Member | ______________________<br>Prof. Akinori Nishihara, Ph.D. |

External Examiner : Assoc. Prof. Chuchart Pintavirooj, Ph.D.

May 2010

# Acknowledgement

The author is sincerely thankful to her advisor, Dr. Itthisek Nilkhamhang, and co-advisor, Asst. Prof. Dr. Toshiaki Kondo for their guidance and support throughout the work of this thesis to develop an understanding of the subject. Without their persistent help this thesis would not have been possible.

She also would like to express her deepest appreciation to Assoc. Prof. Dr. Waree Kongprawechnon for giving her encouragement and an opportunity to attend the research project of Industrial Control and Automation (ICA) Laboratory at Thailand National Electronics and Computer Technology Center (NECTEC) and she is grateful to thank Dr. Kanokvate Tungpimolrat for a convenient cooperation of ICA laboratory. Likewise, she is really thankful to Prof. Akinori Nishihara for a kindly respondse and suggestion. In addition, she would like to thank Assoc. Prof. Dr. Chuchart Pintavirooj for his commentary which are advantageous for her works.

Moreover, she wish to express her special thanks to the scholarship grantor, Thailand Advanced Institute of Science and Technology - Tokyo Institute of Technology (TAIST-Tokyo Tech), National Science and Technology Development Agency(NSTDA), Tokyo Institute of Technology (Tokyo Tech),and Sirindhorn International Institute of Technology(SIIT), Thammasat University (TU). Lastly, she offers her regards and blessings to all of those who supported her in any respect during the completion of her works.

# Abstract


Stereo Vision Based Control for a Tele-Cardiac Auscultation System

May 2010

by

Supawan Kumpituck

B.Eng, Sirindhorn International Institute of Technology

Thammasat University, 2006

Cardiac auscultation skills are essential tools that help to obtain a precise cardiovascular evaluation. Although, medical skills of using stethoscope for the cardiac auscultation are still required, in large-scale medical services, the number of the doctor seems to be small. As a result, a tele-cardiac auscultation is proposed to address these inadequate problems. The thesis presents an application of a stereo vision system to the tele-cardiac auscultation system which is aimed at overcoming the problem of insufficient number of doctors compare with the higher number of patients. Stereo vision techniques are able to analyze multiple images of the same scene captured to recover three-dimensional structural information, as a result, the position of stethoscope location on the human body in the tele-diagnostic system can be estimated.

The stereo vision system simulates human vision which the reconstruction of points in three-dimensional space can be done by using the knowledge of a camera calibration, an image rectification, an epipolar geometry, and a correspondence analysis. In addition, the correspondence analysis is the main issue of this study. In this thesis, the procedure of the correspondence analysis has been seperated into two steps, i.e., the human body edge detection and matching, and the human body region matching. A comparative study of four image matching techniques to the correspondence analysis of a stereo image pair has also been presented, including the sum of absolute difference (SAD) matching, the sum of square difference (SSD) matching, the zero mean normalized cross correlation (ZNCC), and the phase only correlation (POC). In this robustness study, an integrated input environment has been set up to acquire the similarity measurement. The precision of the correlation result and the reconstruction of points of these algorithms are the significant factors of selecting the stereo matching algorithm that is appropriate to the proposed tele-cardiac auscultation system.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

A stethoscope is used as a fundamental device for auscultation, i.e., the technical term for listening to the internal sounds of the body, and to use the stethoscope, the substantial clinical skill and experience of doctor are required. However, nowadays, the demand for doctors may increase as new medical developments allow more diseases to be treated and public awareness of what is possible is raised. Moreover, there appears to be a changing relationship between the medical profession and their patients which patients are becoming better informed and more demanding of new treatments. As a result, in a large-scale medical services, the future supply of doctors to match the anticipated demands may be insufficient. Since the cardiac auscultation for the heart valve disease diagnostic also requires those skills and experiences of the doctor, furthermore, the number of skillful doctors is possible to be insufficient due to the larger number of patients, therefore, the tele-cardiac auscultation system is proposed in this work to lessen the required number of doctors by helping a doctor to remotely auscultate and record heart sound signals for diagnosis over distances.

The entire system is composed of an electronic stethoscope mounted robot arm which can be remotely controlled over the internet communication as an overview of the system in Fig. 1.1. In general, computer stereo vision systems have been broadly used in various kinds of application, for example, in machine vision, robotics, and industry. Specifically, this work presents a stereo vision system which is applicable to the proposed tele-cardiac auscultation system. The stereo vision multiples images of the same scene captured at different positions which provide slightly dissimilar viewpoints. By detecting the differences of two or more images the three-dimensional (3-D) structural information on the scene can be recovered. A stereo correspondence analysis is a problem that has been studied in computer vision. The method intends to match each feature in the left image with its counterpart which is the projection of the same scene element in the right image. Consequently, the heart sound recording positions on the human body and also the length from the initial position of robot arm to the patient's chest can be estimated by using these 3-D stereo vision techniques.

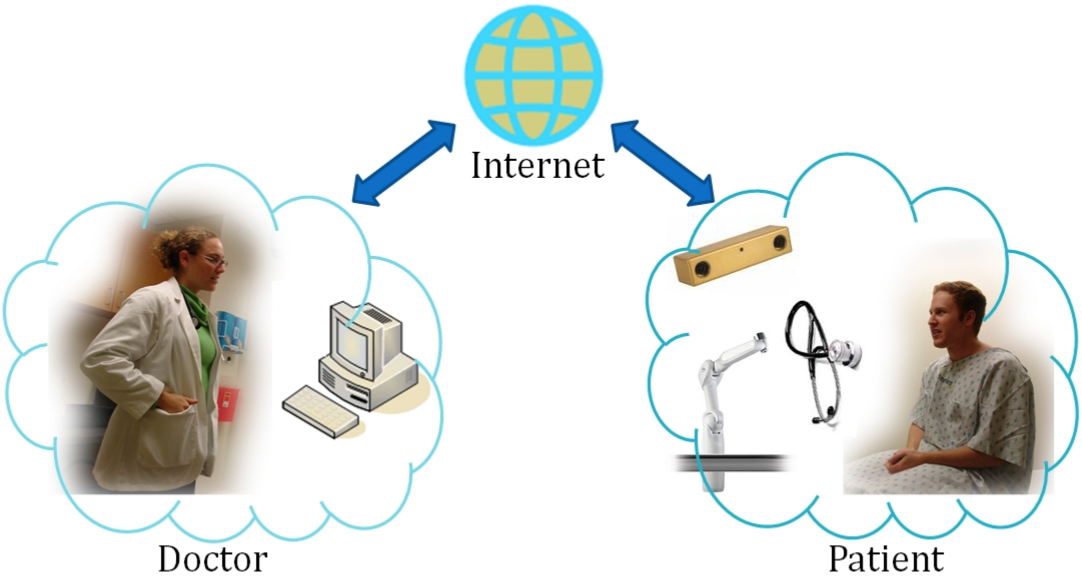



Figure 1.1 An overview of a tele-cardiac auscultation system.

### 1.1 Motivation

Cardiac auscultation is considered as an important tool for doctors in obtaining a precise cardiovascular evaluation. Although the cardiac auscultation using the stethoscope still requires the skill of doctor, the supply of doctors to match the anticipated demands may be insufficient for a large-scale medical services with a large number of patients. By the aim of the proposed system, since the doctors can remotely control the electronic stethoscope which the heart sound signals can be recorded at required positions and those signals can be analyzed, therefore, the system can improve the problem of insufficient number of skilled doctors. Furthermore, the proposed stereo vision based system will simulate human binocular vision which assist the position measurements to be more precise.

### 1.2 Heart Sounds and Measurements

Normal heart sounds are associated with heart valves closing which cause S1 and S2 sounds as shown in Fig. 1.2 S1 is from the closure of the mitral and tricuspid valves after blood has returned to the heart. S2 is, on the other hand, created by the closure of the aortic and pulmonary valves as blood exists from the heart.

The heart valve disease is suspected when the murmur sound detected in systole period and in diastole period as seen in Fig. 1.3. These occur when the valve does not widely opened, or tightly closed [1].

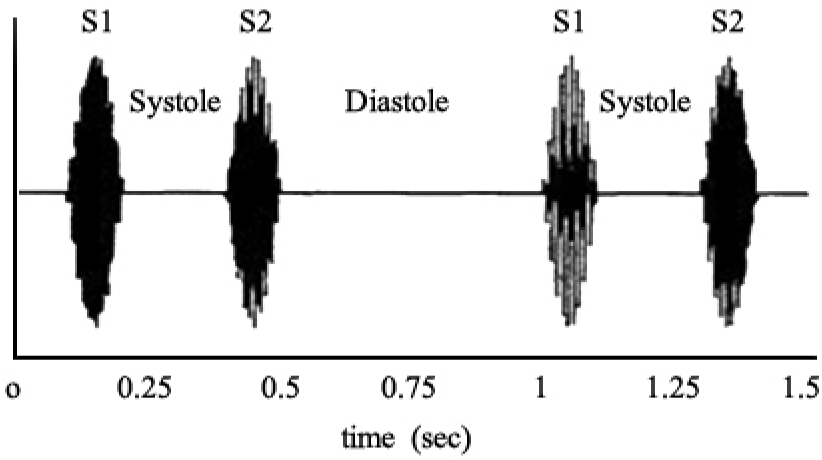



Figure 1.2 The normal heart sound.

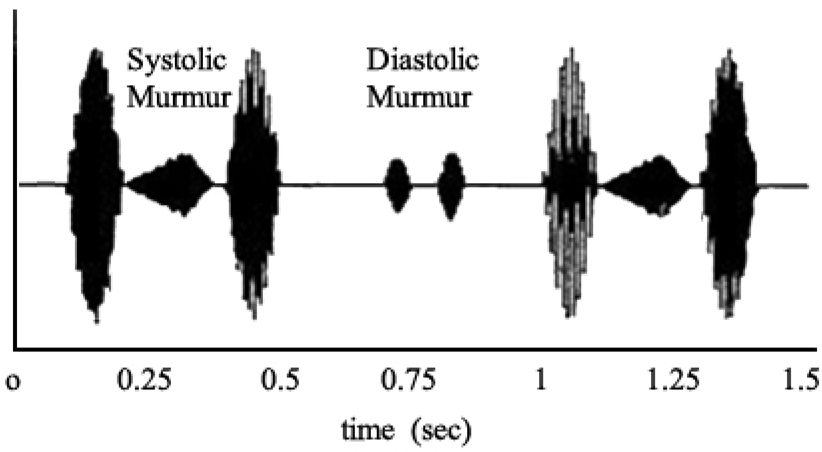



Figure 1.3 The murmur in heart sound.

Typically, there are different features of heart sounds at each recording position. In Fig. 1.4, there are five locations that the stethoscope should be placed to observe the sounds [2].

The aortic area, pulmonary area, mitral area, and tricuspid area are on the surface of the chest where the cardiac sound is auscultated. The opening and closing of the valves is usually much less loud than the sound of the blood rushing through the valve and colliding with the subsequent barrier. Therefore, the auscultation that determine function of a valve is usually not performed at the position of the valve, but at a downstream position where the doctor can best hear the blood colliding after the valve is closed. The features of the cardiac auscultation signal are the variables to determine whether the heart valves work properly or not.

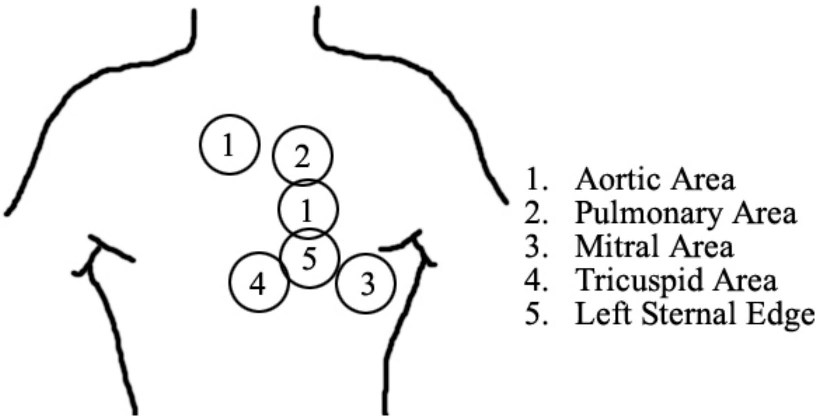



Figure 1.4 The positions of the stethoscope.

## 1.3 Stereopsis

Stereopsis is the process in visual sense leading to an immediate perception of depth from the two slightly different projections of the world onto the retinas of the two eyes. The different perspectives of two eyes lead to slight relative displacements of objects in the two monocular views of scene. Stereopsis is commonly referred to as depth perception.

In computer stereo vision, for example, there are two cameras which can take pictures of the same scene, but they are separated by a distance as well as human's eyes. By shifting the two images together over top of each other to find the parts that match, the computers are able to compare the images which the shifted amount is called the disparity. The disparity at which any interested objects in the image best match is used by the computer to calculate their distance, i.e., depth estimation.

## 1.4 Outline

### Chapter 2

This chapter reviews the stereo vision system which simulates human binocular vision by analyzing multiple images that provide slightly dissimilar viewpoints of the same scene captured at an interested point. By detecting the differences of two or more images, the 3-D structural information on that scene can be recovered. Moreover, a number of steps is involved in recovering 3-D structure from two camera views including camera calibration, image rectification, and correspondence analysis.

### Chapter 3

In this chapter, the data flow of the stereo vision system which will be applied to the tele-cardiac auscultation system is shown. The flow is composed of two main processes for determining the depth image of an observed region, i.e., human chest which are an image preprocessing process, i.e., rectification and edge extraction, and a matching process which is also divided into an edge matching and a region matching.

**Chapter 4**

This chapter shows the results of this work. At the beginning, a comparative study of the stereo matching techniques is presented which shows that each matching technique has different capability under different environment. This chapter also presents the outcome of each process from the flow of stereo vision system shown in previous chapter. Furthermore, the results of the correspondence analysis are initially obtained by manually selected to show that an interested point in the image plane can be converted into the plane of world coordinate. Lastly, the 3-D recontructed results are presented after the input images have been passed through the system to the final step of an automatic correspondence analysis.

**Chapter 5**

The last chapter of this thesis provides the summary of this work which presents the stereo vision system that is applicable to the proposed tele-automatic auscultation system. Throughout this work, disparity map image of the human body optained from the correspondence analysis can be converted to the 3-D structure that is the conversion is done from an image coordinate to world coordinate.

