PERFORMANCE IMPROVEMENT OF A 3D RECONSTRUCTION
ALGORITHM USING SINGLE CAMERA IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

VARLIK KILIÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MECHANICAL ENGINEERING

JULY 2005

Approval of the Graduate School of Natural and Applied Sciences

_____

Prof. Dr. Canan ÖZGEN

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. S. Kemal İDER

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Prof. Dr. Bülent E. PLATİN

Supervisor

Examining Committee Members

Prof. Dr. S. Turgut Tümer (Chairman)	(METU, ME)	_____
Prof. Dr. Bülent E. PLATİN	(METU, ME)	_____
Prof. Dr. R. Tuna Balkan	(METU, ME)	_____
Prof. Dr. Abdulkadir Erden	(ATILIM, MECE)	_____
Asst. Prof. Dr. E. İlhan Konukseven	(METU, ME)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Varlık KILIÇ

ABSTRACT

PERFORMANCE IMPROVEMENT OF A 3D RECONSTRUCTION ALGORTIHM USING SINGLE CAMERA IMAGES

Kılıç, Varlık

M.Sc., Department of Mechanical Engineering

Supervisor: Prof. Dr. Bülent E. Platin

July 2005, 146 pages

In this study, it is aimed to improve a set of image processing techniques used in a previously developed method for reconstructing 3D parameters of a secondary passive target using single camera images. This 3D reconstruction method was developed and implemented on a setup consisting of a digital camera, a computer, and a positioning unit. Some automatic target recognition techniques were also included in the method. The passive secondary target used is a circle with two internal spots.

In order to achieve a real time target detection, the existing binarization, edge detection, and ellipse detection algorithms are debugged, modified, or replaced to increase the speed, to eliminate the run time errors, and to become compatible for target tracking. The overall speed of 20 Hz is achieved for 640x480 pixel resolution 8 bit grayscale images on a 2.8 GHz computer

A novel target tracking method with various tracking strategies is introduced to reduce the search area for target detection and to achieve a detection and reconstruction speed at the maximum frame rate of the hardware.

Based on the previously suggested lens distortion model, distortion measurement, distortion parameters determination, and distortion correction methods for both radial and tangential distortions are developed. By the implementation of this distortion correction method, the accuracy of the 3D reconstruction method is enhanced.

The overall 3D reconstruction method is implemented in an integrated software and hardware environment as a combination of the methods with the best performance among their alternatives. This autonomous and real time system is able to detect the secondary passive target and reconstruct its 3D configuration parameters at a rate of 25 Hz. Even for extreme conditions, in which it is difficult or impossible to detect the target, no runtime failures are observed.

ÖZ

ÜÇ BOYUTLU UZAYDA CİSİMLERİN KONUMLARININ TEK
KAMERA GÖRÜNTÜSÜ KULLANILARAK BELİRLENMESİ İÇİN
GELİŞTİRİLMİŞ YÖNTEMİN PERFORMANS İYİLEŞTİRİLMESİ

Kılıç, Varlık

Yüksek Lisans, Makina Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Bülent E. Platin

Temmuz 2005, 146 sayfa

Bu çalışmada, pasif ikincil bir hedefin tek bir sayısal görüntüsünü kullanarak cisimlerin 3 boyutlu uzaydaki konumlarını belirleyen bir yöntemde kullanılan görüntü işleme tekniklerinin iyileştirilmesi hedeflenmiştir. Daha önceki çalışmalarda, 3 boyutlu uzayda cisimlerin konumlarını belirleyen kuram geliştirilmiş, sayısal kamera, bilgisayar ve hassas konumlandırma cihazından oluşan deney düzeneği üzerinde uygulanmıştır. Sayısal görüntü içindeki ikincil hedefi otomatik olarak saptayan yöntemler uygulamaya dahil edilmiştir. Kullanılan pasif ikincil hedef, üzerinde iki benek olan bir dairedir.

Yöntemin gerçek zamanlı çalışabilmesi için, mevcut siyah-beyazlaştırma, kenar belirleme ve elips tanıma algoritmaları hatalardan arındırılmış, hızlandırılmış veya tamamen değiştirilmiş, hedef izlenmesi için uygun hale getirilmiştir. 640x480 piksel boyutunda 8 bitlik sayısal görüntü için hedef saptama yönteminin 2.8 GHz işlemcili bir bilgisayarda çalışma hızı ortalama 20 Hz olarak ölçülmüştür.

Farklı takip yaklaşımları içeren yeni bir hedef izleme yöntemi kullanılarak, hedef saptama yönteminin tarama alanı küçültülmüş ve çalışma hızı donanımın desteklediği en yüksek görüntü yakalama hızına eriştirilmiştir.

Daha önceki çalışmada önerilen mercek bozukluğu modeli kullanılarak, hem radyal hem de açısal bozuklukları ölçme, bozukluk modelinin parametrelerini belirleme ve bozukluk düzeltme için bir yöntem geliştirilmiştir. Bu yöntemin uygulanmasıyla 3 boyutlu konum belirleme yönteminin hassasiyeti iyileştirilmiştir.