# Chapter 2

# Stereo Vision System

## 2.1 Stereo Vision

Computer stereo vision system has been broadly used in industry and agriculture. Autonomous surveillance and monitoring of structures is one of the studies that employ stereo based tracking systems [3], [4]. The system simulates human binocular vision by analyzing multiple images of the same scene captured at different positions which provide slightly dissimilar viewpoints. By detecting the differences of two or more images the 3-D structural information on the scene can be recovered.

From a computational standpoint, the stereo vision system must solve two problems [5]:

$The\ correspondence\ problem$: determining which parts of the left and right images are projections of the same scene element.

$The\ reconstruction\ problem$: the geometry of the stereo system and the corresponding parts of the point that is interested from the left and right image (two-dimension (2-D) view) are converted to the 3-D map of that point.

eCamera calibration can be done to establish the relationship between the 3-D coordinates and their corresponding 2-D image coordinates. There is a number of steps associated with recovering 3-D structure from two camera views including camera calibration, image rectification, and disparity computation [6], [7].

### 2.1.1 Epipolar Geometry

Epipolar geometry is usually motivated by considering the search for corresponding points in stereo matching [8]. The general epipolar geometry is shown in Fig. 2.1. The relative position of both cameras is known as $O_l$ and $O_r$ which indicate the optical centers of each camera. The straight line connecting both optical centers is called baseline. The point $P$ observed by the two cameras and the two corresponding light rays through the optical centers form an epipolar plane. The lines where the epipolar plane intersects the image planes are conjugated epipolar lines. The epipolar constraint is that, for a given point in the left image plane, its possible matches in the right image plane must lie on a line. Moreover, the constraint is of course symmetric and, for a point in th right image plane, its possible matches in the left image plane must lie on a line which is an epipolar line through the epipole $e_l$ and $e_r$ in Fig. 2.1. Since the corresponding points all lie on these conjugated epipolar lines, therefore, the dimension of the search space can be reduced from two dimensions to one [9].

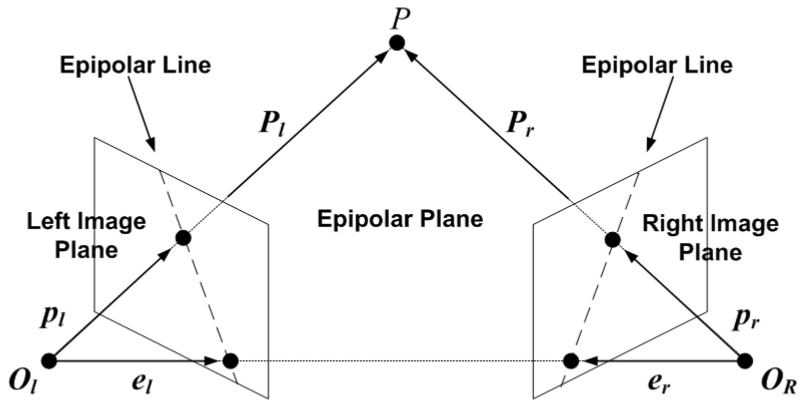



Figure 2.1 The epipolar geometry.

The relation between $P_l$ and $P_r$ which is the projection of $P$ in the left and right image frames respectively is given by

$$P_r = R(P_l - T), \tag{2.1}$$

where $T = (O_r - O_l)$ is the translation vector and $R$ is the rotation matrix.

The relation between 3-D points and their projections can be defined in the equations:

$$p_l = \frac{f_l}{z_l} P_l, \tag{2.2}$$

$$p_r = \frac{f_r}{z_r} P_r, \tag{2.3}$$

where $f_l$ and $f_r$ are the focal lengths and $Z_l$ and $Z_r$ are the length from projected point to the center of baseline [5].

### 2.1.2 Camera Calibration

Camera calibration is an essential preliminary step of stereo vision system where the geometric information of camera setup is found. The values of the intrinsic and extrinsic parameters can be estimated. Two types of parameters, i.e., intrinsic and extrinsic parameter, will be necessary for recovering the exact location of objects in 3-D space.

$Intrinsic\ parameters$: parameters that are necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame, for example, the focal length of the lens, the position of the principal point, and the geometric distortion introduced by the optics.

$Extrinsic\ parameters$: parameters that identify uniquely the transformation between the unknown camera reference frame and the known world reference frame. The parameters are the transition, $T$ which is a vector between the relative positions of the origins of the two reference frames and the rotation, $R$ which is a matrix that brings the corresponding axes of the two frames into alignment with each other [5], [6].

To obtain both intrinsic and extrinsic parameters, existing calibration procedures for cameras can be employed, including the procedures developed by Zhengyou Zhang [10], and Jean-Yves Bouguet [11]. In these procedures, a planar checkerboard is used as a calibration pattern and the cameras are calibrated by using the camera calibration toolbox [11]. In order to perforem the calibration process, a set of 3-D reference points is needed. It can be achieved by using some sort of regular pattern such as the one shown in Fig. 2.2.

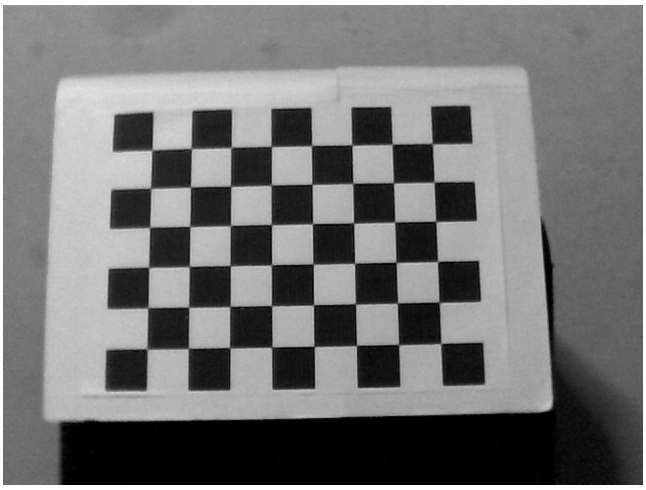


Figure 2.2 Calibration pattern.

In [11], the steps of calibration are as follows:

- Image acquisition

- Checkerboards grid corner extraction

- Parameter computation

The calibration steps are composed of left camera calibration, right camera calibration, and stereo system calibration. First, the images of calibration pattern are taken from each camera that needs to be calibrated, for example, twenty images from both left and right cameras are shown in Fig. 2.3 and Fig. 2.4, respectively. Then the checkerboards corners in each image are extracted as shown in Fig. 2.5 then the intrinsic and extrinsic parameters are calculated. Intrinsic parameters can be individually estimated for each camera. After that, the extrinsic parameters will be estimated by a method called stereo calibrating which used the initial values of intrinsic parameters from previous processes.

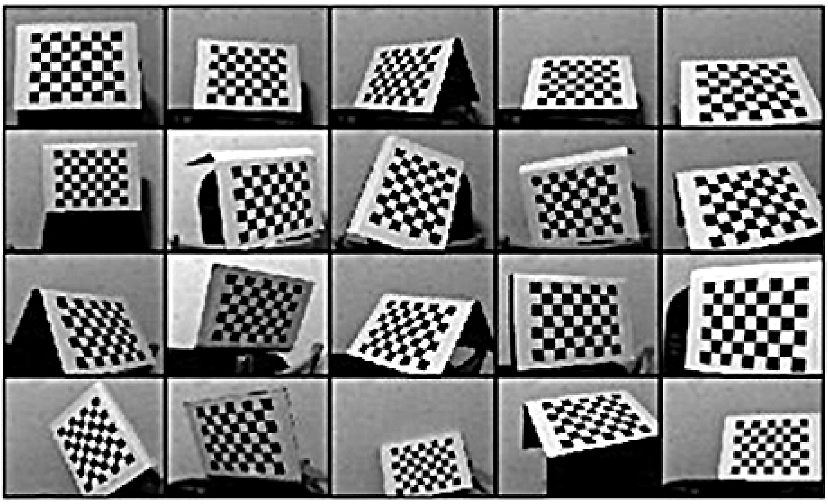


Figure 2.3 Calibration images from left camera.

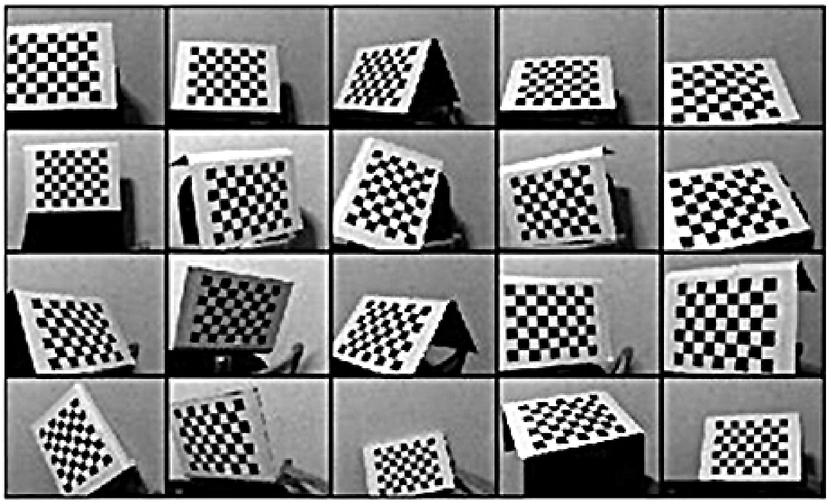


Figure 2.4 Calibration images from right camera.

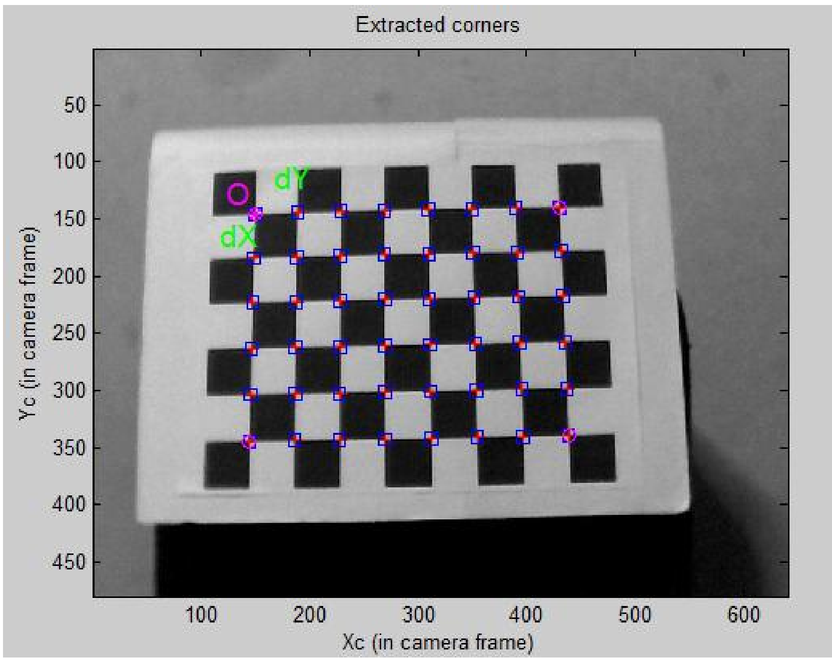



Figure 2.5 The calibration pattern with extracted corners.

### 2.1.3 Image Rectification

After all intrinsic and extrinsic camera parameters are calculated they can be used to rectify images according to the epipolar constraint. Image rectification process is necessary to transform each image such that pairs of conjugate epipolar lines and features are aligned along corresponding horizontal or vertical axes. Therefore, the corresponding point searching process can be done only in 1-D instead of 2-D. In Fig. 2.6, the corresponding points of the rectified image pair are adjusted to align on the same horizontal line.

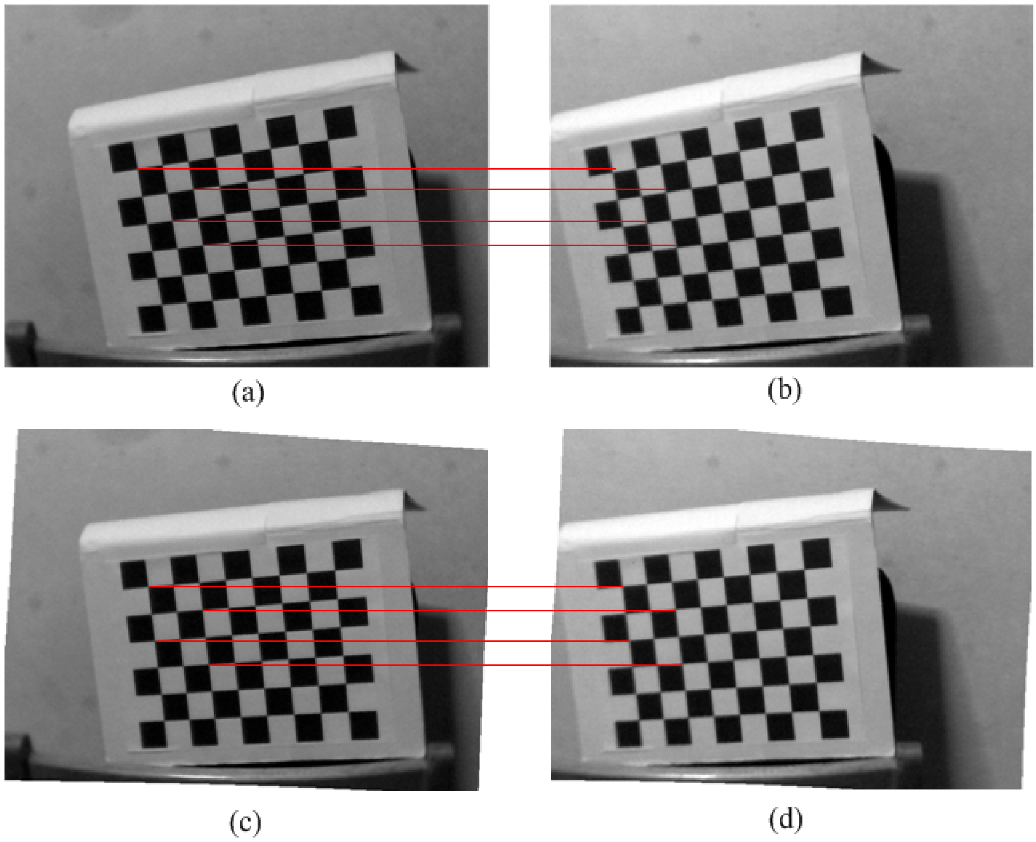



Figure 2.6 Image pairs from rectification process, (a) Unrectified left image, (b) Unrectified right image, (c) Rectified left image, and (d) Rectified right image.

### 2.1.4 Correspondence Analysis

In correspondence analysis, the matches between projections of 3-D points in the stereo pairs are found and the disparity between these two corresponding points can be computed. Many correspondence algorithms use the advantage of the epipolar constraint which states that the corresponding point in the right image of a given pixel in the left image must lies on the conjugated epipolar line. The constraint reduces the search space of the correspondence point matching algorithms. In addition, it is very convenient to use the images that have been rectified because the conjugated epipolar lines of both images are horizontally aligned and the matching process is reduced from 2-D in the image plane to 1-D along these epipolar lines. Mainly, the algorithms can be divided into area based and feature based.

In area based algorithms, the correspondence problem is solved for every single pixel in the image. The elements to match are image windows of fixed size and the similarity is measured between windows in two images. A corresponding pair is given by the window that maximizes the similarity within a search region. These algorithms result in dense depth maps as the depth is known for each pixel. But selecting the right block size is difficult because a small neighbourhood will lead to less correct maps but short run times whereas a large neighbourhood leads to more exact maps at the expense of long run times.

Feature based algorithms on the other hand try to match the extracted features which should be unique within the images, e.g., edge, and corner, between each image pair. The resulting maps will be less detailed as the depth is not calculated for every pixel as in area based. Since it is much more unlikely to match a feature incorrectly because of its detailed description, feature based algorithms are less error sensitive and result in very exact depth maps [7].

Aside from these algorithms, there are some phase based algorithms which the correspondence is analyzed in frequency domain instead of spatial domain. In this work, four correspondence algorithms have been reviewed, i.e., sum of absolute difference (SAD), sum of squared difference (SSD), zero mean normalized cross correlation (ZNCC), and phase only correlation (POC).

### 2.1.5 3-D Reconstruction

Consider the problem of estimating depth from two images taken from a pair of stereo cameras. The most common approach for doing so is stereopsis (stereo vision), in which depths are estimated by triangulation using the two images [12]. After the correspondence between points in an image pair and the intrinsic and extrinsic parameters are known, the location of the point in 3-D can be computed by using the triangulation with the correspondence points and the camera projections, $p_l$ and $p_r$. The stereo geometry shown in Fig. 2.7 is used to recover the position of $P$ from its projections by using equations:

$$x_l = f\frac{x_l}{z_l} \; or \; X_l = \frac{x_l Z_l}{f}, \tag{2.4}$$

$$x_r = f\frac{x_r}{z_r} \; or \; x_r = \frac{x_r z_r}{f}, \tag{2.5}$$

In general, the two cameras are related by the transformation in Eq. 2.1. By using $Z_r$ = $Z_l$ = $Z$ and $X_r$ = $X_l - T$, The depth information, $Z$, can be recovered from Eq. 2.6 and Eq. 2.7.

$$\frac{x_l Z}{f} - T = \frac{x_r Z}{f}, \tag{2.6}$$

$$Z = \frac{Tf}{d}, \tag{2.7}$$

where $d = x_l - x_r$ is the difference in the position between the corresponding points in the two images called disparity.

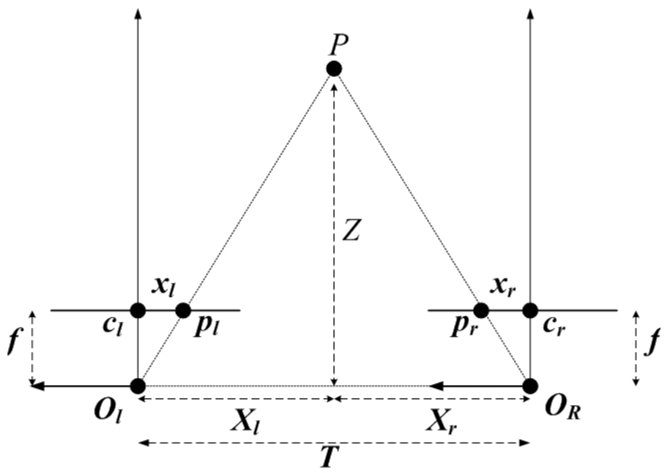



Figure 2.7 The elementary epipolar geometry.

However, since the parameters are approximately calculated and there may be errors in the location of the corresponding features, the two projecting rays will not actually intersect in space and their intersection can be estimated as the point of minimum distance from both rays which is the midpoint $P$ of line segment being perpendicular to both rays as shown in Fig. 2.8. The parametric representation of the line passing through $P_1$ and $P_2$ is given by

$$P(c) = P_1 + c(P_2 - P_1), \tag{2.8}$$

where $P_1 = a_0 p_l$ and $P_2 = T + b_0 R^T p_r$.

The desired midpoint $P$ can be computed when $c = \frac{1}{2}$ [5].

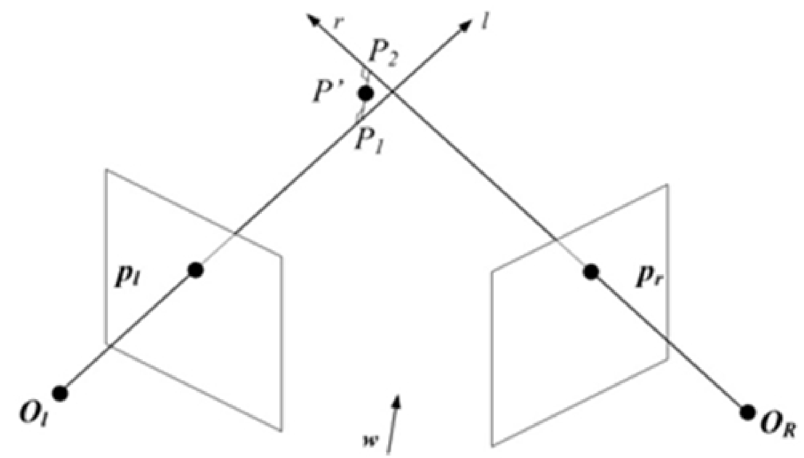



Figure 2.8 Triangulation with nonintersecting rays.

# Chapter 3

# Application to Tele-Cardiac Auscultation System

The diagram shown in Fig. 3.1 is the data flow of the stereo vision system which can be applied to the tele-cardiac auscultation system. There are three main steps for determining the depth image of an observed region, i.e., human chest. The first step is the image pre-processing including the camera calibration, the rectification, and the edge feature extraction. The second step does stereo matching which is also divided into an edge matching and a region matching. Then the last step is the reconstruction. Initially, the pre-processing step takes two images of human body obtained from the stereo camera and then pass the input images into rectification process by using the parameters of the camera previously obtained from the camera calibration process. The rectified images are afterwards converted into edge images then the edge and region disparity map are obtained in the corresponding analysis process. After reconstruction, the final result is the 3-D map of human body which the distance from the camera to the chest can be estimated.

## 3.1 Rectification

Lenses often cause distiortions in raw images, for example, straight lines in the scene will often appear curved in the raw images as shown in Fig. 3.2 (a) and (b). Rectification is the process of correcting input images from the distortions of the lenses. In addition, rectified imags will be corrected so that the rows of mages digitized from horizontally displaced cameras are aligned as can be seen in Fig. 3.2 (c) and (d), and similarly that the columns of images obtained from vertically displaced cameras are aligned. Without this step, searching along the rows and columns may produce the incorrect results. As a result, the input images will be previously rectified before entering the process of corresponding analysis.

## 3.2 Edge Feature Extraction

In typical images, edges characterize object boundaries and are therefore useful for segmentation, registration, and identification of objects in a scene. In most image processing applications used edge detection to detect outlines of an object and boundaries between objects and the background in the image. That is meant to extract salient features of the scene.

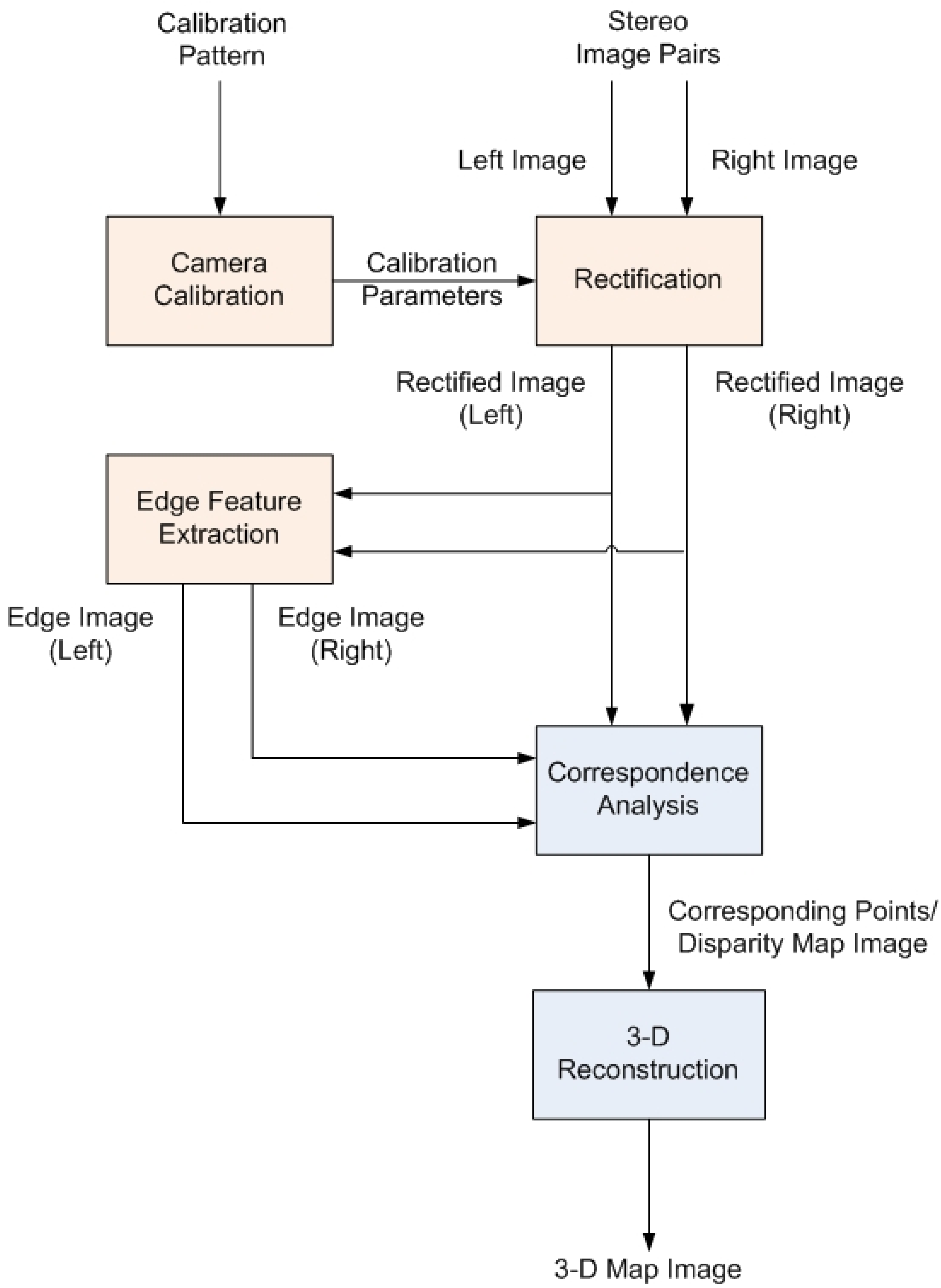



Figure 3.1 The diagram of the stereo vision system for detecting human body position.

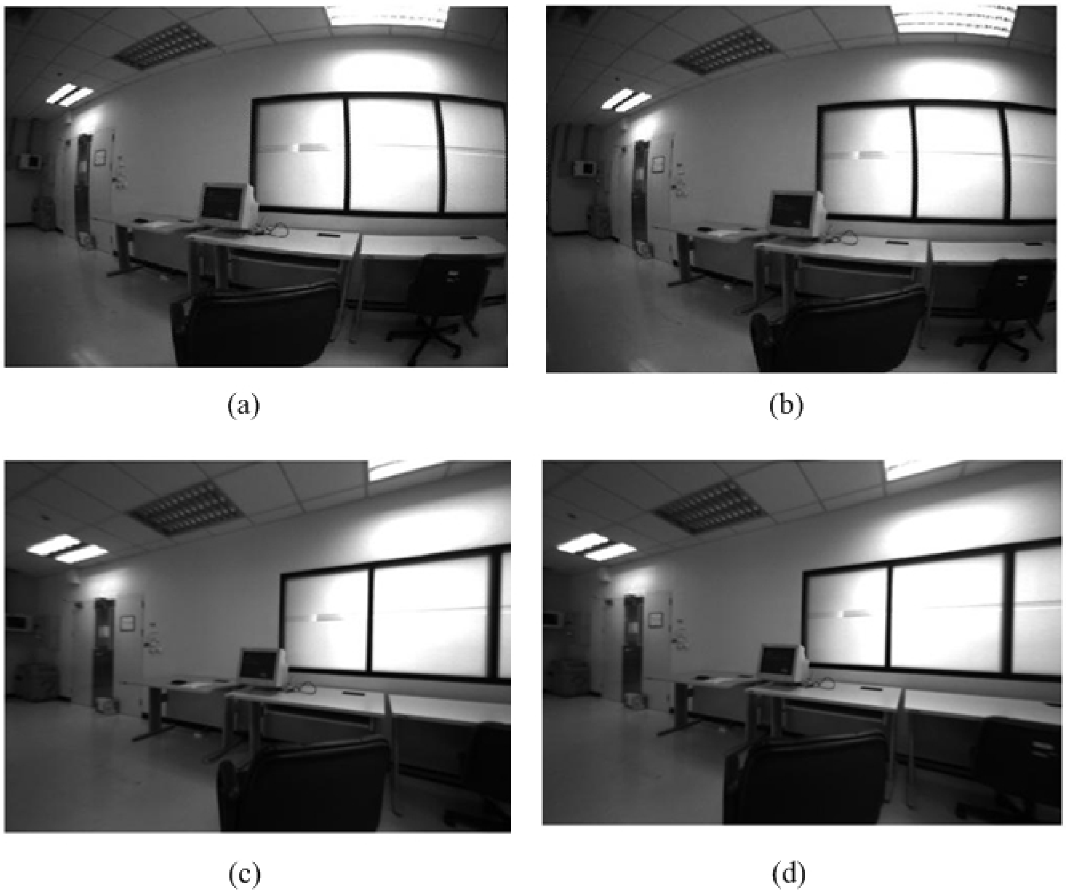



Figure 3.2 The raw and rectificied images from the stereo camera, (a) Unrectified left image, (b) Unrectified right image, (c) Rectified left image, and (d) Rectified right image.

The salient features are expected to be the boundaries of objects that tend to produce sudden changes in the image intensity. The goal of edge detection is to convert a 2-D image into a set of curves, i.e., a line drawing.