Çeşitli seçenekler arasından en iyi performansa sahip yöntemler birleştirilerek oluşturulan 3 boyutlu konum belirleme yöntemi bütünleşik bir yazılım-donanım ortamında uygulanmıştır. Tümüyle otonom ve gerçek zamanlı çalışan bu sistem ikincil bir hedefi algılayabilmekte ve üç boyutlu konum bilgilerini 25 Hz'lik bir hızda hesaplayabilmektedir. Hedefin zor fark edildiği ya da fark edilemez olduğu en uç durumlarda bile çalışma zamanı hatası gözlenmemiştir.

Anahtar Kelimeler: otonom siyah-beyazlaştırma, elips tanıma, hedef izleme, mercek bozuklukları, 3 boyutlu konum belirleme.

To my family…

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Bülent E. Platin for his instructive guidance and insightful comments throughout the study. I would also like to thank him for his model academic attitude, which will guide me through the rest of my carrier.

I am indebted to Asst. Prof. Dr. Buğra Koku. His patient and valuable hardware support provided me with the chance to make this thesis complete.

I would like to express my appreciation to Mr. İ. Alpar Kılınç, Dr. Ercan U. Acar, and Ms. Sibel Özkılıç. Without their contributions, this study would not have been what it is now.

I owe many thanks to my friends Mr. Andaç T. Şamiloğlu, Mr. Soner Büyükatalay, and Ms. Elçin Özgür for their precious support and presence.

Lastly, I am deeply grateful to my family who supported and encouraged me through all my life.

TABLE OF CONTENTS

LIST OF FIGURES

xv

LIST OF TABLES

LIST OF SYMBOLS

$\mathbf{R_x}$	*Rotation about x-axis*
$\mathbf{R_y}$	*Rotation about y-axis*
$\mathbf{R_z}$	*Rotation about z-axis*
$\mathbf{T_x}$	*Translation along x-axis*
$\mathbf{T_y}$	*Translation along y-axis*
$\mathbf{T_z}$	*Translation along z-axis*
d_0	*Image plane distance*
f	*Focal length*
R	*Main target radius*
r_0	*Spot radius*
r_f	*Spot offset*
I	*Grayscale image array*
B	*Binary image array*
t^*	*Threshold value*
$h(g)$	*Histogram function*
μ_1, μ_2	*Mean of Gaussian distribution*
σ_1, σ_2	*Variance of Gaussian distribution*
P_1, P_2	*Proportion of Gaussian Distribution*
$J(t)$	*Criterion function for minimum error threshold selection*
$F(t)$	*First moment vector of histogram function*
$S(t)$	*First moment vector of histogram function*
$\mathbf{S_h}$	*Horizontal Sobel filter*
$\mathbf{S_v}$	*Vertical Sobel filter*
$\mathbf{G_x}(x, y)$	*Gradient vector horizontal component array*
$\mathbf{G_y}(x, y)$	*Gradient vector vertical component array*

$\mathbf{E}(x, y)$	Edge pixels array
$C(x, y)$	Conic equation
$L(x, y)$	Equation of line connecting two points on an ellipse
$I_1(x, y)$, $I_2(x, y)$	Tangent lines to an ellipse
$F(\mathbf{a}, \mathbf{x})$	Conic equation
$\mathbf{a}$	Parameter vector for conic equation
$\mathbf{C}$	Constraint matrix
E	Total error function for direct ellipse fitting
$\mathbf{S}$	Scatter matrix
E_{rel}	Relative error criterion for direct ellipse fitting
$(\lambda_i, \mathbf{u_i})$	Eigenvalue, eigenvector pair
$P_m(t)$	m^{th} degree polynomial of time
$\mathbf{c}$	Coefficient vector of a polynomial
$\mathbf{r}$	Correlation coefficient of fitting
e	Estimation error
$\tilde{x}$	Estimated value
$\delta u_i^{(r)}$	Horizontal component of radial distortion
$\delta v_i^{(r)}$	Vertical component of radial distortion
$u_i^{(ud)}$	Undistorted horizontal image coordinate
$v_i^{(ud)}$	Undistorted vertical image coordinate
k_1, k_2	Radial distortion coefficients
r_i	Radial distance of the undistorted image coordinates to the origin of the image plane
$\delta u_i^{(t)}$	Horizontal component of tangential distortion
$\delta v_i^{(t)}$	Vertical component of tangential distortion
p_1, p_2	Tangential distortion coefficients
δu_i	Cumulative horizontal distortion
δv_i	Cumulative vertical distortion

u_i	*Distorted horizontal image coordinate*
v_i	*Distorted vertical image coordinate*
l	*Length of pendulum*
T	*Period of pendulum*
g	*Gravitational acceleration*
u_3	*Surface normal vector*

Acronyms

VCR	*Video cassette recorder*
CMOS	*Complimentary metal oxide semiconductor*
CCD	*Charge coupled device*
CID	*Charge induced device*
GT	*Global thresholding*
LT	*Local thresholding*
CN0, CN1	*Connectivity Analysis Methods*

CHAPTER 1

INTRODUCTION

1.1 Machine Vision

Vision is obviously the most complex and most powerful sense of human and its importance for the development and operation of human mind is indisputable. Any intelligent machine unaware of this powerful sense can not even be imagined. The problem of developing a machine system that gathers information about the 3D scenes from 2D images is called as the machine vision.