Edge is a boundary between two regions of relatively uniform internsity, indicated by a strong gradient of a discontinuity in the intensity function. Consequently, the intisity profile in a direction perpendicular to the edge will behave as a step function, on the other hand, the intensity can vary smoothly in the direction along the edge. Edge detection is advantangeous in correspondence analysis in term of it allows matching on the changes in the brightness rather than the absolute values of the pixels in the images. The edge that can be detected in the input image is the boundary of human body seperated from the background. In this work, the canny edge detection algorithm which is known to many as the optimal edge detection technique is used to extract the desired boundary.

### 3.2.1 Canny Edge Detection

The Canny edge detection operator [13] is one of the standard edge detection methods at present. John F. Canny intented to enhance the many edge detectors that were already existed at the time he started his work. His ideas and methods which he was very successful in achieving his goal can be found in the paper, ”A Computational Approach to Edge Detection” [14]. In his work, he suggested an edge detection algorithm that is optimal with regards to the following criteria:

$Ability\ to\ handle\ noise$: The probability of detecting real edge points should be maximized while the probability of falsely detecting non-edge points should be minimized. This corresponds to maximizing the signal-to-noise ratio which the output should not be overly sensitive to noise.

$Accurate\ location\ of\ the\ true\ edge$: The edge points should be well localized which the detected edges should be as close as possible to the real edges.

$Uniqueness\ of\ the\ response$: If there is only one real edge in a given neighborhood, then there should not result in more than one detected edge.

Based on these criteria, the canny edge detector first smoothes the image to eliminatenoise. It then finds the image gradient to highlight regions which the edges should be marked where the gradients of the image has large magnitudes. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum, i.e., only local maxima should be marked as edges. Then, the potential edges are determined by thresholding. Although, the edge pixels remaining after the non-maximum suppression step are marked with their strength, many of them will probably be true edges in the image, but some may be caused by noise or color variations for instance due to rough surfaces. The simplest way to discern between these would be using a threshold, so that only edges stronger than a certain value would be preserved.

The Canny edge detection algorithm uses double thresholding which edge pixels stronger than the high threshold are marked as strong, on the other hands, edge pixels weaker than the low threshold are suppressed and edge pixels between the two thresholds are marked as weak. Then, the gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels which the strong edges can immediately be included in the final edge image while the weak edges can either be due to true edges, noise, or color variations which will be included if and only if they are connected to strong edges. Fig. 3.3 shows the output edge images that are extracted by using Canny method with different thresholding. Comparing between Fig. 3.3 (b) and (c), while the high threshold remains the same, the edge pixels that are weaker than the low threshold are suppressed which the number of suppressed pixels are lower when the low threshold is set to be smaller as seen in Fig 3.3 (a). Furthermore, comparing beteen Fig. 3.3 (c) and (d), while the low threshold remains the same, since the edge pixels that are stronger than the high threshold will be immediately added into the extracted edge output, therefore, the number of extracted edges are higher when the high threshold is set to be smaller as seen in Fig 3.3 (c).

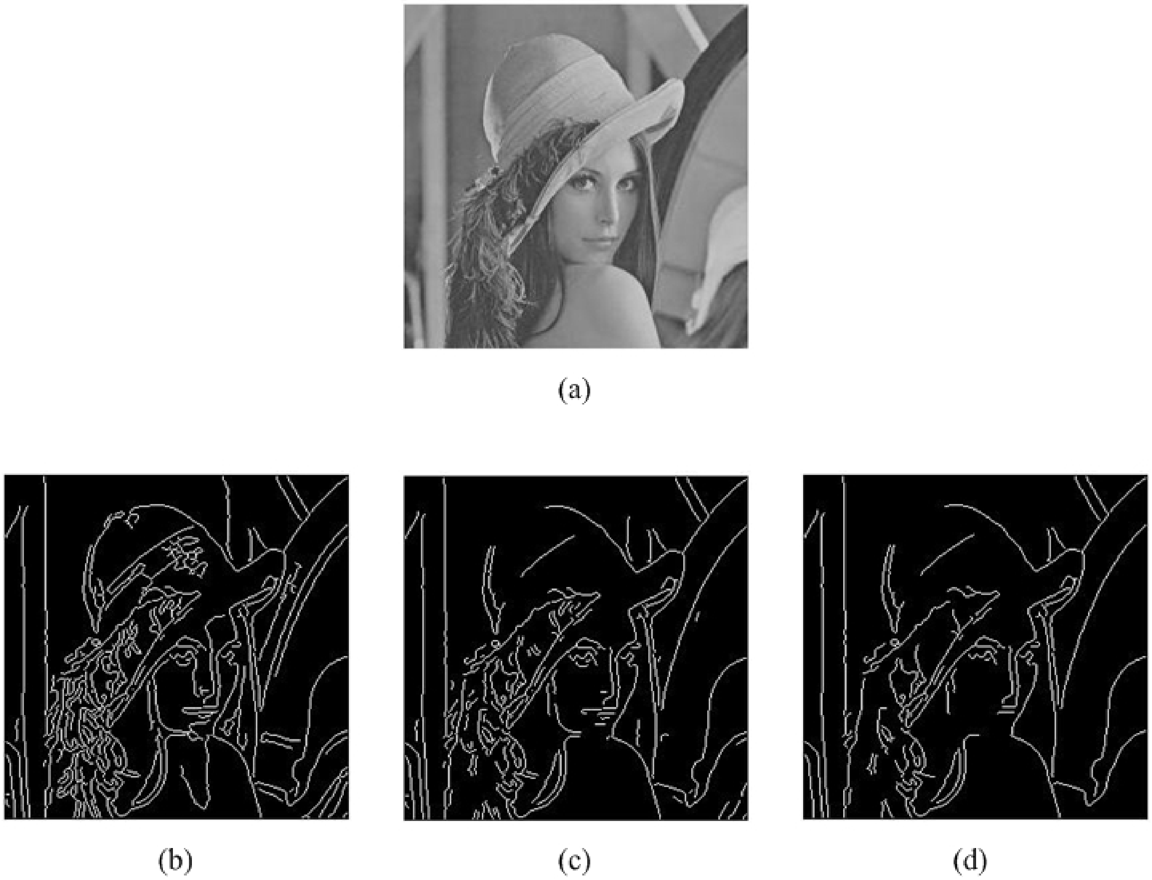


Figure 3.3 The original image (a) and the edge image generated by the Canny edge detector with,(b) 0 low threshold and 0.2 high threshold, (c) 0.15 low threshold and 0.2 high threshold, (d) 0.15 low threshold and 0.3 high threshold.

### 3.3 Correspondence Analysis for Tele-Cardiac Auscultation System

Stereo correspondence analysis is a problem that has been studied in computer vision. The method intends to match each feature in the left image with its counterpart which is the projection of the same scene element in the right image. The correspondence analysis for the tele-cardiac auscultation is divided into two categories, i.e., edge matching and region matching.

In addition, the result obtained from this step is a disparity range which is the range of pixels that the stereo algorithm searches in order to find the best match of the corresponding points. Initially, the disparity range of the edge points will be obtained. After that, the disparity value of the observed region will be calculated. Correspondence searching in this system is done along horizontal direction by applying standard block matching to the input image. In the system , the maximum disparity defines the closest position of a reiogn that is to be determined. When the disparity range is reduced, it allows the system to run faser and reduce the chance of mismatching, therefore, in this work the disparity range is limited to be within the boundary of the human body which is extracted from the edge detection process.

### 3.3.1 Block Matching Algorithm

The block matching method is one of the most popular local methods because of its simplicity in implementation [15]. The idea behide block matching is to divide the image into a matrix of macro blocks that are then compared with corresponding block and its adjacent neighbors in the other stero image of a pair [16]. Furthermore, a similarity check is done between two equal sized blocks (n x m-matrices) in the left and the right image (area-based stereo) [17].

In addition, to estimate the disparity of an interested point (pixel) in the reference image, for example, left input image, firstly, the method defines a reference block with a pre-defined window size surrounding that point and compare this neighborhood to a number of neighborhoods in the other image. The closest matched block is found within a search range in the right image, using a pre-specified matching criterion. The disparity of the point being evaluated is the relative displacement between the reference block and the closest matched block. The matching criteria for the similarity measurement considered in this work are the sum of absolute differences (SAD), the sum of squared differences (SSD), the zero mean normalized cross correlation (ZNCC), and the phase only correlation (POC).

### 3.3.2 Stereo Matching Criteria

$Sum\ of\ absolute\ difference\ (SAD)$: SAD computes the intensity differences for each center pixel in a window by subtracting pixels within a square neighborhood between the reference image $I_1$ and the target image $I_2$ then aggregating absolute differences within the square window as in Eq. 3.1. The minimum SAD is chosen to be the best matching pixel. Correspondence analysis techniques related to SAD are mentioned in [18], and [19].

$$SAD = \sum_{(i,j)\in W} |I_1\left(i,j\right) - I_2\left(x+i, y+j\right)| \tag{3.1}$$

$Sum\ of\ square\ difference\ (SSD)$: SSD computes the intensity differences for each pixel in a window then the differences are squared and aggregated within the square window as in Eq. 3.2. The minimum SSD is chosen to be the best matching pixel. Correspondence analysis technique related to SSD is mentioned in [19].

$$SSD = \sum_{(i,j)\in W} \left[I_1\left(i,j\right) - I_2\left(x+i, y+j\right)\right]^2 \tag{3.2}$$

$Zero\ mean\ normalized\ cross\ correlation$ (ZNCC): ZNCC computes the degree of similarity between a template and a portion of the image [20]. The ZNCC between two images is defined as in Eg. 3.3. Maximum correlation score is chosen to be the best matching template. Correspondence analysis technique related to ZNCC is mentioned in [20].

$$ZNCC = \frac{\sum_{(i,j)\in W}[I_1(i,j) - \mu(I_1)]\cdot[I_2(x+i,y+j) - \mu(I_2c(x,y))]}{\sqrt{\sum_{(i,j)\in W}[I_1(i,j) - \mu(I_1)]^2}\cdot\sqrt{\sum_{(i,j)\in W}[I_2(x+i,y+j) - \mu(I_2c(x,y))]^2}}, \tag{3.3}$$

where $\mu(I_1)$ and $\mu(I_2c(x,y))$ are the mean intensity values computed on $I_1$ and $I_2c$ sub-pixel of image $I_2$ at position $(x,y)$.

$Phase\ only\ correlation\ (POC)$: POC is a frequency domain approach to estimate the similarity and translative movement between two images. POC function is calculated from the phase information in 2-D Discrete Fourier Transforms (2-D DFTs) of given images as in Eg. 3.4.

$$POC = \frac{I_1 I_2^*}{|I_1 I_2^*|}, \tag{3.4}$$

where $I_1$ is a 2-D Discrete Fourier transform of input Image $I_1$, and $I_2^*$ is a conjugation of a 2-D Discrete Fourier transform of input Image $I_2$. The maximum correlation score is chosen to be the best matching output. Image matching applications based on POC are mentioned in [21], and [22].

### 3.3.3 3-D Reconstruction

The reconstruction process converts disparity map image optained from the correspondence analysis to 3-D map. The conversion is done from an image coordinate to world coordinates. Each pixel in the disparity map image has a label of disparity range assigned to it. In addition, not only the disparity range of each pixel but the focal length and the baseline of the camera are also have an important role in image reconstruction. The final result from this step is the 3-D plot of human body which the distance from the camera to the body is obtained.

# Chapter 4

# Results and Discussion

In this chapter, the results of the algorithms that have been studied in this work will be presented. At the beginning, a comparative study of the stereo matching techniques is presented [23] which shows that each matching technique has different capability under different environment. This chapter also presents the outcome of each process from the flow of stereo vision system which previously described in Chapter 3. Furthermore, the results of the correspondence analysis are initially obtained by manually selected [24] to show that an interesting point in the image plane can be converted into the plane of world coordinate. Finally, the 3-D recontructed results are presented after the input images have been passed through the system to the final step of automatic corresponence analysis.

## 4.1 Comparative Study of Stereo Matching Techniques

In this work, stereo matching algorithms, namely, SAD, SSD, ZNCC and POC have been implemented under different environment of input images. The similarity between two input images can be estimated by calculating the correlation values using SAD, SSD, and ZNCC which are based on the images spatial domain and POC which is based on the images frequency domain. For each input image pair, using MATLAB, the comparisons are made among the correlation results of each algorithm. In this experiment, the image of size 256x256 pixels shown in Fig. 4.1 is used as an original input. To evaluate the result of each input pair, the range of the correlation value is between -1 and 1 which states the most dissimilarity, -1 to the most similarity results, 1. In addition, there are four different circumstances that are concerned in the experiments of the similarity measurement which are the variation in image sharpness, noise, illumination, including, the effect of geometric transformation.

To estimate the correlation values between the different sharpness inputs, the average, median, and Gaussian filters are applied to one of the original input as shown in Fig. 4.2. Table 4.1 shows the correlation results between the original and each distorted input. SSD generates the highest correlation peak which means the algorithm is less sensitive to these kinds of distortion. In contrast, POC is the most sensitive to this experiment.

In addition, the correlation values between original and noisy images are estimated by

Figure 4.1 The original input image.

adding some noises to the original image, e.g., salt and pepper noise, Gaussian noise, and periodic noise, as shown in Fig. 4.3 and the correlation results are in Table 4.2. From the results SSD works better than others under salt and pepper and Gaussian noise while POC works better under the periodic noise condition.

Moreover, the correlation values between original and illumination variation images are estimated by interfering the original input with uniform shading, Gaussian shading, periodic shading and mixed Gaussian and periodic shading as in Fig. 4.4. The results in Table 4.3 show that ZNCC works better than the others under uniform shading environment. In contrast, POC works better under other forms of illumination changing environment. Lastly, the correlation values between original and images that change in position and orientation are estimated by zooming, rotating and shifting the original inputs shown in Fig. 4.5. Table 4.4 is the correlation result which SSD generates the highest correlation values under a slightly change in position input as zooming and rotating. However, while POC has a limitation when correlating with zooming and rotating inputs, it produces the highest correlation for image shifting which its correlation peak is also shifted at the same amount as the input image.


Figure 4.2 Test images that are different in sharpness by applying (a) 5x5, (b) 11x11 average filter, (c) 5x5, (d) 11x11 median filter, (e) 5x5, and (f) 11x11 Gaussian low-pass filter.

Table 4.1 Correlation results of different sharpness inputs. ($\sigma$ is the standard deviation of the Gaussian filter)

| Input image pairs (original and filtered images) | SAD | SSD | ZNCC | POC |
|---|---|---|---|---|
| 5x5 Average filter | 0.9515 | 0.9958 | 0.9377 | 0.0917 |
| 11x11 Average filter | 0.9141 | 0.9904 | 0.8535 | 0.0634 |
| 5x5 Median filter | 0.9700 | 0.9980 | 0.9708 | 0.2373 |
| 11x11 Median filter | 0.9481 | 0.9950 | 0.9241 | 0.1542 |
| 5x5 Gaussian filter, $\sigma$ = 1 | 0.9682 | 0.9980 | 0.9703 | 0.5283 |
| 11x11 Gaussian filter, $\sigma$ = 5 | 0.9203 | 0.9915 | 0.8713 | 0.0789 |

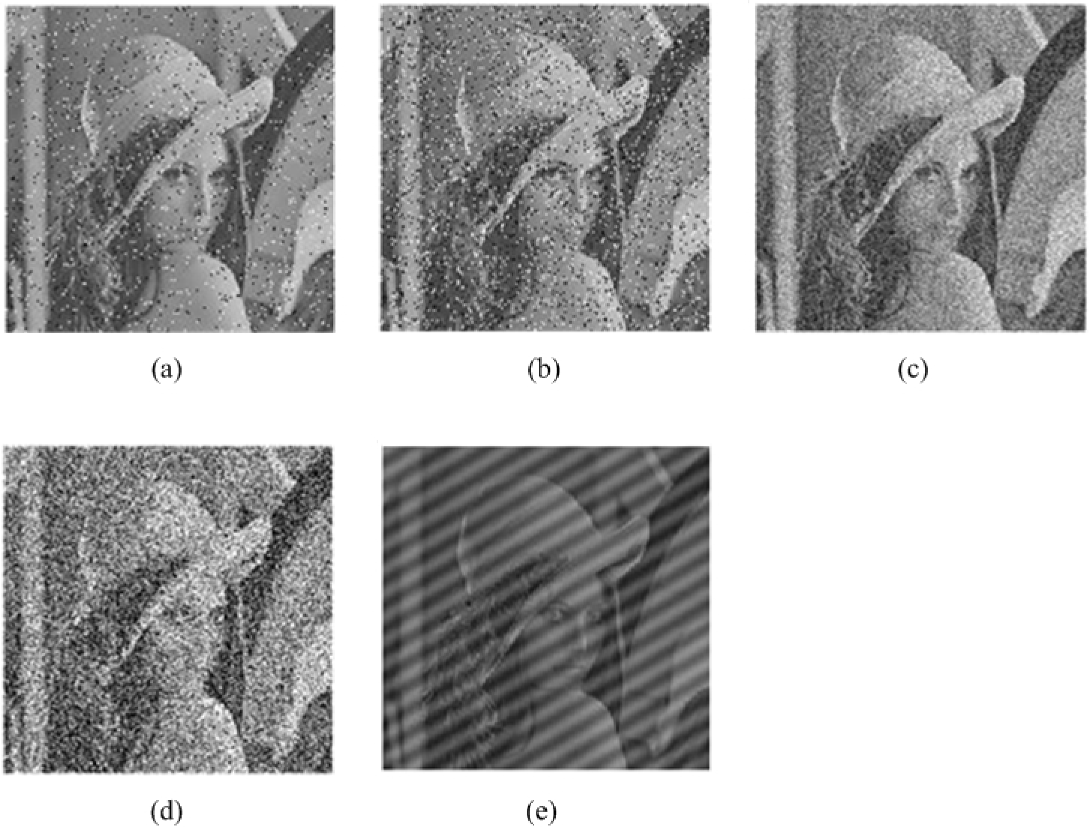



Figure 4.3 Comparison inputs which are noisy by adding salt and pepper noise with (a) density of 0.05, (b) density of 0.1, Gaussian noise of zero mean and variance of (c) 0.01, (d) 0.05, and (e) periodic noise.

Table 4.2 Correlation results of different noisy inputs. ($d$ is the noise density, $\mu$ is the mean and $\sigma^2$ is the variant of Gaussian noise)

| Input image pairs (original and noisy images) | SAD | SSD | ZNCC | POC |
|---|---|---|---|---|
| Salt and pepper, $d$ = 0.05 | 0.9506 | 0.9739 | 0.7408 | 0.2229 |
| Salt and pepper, $d$ = 0.1 | 0.9003 | 0.9469 | 0.5910 | 0.1582 |
| Gaussian noise,$\mu$ = 0, $\sigma^2$ = 0.01 | 0.8407 | 0.9801 | 0.7964 | 0.2501 |
| Gaussian noise,$\mu$ = 0, $\sigma^2$ = 0.05 | 0.6528 | 0.9086 | 0.5085 | 0.1315 |
| Periodic noise | 0.6921 | 0.9366 | 0.6824 | 0.9737 |

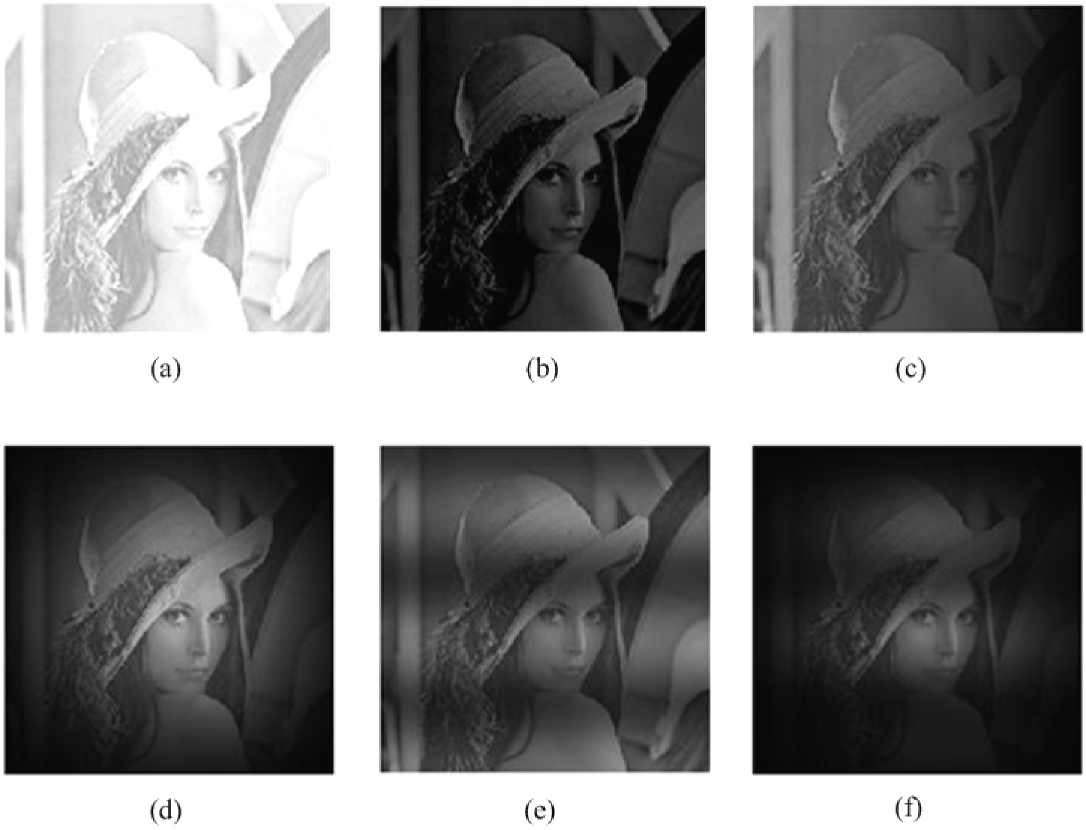



Figure 4.4 Comparison inputs which are different in illumination, (a) uniform (brighter), (b) uniform (darker) (c) linear, (d) Gaussian, (e) periodic, and (f) mixed Gaussian and periodic shading.

Table 4.3 Correlation results of different illumination inputs

| Input image pairs (original and illumination variation images) | SAD | SSD | ZNCC | POC |
|---|---|---|---|---|
| Uniform shading, brighter | 0.2426 | 0.7109 | 0.9724 | 0.8807 |
| Uniform shading, darker | 0.2531 | 0.7181 | 0.9695 | 0.7548 |
| Linear shading | 0.4878 | 0.8079 | 0.2161 | 0.8705 |
| Gaussian shading, $\sigma = 64$ | 0.3507 | 0.7392 | 0.3387 | 0.7816 |
| Periodic shading | 0.5052 | 0.8587 | 0.6775 | 0.9010 |
| Gaussian and periodic shading | 0.1804 | 0.6306 | 0.3247 | 0.7384 |


Figure 4.5 Comparison inputs which are different in position, (a) 105and (d) 5 degree counter clockwise rotating, and (e) horizontal shifting.

Table 4.4 Correlation results of geometric transformation inputs

| Input image pairs (original and geometric transformed images) | SAD | SSD | ZNCC | POC |
|---|---|---|---|---|
| Zooming (105%) | 0.8885 | 0.9821 | 0.7426 | 0.0412 |
| Zooming (120%) | 0.7872 | 0.9562 | 0.3865 | 0.0143 |
| Rotating (2 degrees) | 0.8984 | 0.9799 | 0.7435 | 0.0530 |
| Rotating (5 degrees) | 0.8244 | 0.9562 | 0.5144 | 0.0269 |
| Horizontal shift | 0.6812 | 0.9084 | 0.1297 | 0.9635 |

## 4.2 Implementation of Stereo Vision System

At the beginning of this work, the reconstruction results of an interested point in 3-D space are obtained by selecting the corresponding points manually following the markers on the body in Fig. 4.6 and Fig. 4.7. The input images are rectified to transform each image such that pairs of conjugate epipolar lines are horizontally aligned. After that four corresponding points are manually selected in both rectified images. As a result, the correspondence between points in an input image pair is determined. By using the camera calibration parameters for reconstruction, the location of selected points in 3-D space could be found as shown in Fig. 4.8 which the points in 3-D space of the first position are plotted beneath the second position and the depth information of both positions could be observed.

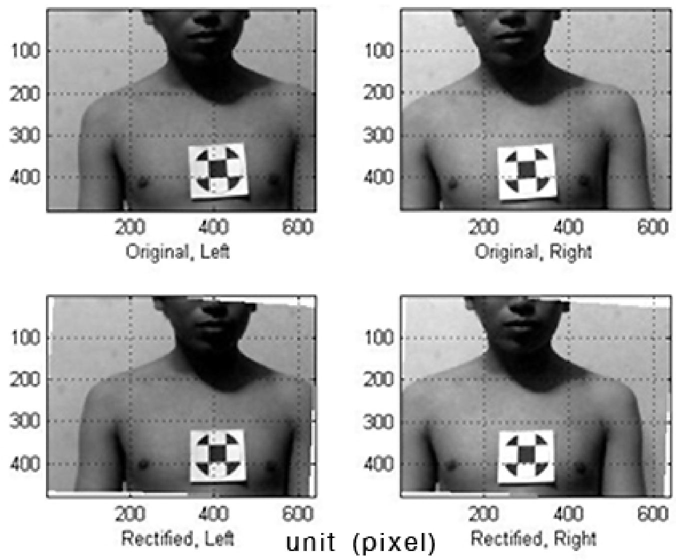



Figure 4.6 Stereo image pair of first position.

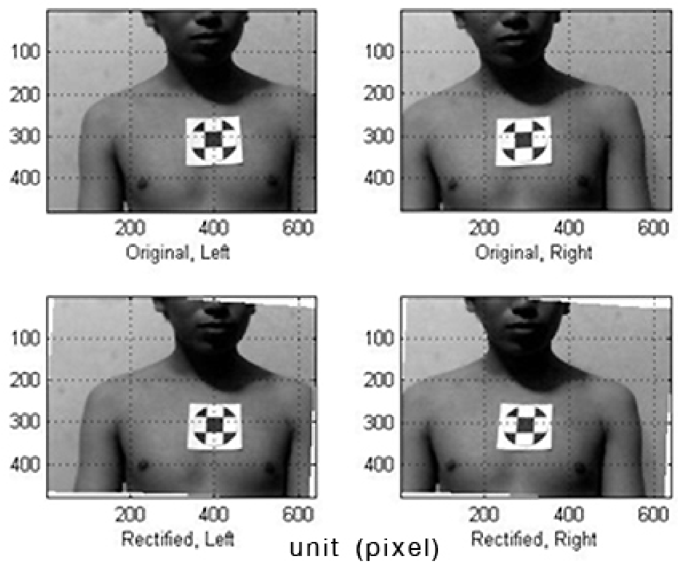



Figure 4.7 Stereo image pair of second position.

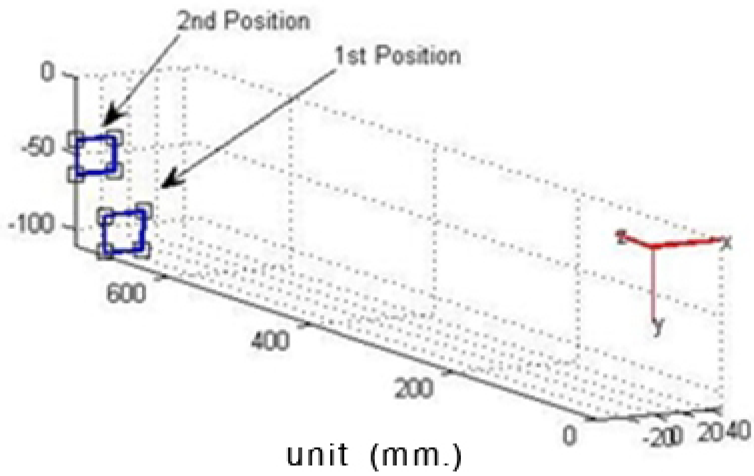



Figure 4.8 The 3-D plots of reconstructed points.

In automatic analysis experiment, the input images are captured from a stereo camera provided by Point Grey Research, Inc., i.e., Bumblebee2 as shown in Fig. 4.9 and the arrangement of the camera and the manikin is shown in Fig. 4.10. The Bumblebee2 camera uses two CCD image sensors to provide a balance between 3D data quality which its specifications are a baseline of 12 cm. and a 2.5 mm. focal length. Moreover, Bumblebee2 is pre-calibrated camera against distortion and misalignment and the calibration value can be retrieved from the camera by using the provided software for Bumblebee2 camera, i.e., The Triclops Stereo Vision Software Development Kit. In addition, the actual calibrated parameters used in the reconstruction process are the baseline of 11.88 cm. and the focal length of 2.5 mm or 141.07 pixels.

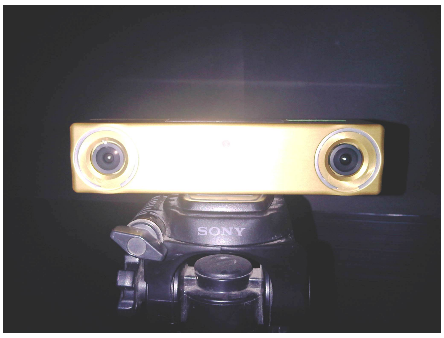



Figure 4.9 The Bubblebee2 stereo camera.