Unfortunately, almost all machine vision systems developed up to now are application dependent and can not be used for any other application except the one it is designed for and this fact seems to preserve its validation for some years.

The problem of machine vision can be classified into two main parts: pattern recognition and scene reconstruction. Pattern recognition deals with identifying and classifying objects in an image, while the reconstruction aims to obtain 3D configuration of these objects. Whatever the application or problem is, the solution relies very much on the fundamentals of the image formation and obviously to the human vision sense.

The process of image formation is well known especially in the human vision system, where the photons or electromagnetic waves reflected from the objects are bended by the eye lens and focused on the retina tissue, and finally produce some electric signals which carry the color or intensity information. This process is exactly the same for the cameras and called as the perspective projection.

The main problem under perspective projection from 3D world to 2D image plane is that the depth information is lost and the images of the foreground and background objects appear in the same plane.

Binocular or so-called stereo vision is the solution of the evolution to this problem. Using disparity of the images of the same scene taken by two cameras (Figure 1.1), the depth information is generated.

Figure 1.1. Stereo Imaging (http://www.elec.qmul.ac.uk/mmv/rendering.html)

On the other hand, it is well known that humans can also generate the depth information and even more from a single image (Figure 1.2) using their life time gathered experience. It is not expected from machines to use human feelings in processing an image, but it is possible to teach or implement some basic rules to reconstruct 3D information such as:

- Closer objects partially or fully occludes farther ones.
- Images of the closer objects are larger than images of farther objects of similar size.
- The image of a closer object on the ground plane occurs in a lower position on the image plane.
- Color and intensity values and changes in these values within an image plane give information about the shape, count, and orientation of objects.

Naturally, all rules listed above depend on the basics of perspective projection and some preliminary knowledge or experience about the scene as most studies on machine vision systems do.

Figure 1.2. Porters, Ara Güler, 1959.

1.2 Previous Studies

Although the history of image formation fundamentals goes beyond centuries, the intensive studies on machine vision systems have started few decades ago when the dedicated hardware become available. With developments in the technology, the use of vision systems spreads in many application areas such as production lines, autonomous vehicles, medicine, weapon systems, criminology and security.

In Mechanical Engineering Department of Middle East Technical University, the research on machine vision systems has started in early 90's in the four different laboratories, namely Control Systems, Mechatronics Design, Biomechanics, and BİLTİR (CAD/CAM and Robotics Center).

Control Systems Laboratory: Kılınç [1] proposed a 3D reconstruction method which utilizes a single image of a secondary passive target to determine the 3D configuration parameters. Acar [2] constructed a physical experimental setup and tested the accuracy of the 3D reconstruction method developed by Kılınç [1]. Özkılıç [3, 4] proposed some image processing and target detection methods to avoid operator guidance in the implementation of the 3D reconstruction method developed by Kılınç [1].

Mechatronics Design Laboratory: Arslan [5] developed a method utilizing low and intermediate level image processing techniques to convert images of machine elements to engineering drawings. Çokel [6] and Özdemir [7] proposed a method to estimate the position of an indoor robot using a single image of a door with known dimensions. Çetin [8] modified this method to improve the accuracy of position estimation.

Biomechanics Laboratory: Shafiq [9] proposed and developed a stereo vision system for gait analysis. Güler [10] extended this method to use 6 cameras and conducted experiments for lower extremity modeling. Karpat [11] introduced some techniques on the internal camera calibration and linearization, and measured the accuracy of the setup developed by Shafiq [9] and Güler [10]. This setup has been successfully used and is still being used by researchers to investigate lower extremity disorders and to crosscheck mathematical joint and tissue models.

BİLTİR: Konukseven [12] proposed a vision based method for recognizing and tracking an object on a moving conveyor belt. Şeran [13] developed a stereo vision system to assemble basic geometric primitives into their respective templates. Bayraktar [14] proposed a stereo vision system which locates and classifies industrial parts according to their geometric properties. Fidan [15] proposed a method to construct 3D image map of an indoor environment utilizing a laser range finder.

1.3 Objective of the Study

This study is the latest part of the project conducted in the Control Systems Laboratory of Mechanical Engineering Department at METU. The aim of the project was to develop a vision based non-iterative, accurate and robust sensing system.