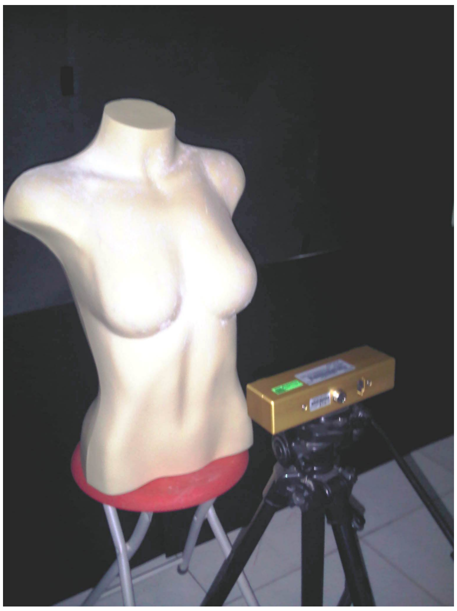


Figure 4.10 The arrangement of the camera and the manikin.

The manikin, which represents the human body, and the camera are seperated from each other which five different distances have been selected in this experiment which are 20, 25, 30, 35, and 40 cm.. At each distance, a pair of human body images captured is used as an input pair to the stereo vision process which follows the diagram in Fig. 3.1. However, the camera calibration can be separately processed in an off-line mode which the camera parameters can be employed in the image rectification and 3-D reconstruction processes afterward. The stereo camera used in this work is composed of two lenses which each lense can capture image simutaneously. In the pre-processing stage, the left and right raw images of human body shown in Fig. 4.11 are primarily captured and used as inputs of the image rectification process which are remapped to the images that fit the pin-hole camera model and their rectified outputs are shown in Fig. 4.12. Next to the rectification, the rectified outputs are sent to the edge detection process to extract their feature points. The low and high thresholds of Canny edge detection are properly adjusted until the boundary of the body of each input is obviously seen. The outputs from this process can be seen in Fig. 4.13. After that the left and right most boundaries of the edge images are selected as shown in Fig. 4.14. Then, in edge matching the disparity value between each pixel of the extracted boundary from left and right images is calculated along the horizontal line which an example of disparity calculation is shown in Fig. 4.15 and the disparity result of each pixel of the edge boundary is kept in the disparity map image of the edge boundary shown in Fig. 4.16.

In region matching, the zero mean normalized cross correlation (ZNCC) is selected as the matching algorithm for estimating the disparity value of each pixel within the boundary of the body since in the previous comparative study it performs a high correlation result in many cases, especially, the uniform shading which is very likely to occur between a pair of stereo input image. Correspondence searching in region matching is also done along horizontal direction by applying standard block matching to the input image, as a result, the disparity map of the body region is shown in Fig. 4.17. Finally, Fig. 4.18 and Fig. 4.19 show the reconstruction result of the body region which the unit of each axis is in millimeter. The distance from the reconstruction results are close to the real distance when the distance is at least approximately 30 cm. as shown in Fig. 4.19 (c). In addition, there are some mismatching results at 25 cm. shown in Fig. 4.19 (b). and higher at the distance of 20 cm. shown in Fig. 4.19 (a) because in the shorter distance the disparity between two images are higher which leads to the difficulties in obtaining the exact value of the correspondence point. On the other hand, the results are still acceptable at the distances of 35 and 40 cm. shown in Fig. 4.19 (d) and Fig. 4.19 (e), respectively. For the application of the tele-cardiac auscultation, the frame of the camera should particularly capture the chest of the patient, therefore, the camera is not necessary place at a long distance apart from the body and the experiment shows that the distance that is optimized is minimum at approximately 30 cm..

Although, the projection of the 3-D contour from 2-D space may confront with an accuracy problem due to, for example, an occlusion problem, the purpose of using the contour of the body in edge matching is to, initially, estimate the approximate distance from the camera to the contour points which will be, afterwards, used as the limit in region matching since the characteristic of the human body, the body region should be closer to the camera than its contour. Also, an invarient problem of the projection of 3-D contour on 2-D space can be neglected in an experiment that uses a small baseline of the stereo camera as in this experiment. Moreover, from the experiment, there are some mismatch pixels that generate the poor reconstruction results after using only region matching. However, by combining the edge matching with the region matching the results are more improved since the mismatch results from the pure region matching are bounded by the mapping value of the contour points from the edge matching process.

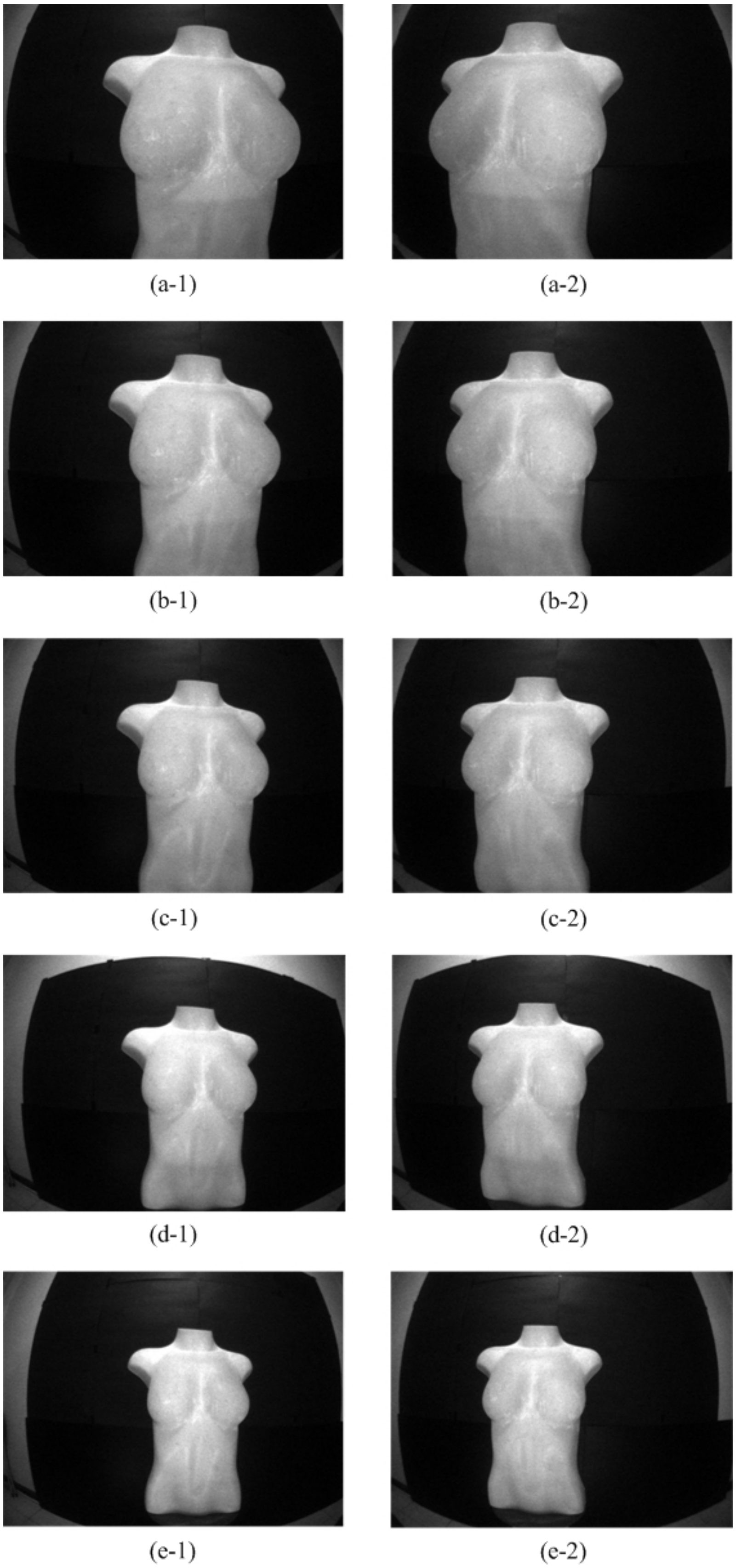



Figure 4.11 The raw images taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

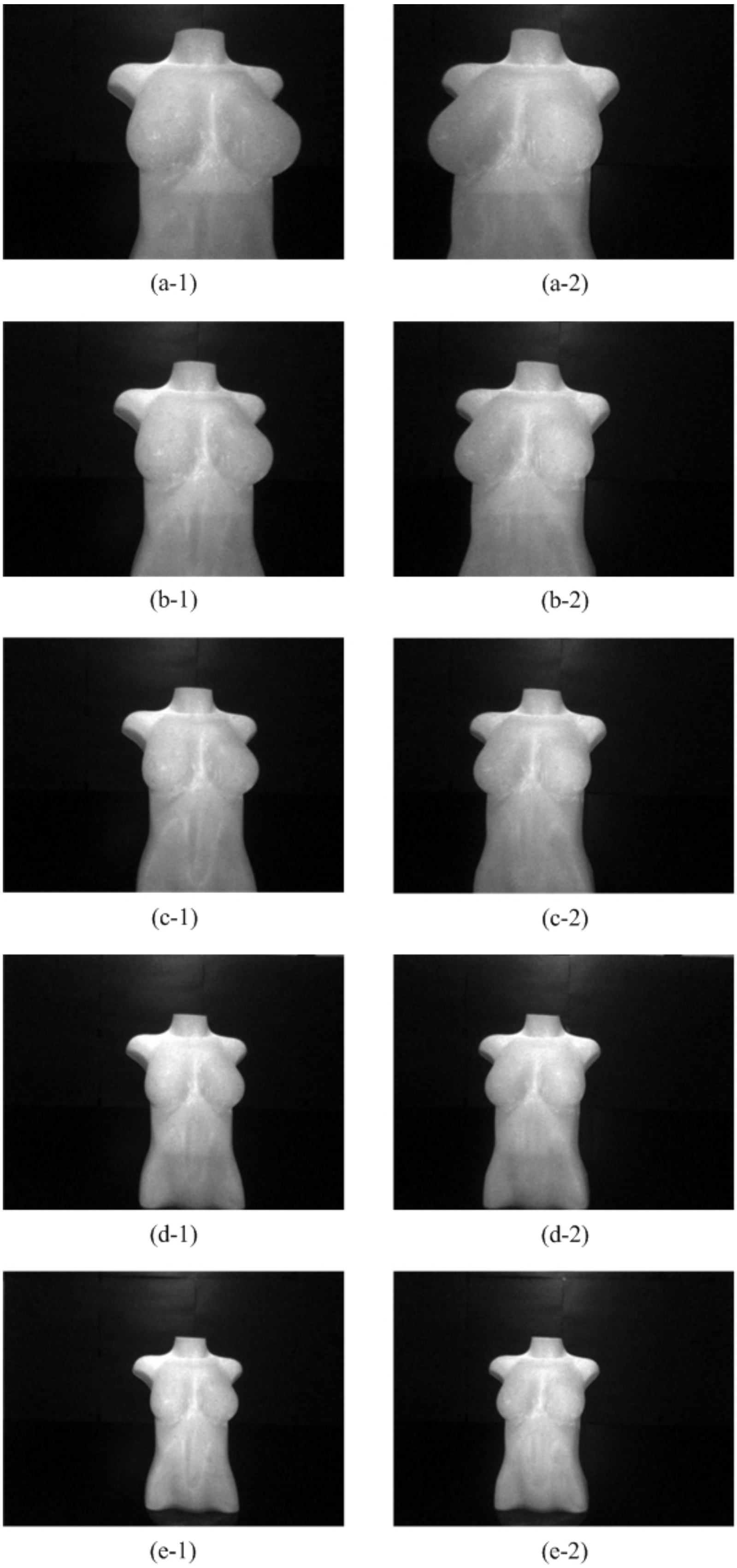



Figure 4.12 The rectified images taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

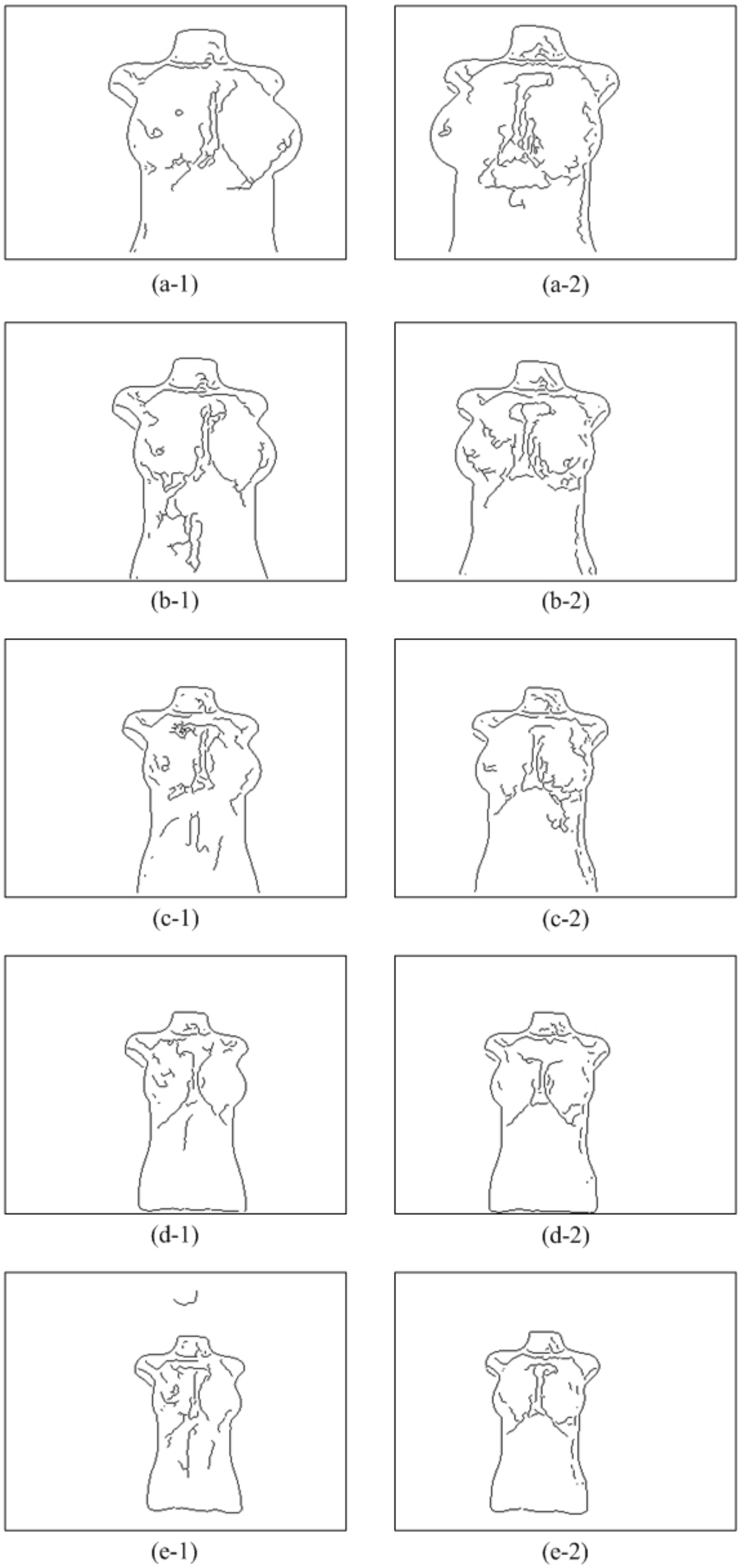
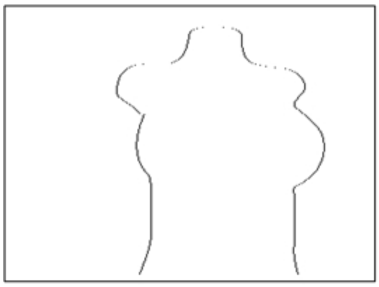
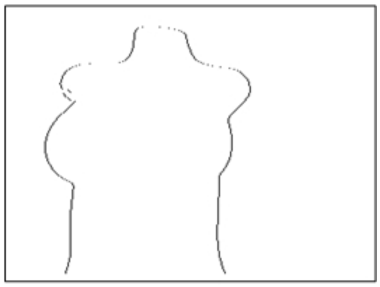
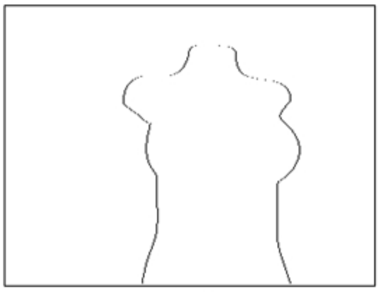
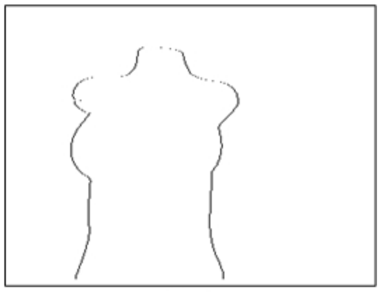