In the previous studies of the project, a reconstruction algorithm was developed and its theoretical limits were tested [1]. This reconstruction algorithm was implemented by using a digital imaging system and the validity of the theoretical limits of the algorithm was examined through a set of experiments [2]. Digital image processing techniques such as automatic thresholding and ellipse detection were implemented in the studies to avoid the human guidance [3, 4].

This study aims to investigate the reconstruction algorithm developed in the previous study [1], the digital imaging system constructed in the previous study [2], and the automatic target detection method implemented in the previous studies [3, 4], in terms of their accuracy, reliability, and speed; and in addition to search for further improvements.

The main goal of this thesis is to build a robust vision based sensing system that works in real time[1], by using the results of the previous studies [1-4]. In order to achieve robustness and real time working speed, all methods and algorithms discussed and implemented in the previous studies [1-4] are re-examined.

1.4 Methodology

It is necessary to identify the design limitations and characteristics of the methods developed in previous studies [1-4] to clarify the aim, constraints and scope of this study.

---

[1] working at a speed more than or equal to the maximum frame rate of the imaging hardware

The idea of using a secondary circular passive target and projection based reconstruction goes beyond the previously cited studies [1-4] and first posed by Olgaç et al. [16]. That study [16] used a circular secondary target and orthographic projection model to reconstruct a reduced set of 3D configuration parameters. The method used by Platin [17] and Olgaç [18] is similar except that the orthographic projection model is replaced by perspective projection model. However these studies made the assumption that the target center lies on the optical axis of the camera. Besides this constraint, the use of a circular target makes the determination of the in-plane rotation, impossible.

The 3D reconstruction method proposed by Kılınç [1] utilizes a single image of a secondary passive planar target (Figure 1.3) to reconstruct the complete set of 3D configuration parameters of the target. The target is a circle with two internal spots, one is located at the center of circle and all target dimensions are known. The proposed method reconstructs the three rotations and three translations of the target analytically by the inverse perspective projection, using the equation of the ellipse and 2D coordinates of the spots on the image plane.
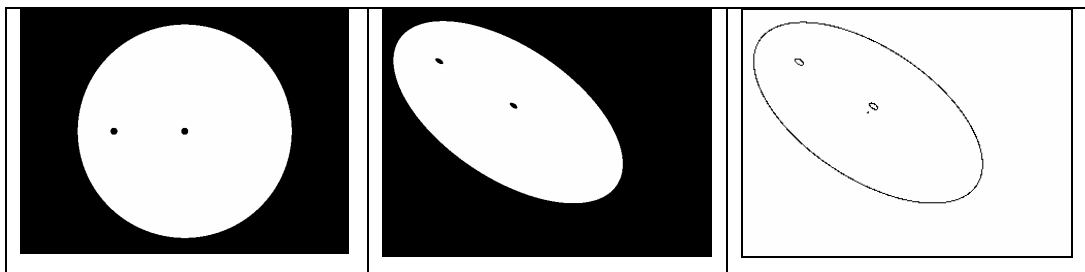
Figure 1.3. Secondary Passive Target as Utilized by Kılınç [1].

The image processing sequence followed in the previous study [2] (Figure 1.4), was designed considering required inputs of 3D reconstruction method. The steps except 3D reconstruction require operator guidance. In the image grabbing step, the operator adjusts the lighting conditions and captures a single grayscale image using the utility tools of the frame grabber device. Then a global threshold value is to be selected by the operator in the segmentation step, and the binary image is formed

containing foreground objects only. Finally, the smallest rectangular region in the binary image is marked by the operator in the target detection part, and the pixels corresponding to the contour and the spots of target are detected automatically and send to the 3D reconstruction method.
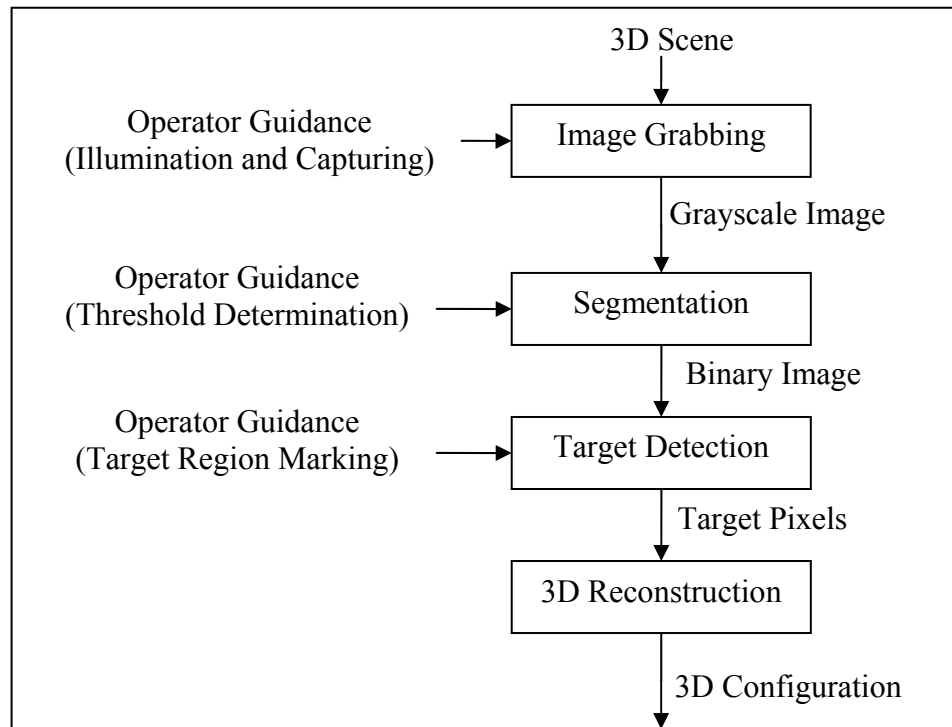
Figure 1.4. Image Processing Sequence Formed by Acar [2].

In the pervious studies [3, 4], autonomous thresholding, ellipse detection and some complementary methods were added into segmentation and target detection steps to avoid the operator guidance. The image processing sequence was kept unchanged (Figure 1.5). An internal camera calibration method was also proposed to increase the reconstruction accuracy. However, due to the high computational load and implementation bugs of the proposed ellipse detection method, direct frame grabbing could not be implemented in the study [4].

Since this study aims to improve the performances of the methods proposed in the previous studies [1-4] as briefed above, the design objectives of the previous studies are also valid for this study.
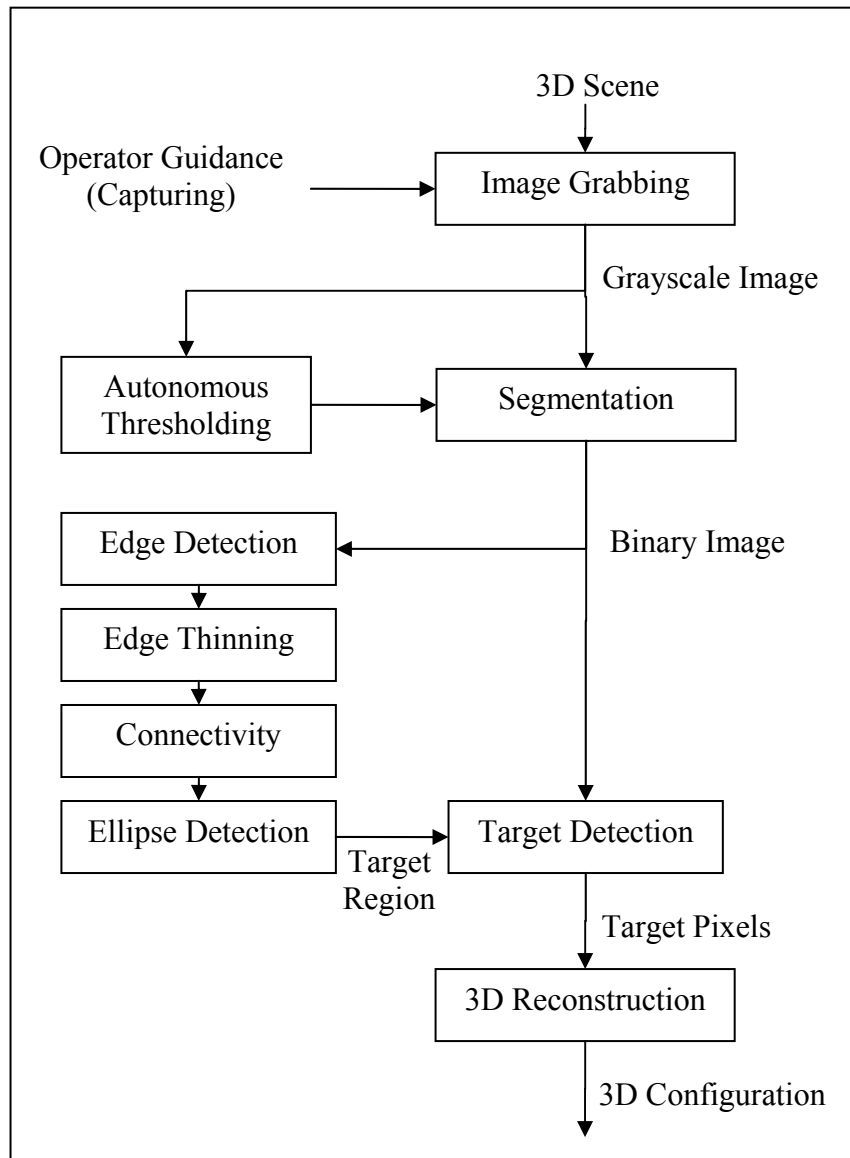
Figure 1.5. Image Processing Sequence Followed by Özkılıç [4].