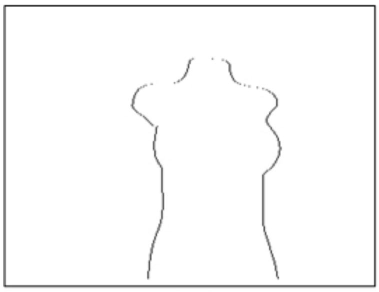

Figure 4.13 The edge images taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

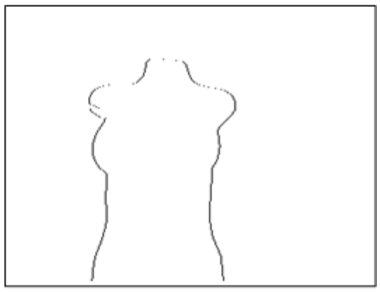

(a-1)

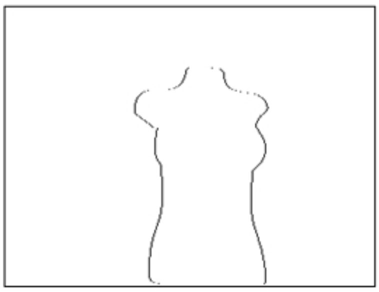

(a-2)

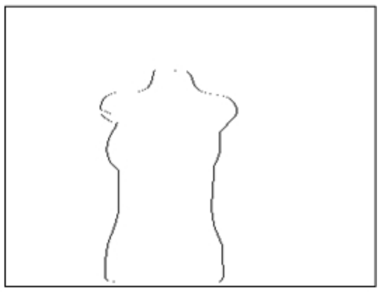

(b-1)

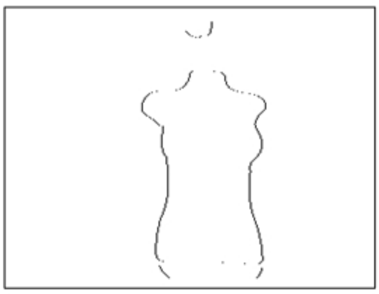

(b-2)


(c-1)


(c-2)


(d-1)


(d-2)


(e-1)

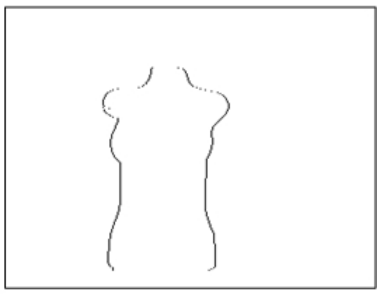

(e-2)

Figure 4.14 The boundary images taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

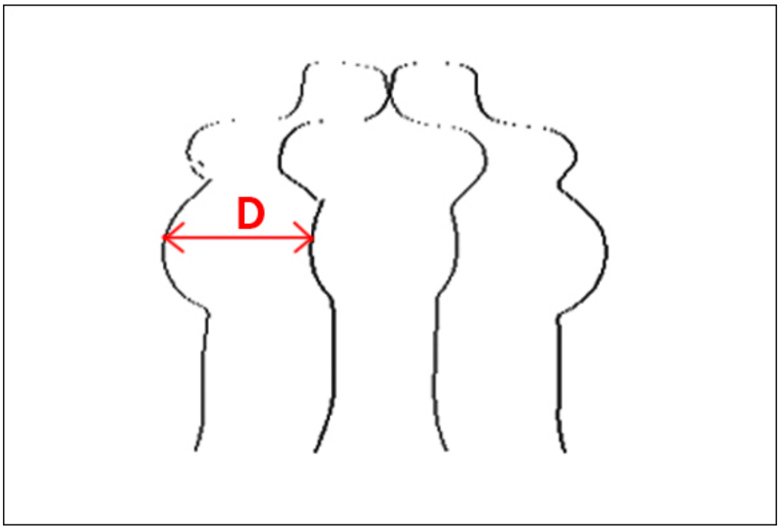



Figure 4.15 An example of disparity calculation.

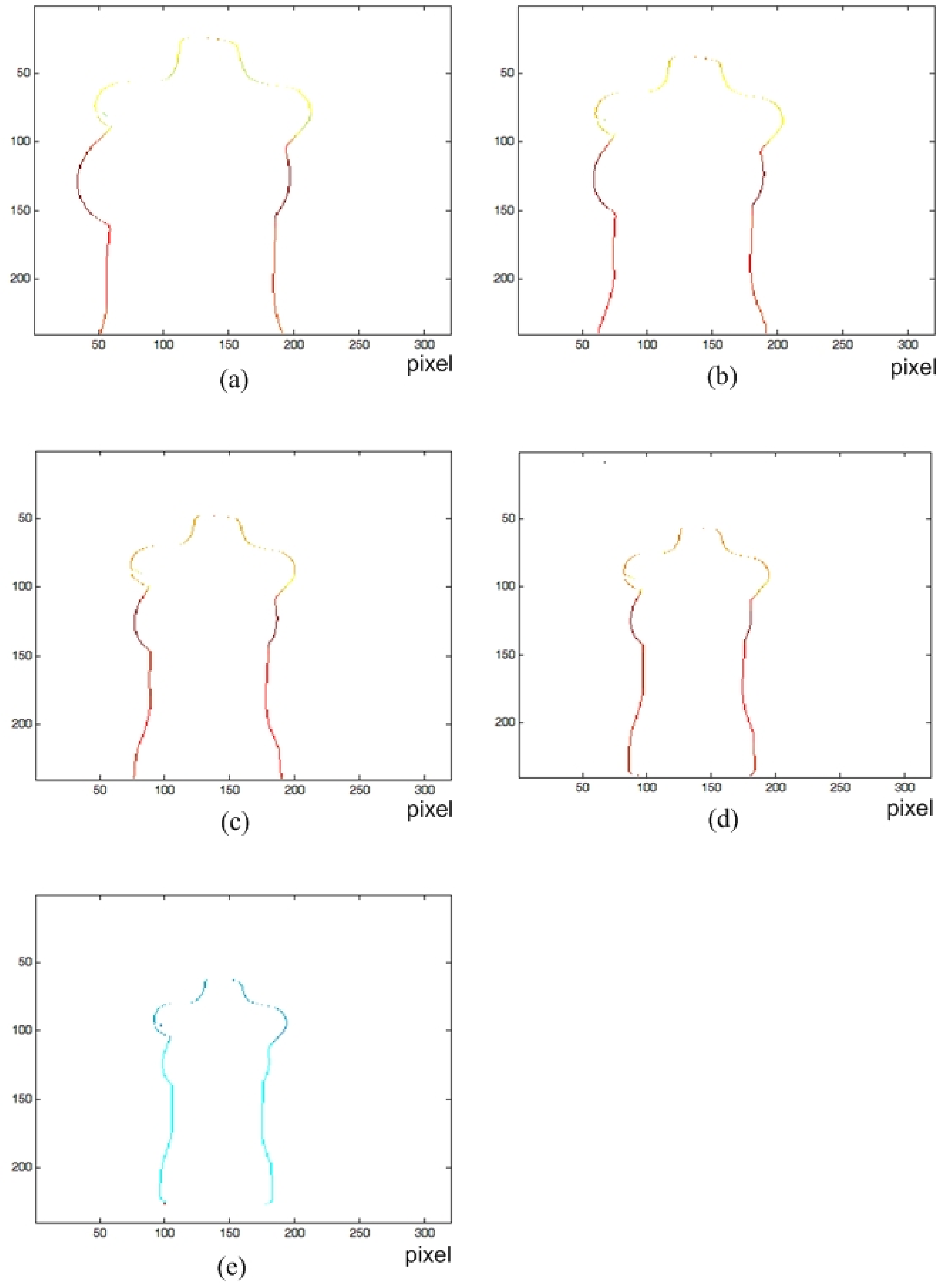



Figure 4.16 The disparity map images of the edge boundary taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

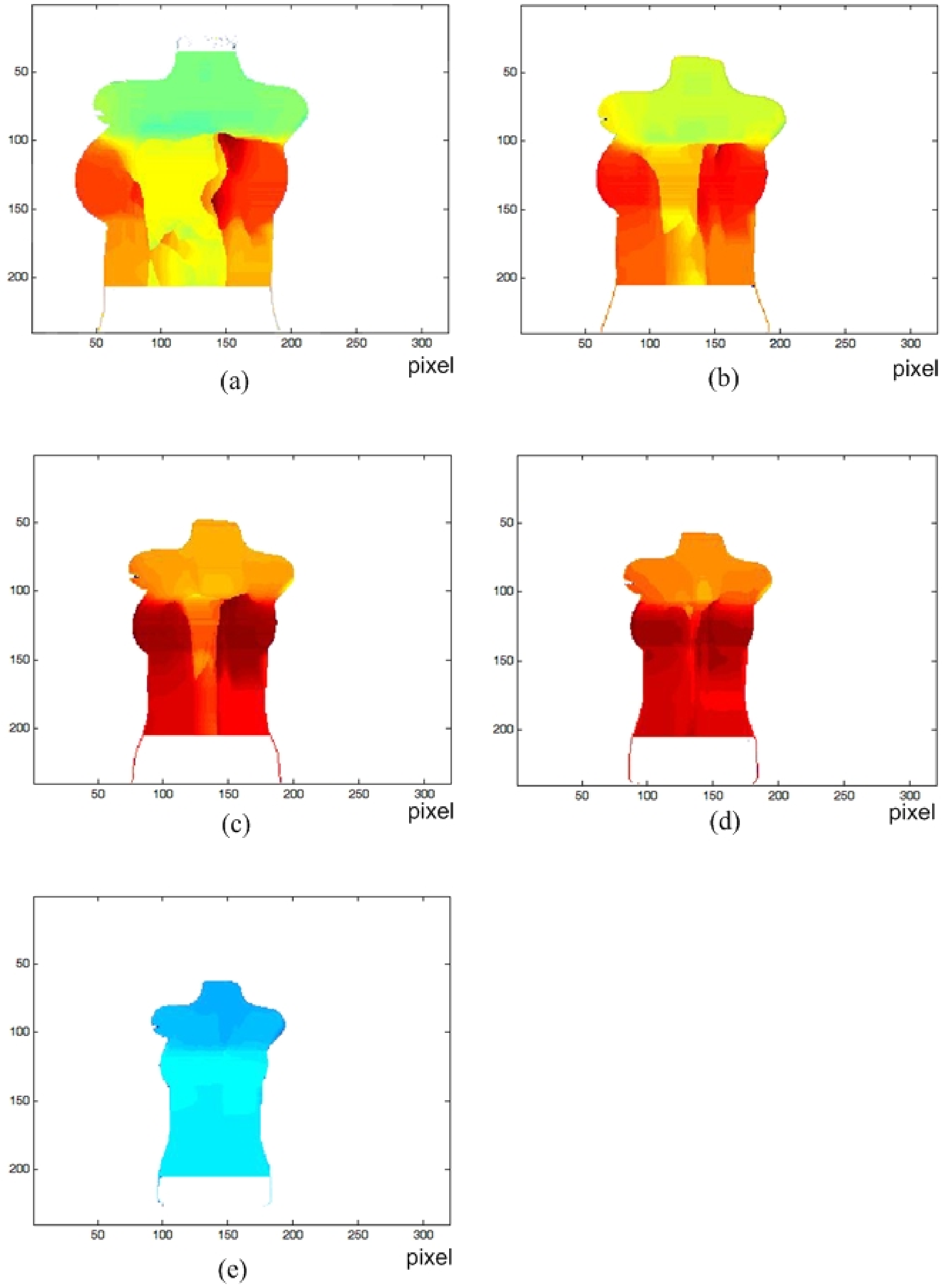



Figure 4.17 The disparity map images of the body region taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