The 3D reconstruction method used in previous studies [1-4] was monocular. In this study, the sequential images obtained from a single camera are utilized to reconstruct 3D configuration of a moving target. However, no complementary method or modification should be proposed in this study to use disparity of these sequential images. Therefore, the monocular feature of the project is to be maintained.

8

Low computational load and memory usage are the main advantages of the monocular vision to stereo vision. Thus, any method proposed in this study must prefer analytical solutions to iterative ones to keep this advantage. High memory and CPU usage should be avoided to ease the implementation of the method on digital image processing integrated circuits.

The method is aimed to be fully autonomous, so there should be no operator guidance during the operation of the methods or in their possible modifications in this study. However, to increase the reconstruction performance and to adapt different application environments, all parameters of the hardware and the software are aimed to be adjustable.

Robustness is another important goal of this study, so the singularities or ambiguities of the methods proposed in this study should be clarified and all methods should check its inputs and outputs for possible singularities and ambiguities and return a success or error code.

1.5 Outline of the Study

In this study, the developed vision based sensing system is divided into implementation modules and sub-modules and the following chapters are organized to investigate each module separately.

In Chapter 2, the target detection is discussed as the first module. Its sub-modules imaging, autonomous segmentation, edge detection, connectivity analysis, and finally ellipse detection methods are presented. This sequence is also identical to the image processing sequence of this study.

In Chapter 3, the target tracking method proposed to estimate the target position in the following frames is presented. Although the 3D reconstruction method comes before the target tracking method in the processing sequence, it is not included in

this study since it contains no remarkable modifications compared to the version in previous studies [1-4].

Chapter 4 identifies the internal camera parameters and lens distortion model. Determination of distortion parameters and distortion correction methods are also presented.

In Chapter 5, the overall structure of the 3D reconstruction method and the used hardware are explained. Two test procedures designed to measure the onsite performance of the method are discussed and the results of these test are presented.

Chapter 6 reviews and concludes the results of the chapters and presents recommendations for future work.

Appendix A explains the general features of the ImagePro 2.0 software, which is developed to test the performances of the methods proposed in this study and in previous studies [1-4]. A user manual for reconstruction and internal camera calibration is also presented.

In Appendix B, the characteristics of the elliptical and rectangular images in the generated synthetic image library are given. This synthetic image library is used to measure and tune the performance of the ellipse detection methods.

CHAPTER 2

TARGET DETECTION

2.1. Overview

The 3D reconstruction algorithm developed in [1] utilized a secondary passive target as shown in Figure 2.1. This target is a white perfect circular disk on a black background, with two black internal spots; one is located at the center of the circle. This algorithm can reconstruct the 3D parameters, namely three rotations, $\mathbf{R_x}$, $\mathbf{R_y}$, and $\mathbf{R_z}$, and three translations, $\mathbf{T_x}$, $\mathbf{T_y}$, and $\mathbf{T_z}$, of the target with respect to the camera coordinate system by using the contour of the elliptical image of this target, and image locations of two internal spots. This chapter covers the methods used starting from obtaining the image of a scene that includes the passive secondary target to determining the target's position in the image.
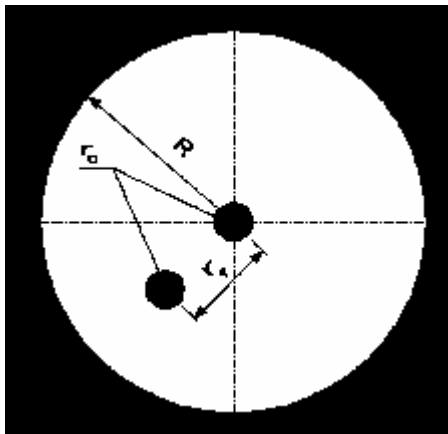
Figure 2.1. Secondary Passive Target