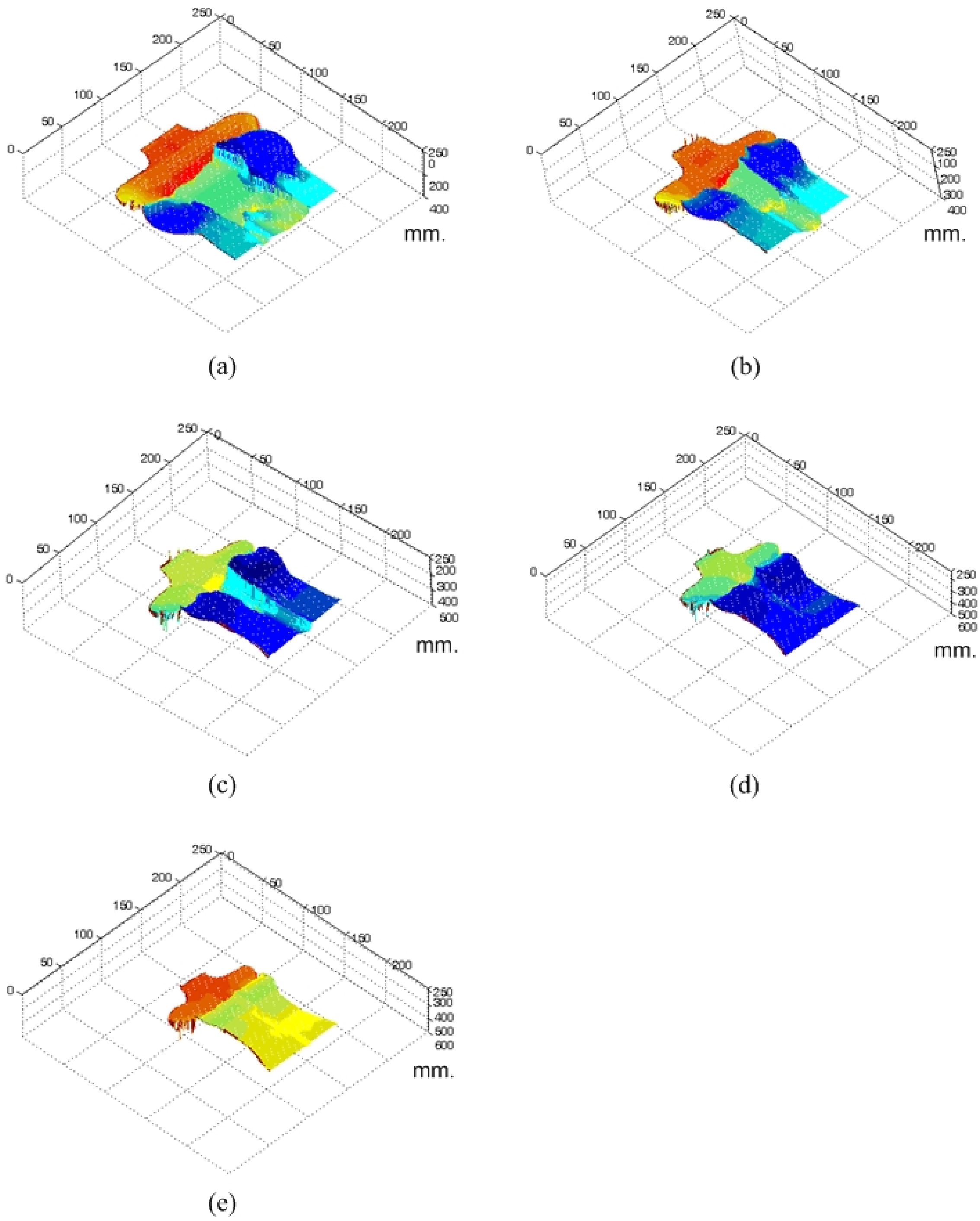



Figure 4.18 The reconstruction results (1st view) of inputs taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

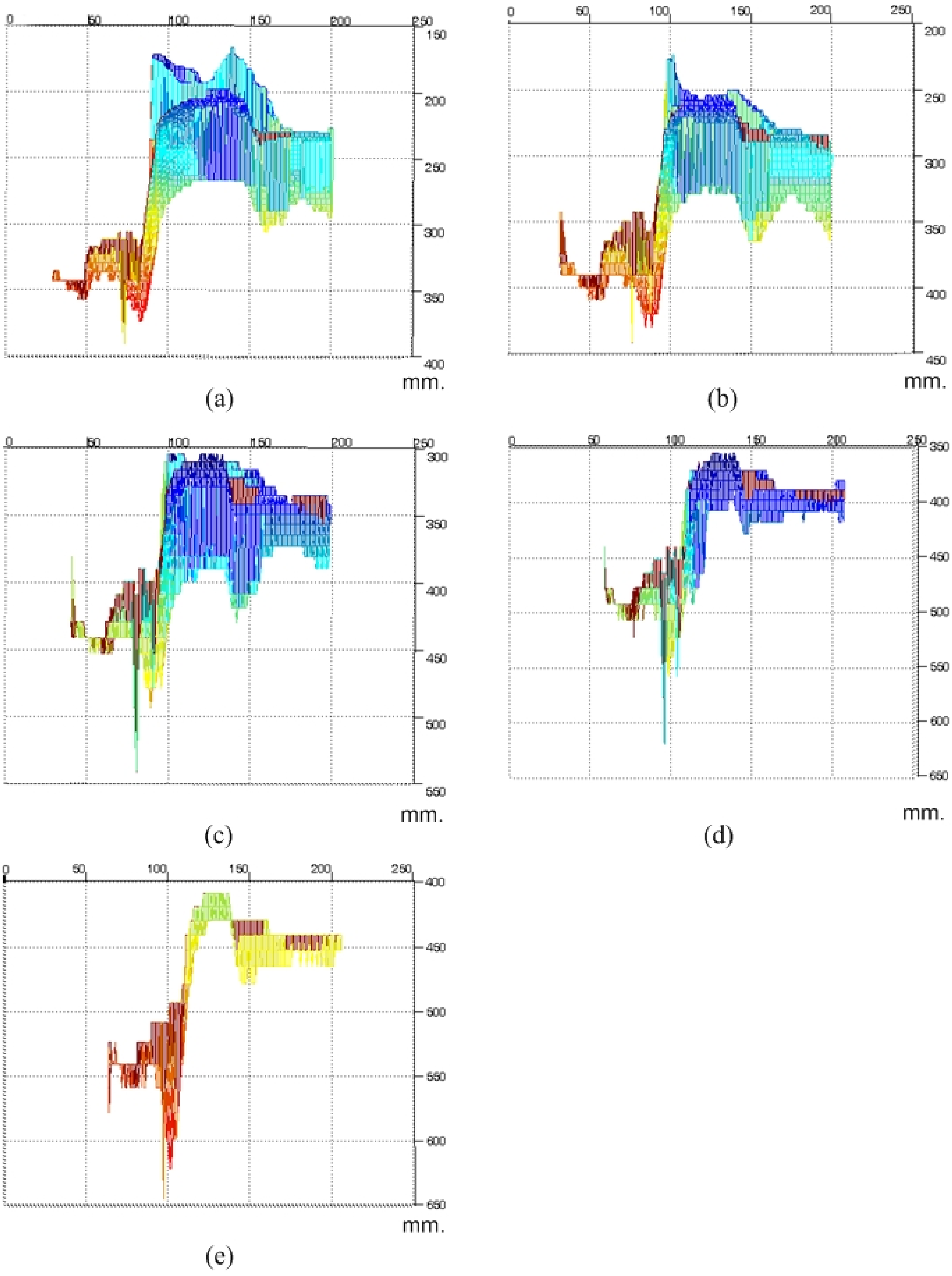



Figure 4.19 The reconstruction results (2nd view) of inputs taken at different distances, (a-1) left image at 20 cm., (a-2) right image at 20 cm., (b-1) left image at 25 cm., (b-2) right image at 25 cm., (c-1) left image at 30 cm., (c-2) right image at 30 cm., (d-1) left image at 35 cm., (d-2) right image at 35 cm., (e-1) left image at 40 cm., and (e-2) right image at 40 cm..

# Chapter 5

# Conclusions

The work presents a stereo vision system which is applicable to an automatic heart sound analysis system. The stereo vision system simulates human binocular vision by analyzing multiple images of the same scene captured at different positions which provide slightly dissimilar viewpoints. By detecting the corresponding points from two or more images the three-dimensional structural information on the scene can be recovered. In addition, the knowledge of a camera calibration, an image rectification, an epipolar geometry, is also utilized in this work.

In the corresponding analysis, The corresponding points to be reconstructed are automatically matched by using proper matching algorithm. From the comparative study of stereo matching algorithms, four stereo matching algorithms have been considered, namely, SAD, SSD, ZNCC and POC which the calculation of the correlation values using SAD, SSD, and ZNCC which are based on the images spatial domain and POC which is based on the imagesfrequency domain. From the comparison results, SSD always less sensitive to the interference than SAD. While ZNCC works well under different sharpness, noisy environment, and also the uniform shading illumination while POC is less sensitive to the other kinds of illumination changing input. Experimental results show that each algorithm works properly under different conditions.

In addition, the process of the correspondence analysis for the tele-cardiac auscultation is divided into two categories, i.e. edge matching and region matching. In edge matching, the corresponding points of the boundary of human body are matched after the edge images have been extracted from the original input images. However, in region matching, ZNCC have been selected as the matching algorithm for analysing the corresponding points since it performs a high correlation result in many cases, especially, the uniform shading which is very likely to occur between a pair of stereo input image. Consequently, the reconstruction process can convert disparity map image of the human body optained from the correspondence analysis to three-dimensional structure which the conversion is done from an image coordinate to world coordinate.

However, there will be some further researches related to this work. For instance, employing a coarse-to-fine technique into the present stereo vision system. The coarse localisation will approximately determine the position of the human body, then the fine localisation will give the system more specific position. Furthermore, this work can be extended to approach the automatic system that is the doctor is able to command the system and the stereo camera will scan the patients body to retrieve the information of the body's size and determine the stethoscope position. Then the corresponding points to be reconstructed will be automatically matched. The robor arm will move the stethoscope to the position that will be auscultated, moreover, an automatic tuning, e.g. PID controller, can be employed to control the positions error that is possible to be occurred. Finally, the heart sound will be recorded and the signal will be generated and sent back to the other site doctor.

# References


[1] W.W. Myint and B. Dillard, "An electronic stethoscope with diagnosis capability", *Proceedings of the 33rd Southeastern Symposium on System Theory*, 2001.

[2] O. Abdel-Alim, N. Hamdy, and M.A. El-Hanjouri, "Heart diseases diagnosis using heart sounds", *Proceedings of the 19th National Radio Science Conference*, 2002.

[3] D. Beymer and K. Konolige, "Real-time tracking of multiple people using stereo", *Proceedings of the IEEE FrameRate Workshop*, 1999.

[4] S. Bahadori, L. Iocchi, and L. Scozzafava, "People and robot localization and tracking through a fixed stereo vision system", *First RoboCare Workshop*, 2003.

[5] E. Trucco, and A. Verri, *Introductory Techniques for 3-D Computer Vision*, 1998.

[6] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2008.

[7] J.A. Kuhl, *Comparison of stereo matching algorithms for mobile robots*, http://robotics.ee.uwa.edu.au/theses/2005-Stereo-Kuhl.pdf.

[8] W. Zhang, Q. Zhang, L. Qu, S. Wei, "A stereo matching algorithm based on multiresolution and epipolar constraint", *Proceedings of the 3rd International Conference on Image and Graphics*, 2004.

[9] O. Faugeras, "Stereo Vision", *Three-Dimensional Computer Vision, A Geometric Viewpoint*, 1993.

[10] Z. Zhang, "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000.

[11] *Camera calibration toolbox for matlab*, http://www.vision.caltech.edu, July 2008.

[12] A. Saxena, J. Schulte, A. Y. Ng,"Depth estimation using monocular and stereo cues", *Proceedings of the International Joint Conferences on Artificial Intelligence*, 2007.

[13] M. S. Nixon, A. S. Aguado, “Low-level feature extracion (including edge detection)”, *Feature Extraction and Image Processing*, 2008.

[14] John Canny, “A computational approach to edge detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679698, Nov. 1986.

[15] Tangfei Tao, Ja Choon Koo, and Hyouk Ryeol Choi, “A fast block matching algorthim for stereo correspondence”, *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, pp. 38-41, 2008.

[16] A. Barjatya, *Block matching algorithms for motion estimation*, Final Project Paper, DIP 6620 Spring 2004.

[17] Andreas Koschan1, Volker Rodehorst, Kathrin Spiller, “Color stereo vision using hierarchical block matching and active color illumination”, *Proceeding of the 13th International Conference on Pattern Recognition, ICPR’96*, Vol. I, pp. 835-839, 1996.

[18] Chunni Dai, Xiang Wu, and Jingao Liu, “Stereo matching using adaptive genetic algorithm”, *Proceedings of the International Conference on Audio, Language and Image Processing*, 2008.

[19] T. Chinapirom, U. Witkowski, and U. Ruckert, *Stereoscopic camera for autonomous mini-robots applied in kheperasot league*, http://robogames.net/papers/2007/07-114-Chinapirom-UnivPaderborn-SteroscopicCameraForAutonomousMiniRobots

[20] L. D. Stefano, S. Mattoccia, and F. Tombari, “An algorithm for efficient and exhaustive template matching”, *Proceedings of the International Conference on Image Analysis and Recognition*, 2004.

[21] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, “3D face recognition using passive stereo vision”, *Proceedings of the IEEE International Conference on Image Processing*, 2005.

[22] K. Miyazawa, and T. Aoki, “A robot-based 3D body scanning system using passive stereo vision”, *Proceedings of the IEEE International Conference on Image Processing*, 2008.

[23] S. Kumpituck, T. Kondo, I. Nilkhamhang, W. Kongprawechnon, K. Tungpimolrut, and A. Nishihara, “A robustness study of stereo matching techniques”, *Proceedings of the International Conference on Information and Communication Technology for Embedded Systems*, 2010.

[24] S. Kumpituck, W. Kongprawechnon, T. Kondo, I. Nilkhamhang, K. Tungpimolrut, and A. Nishihara, "Stereo based vision system for cardiac auscultation tele-diagnostic system", *Proceedings of the International Conference on Control, Automation and Systems*, 2009.

# Publications

[1] S. Kumpituck, W. Kongprawechnon, T. Kondo, I. Nilkhamhang, K. Tungpimolrut, and A. Nishihara, "Stereo Based Vision System for Cardiac Auscultation Tele-Diagnostic System", *Proceedings of the ICROS-SICE International Joint Conference 2009 (ICROS-SICE2009)*, Fukuoka, Japan, 18-21 August 2009, 4 p.

[2] S. Kumpituck, T. Kondo, I. Nilkhamhang, W. Kongprawechnon, K. Tungpimolrut, and A. Nishihara, "A Robustness Study of Stereo Matching Techniques", *Proceedings of the International Conference on Information and Communication Technology for Embedded Systems 2010 (ICICTES2010)*, Pathum Thani, Thailand, 28-30 January 2010, 6 p.