2.2. Image Formation and Grabbing

An optical image formation can be described as a perspective projection of the 3D world onto a 2D image plane by an optical system. The ideal case for a perspective projection occurs in a pin-hole camera as seen in Figure 2.2. The photons reflected by the objects pass through an infinitesimally small hole, fall onto the image plane, and form a very sharp image of the 3D scene. The problem of the pin-hole camera is that the illuminance at the image plane is also infinitesimally small due the infinitesimally small pin-hole, hence it necessitates very sensitive light sensors or very low shutter speeds to detect the incoming light.
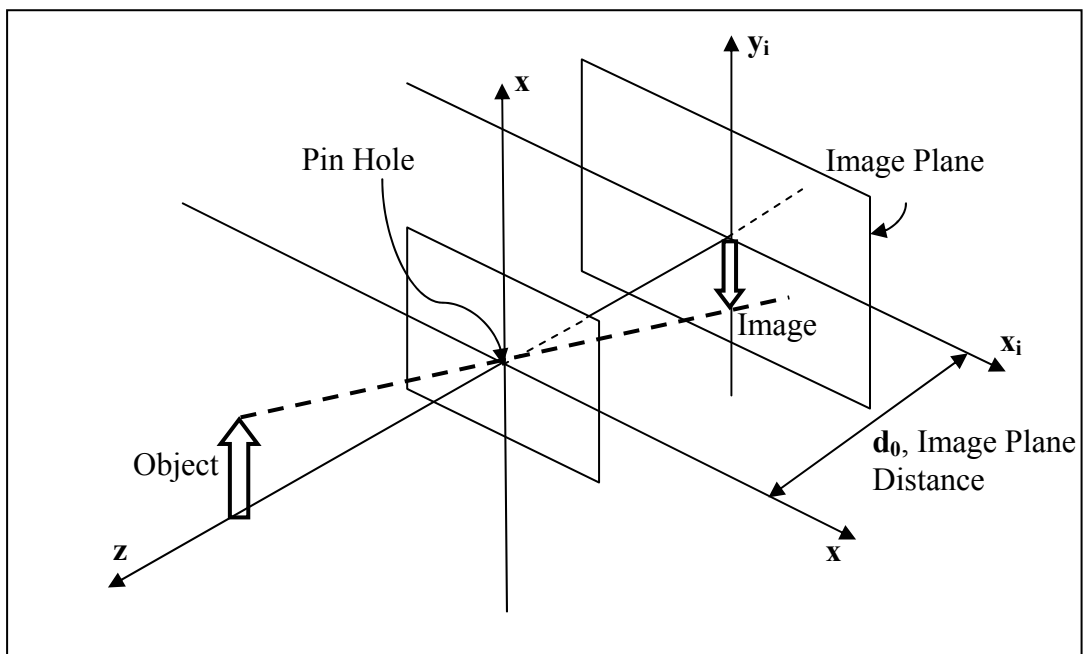
Figure 2.2. Pin-Hole Camera

All vision systems such as human eye, cameras, and VCRs utilize a single lens or a compound lens system to overcome this problem. A lens is an approximation to a pin-hole camera system with a finite amount of light. It collects finite amount of light rays, proportional with its diameter, reflected from a 3D point in a 3D scene into another point in the 3D world which is the image of the point as depicted in Figure 2.3. By relative positioning the image plane and lens such that the image

point falls onto the image plane, called focusing, the image of a 3D scene with a finite amount of illuminance is obtained on the image plane.
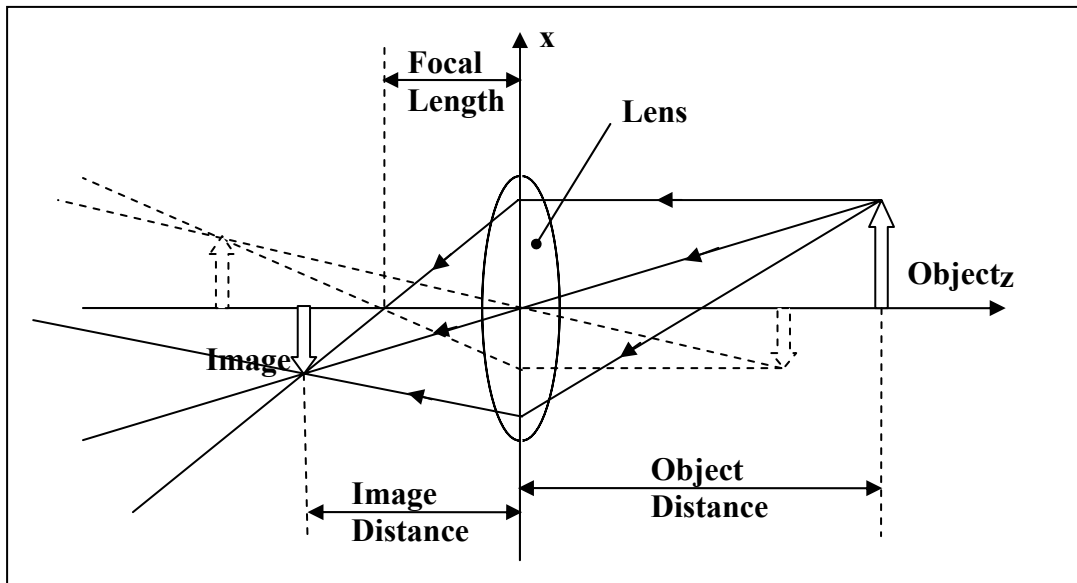
Figure 2.3. Simple Lens System

Lenses are characterized by their focal lengths, f-numbers and depth of fields.

The focal length of a lens is the distance of the focal point, at which the rays parallel to optical axis converge, to the lens. The focal length determines the image plane distance and magnification factor.

The f-number of a lens is the focal length divided by the diameter of the aperture of the lens. The illuminance is inversely proportional with the squared f-number.

The depth of field is the depth range behind and in front of a focused object, in which the other objects are still imaged properly focused. The depth of focus is the depth range around the image plane distance in which the object is still imaged properly focused. The depth of field is inversely proportional with the aperture size.

The image of a 3D scene formed on the image plane can be transformed into electrical information by light sensors for storing or transferring to another device.

13

There are various types of image sensors; but the most common ones are CMOS (complimentary metal oxide semiconductor), CCD (charge coupled device), and CID (charge induced device). Although all these sensors convert the intensity of light to an electric potential, the read-out technique of this potential differs. The CID sensors provide a non-destructive read-out, high quality and low noise images, anti-blooming, and contiguous pixel structure. On the other hand, CID cameras are rarely used and expensive. CMOS sensors are known with their low cost, very low power consumption. But their image quality is low and noise susceptibility is high. CCD type image sensors are the most commonly produced and used sensors. CCD type image sensors provide high quality and low noise images and their costs are reduced as a result of intensive research in recent years. The image data transformed into electric signals by these image sensors are then transferred to the digital environment by frame grabber devices.

This study implements three different image grabbing techniques as a combination of previous studies [1–4]: virtual perspective projection, digital image file, and frame grabbing.

Since the perspective projection is a well known process, the image of a scene can be formed virtually, without using a camera, if the geometry and color of the objects in the scene are known and if 3D configurations of the objects with respect to the camera coordinate system and internal camera parameters are fully identified. In this study, this process is named as "Virtual Perspective Projection".

A binary image (Figure 2.4) is generated by virtual perspective projection using the internal camera parameters and the geometry of secondary passive target for a set of 3D configuration parameters.

This virtual perspective projection was used in [1] to generate stand still images of the secondary target to test the accuracy of the developed 3D reconstruction algorithm. The accuracy levels so obtained give a theoretical lower bound to the expected accuracy levels in actual applications. Because, in the virtual perspective projection, the following effects are not considered:
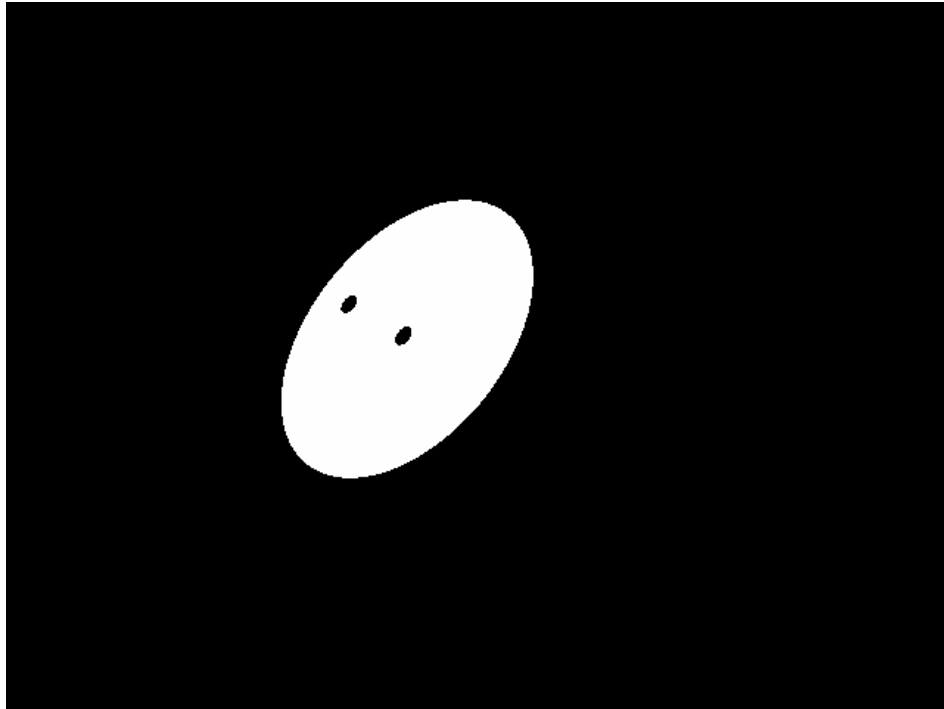
Figure 2.4. Binary Image Generated by Virtual Perspective Projection

- Effects of errors associated with target positioning.
- Effect of the environment illuminance on the image.
- Effect of lens distortion.
- Effects of uncertainties in the lens's focal length and image plane distance.
- Effects of sensor and data acquisition noise.
- Effects of image processing algorithms on the locations of pixels corresponding to the image of target.

In this study, the virtual perspective projection is used to generate multiple frames of a moving secondary passive target to test the 3D reconstruction and target tracking algorithm. These tests, too, give a theoretical lower bound to the expected accuracy levels in actual applications. Because, in addition to the effects listed above, the followings are not considered, either:

- Change in illuminance of the target due to its position change.
- Motion blurring for fast motions of the target.

The output of the implemented virtual perspective projection method is a binary (1-bit) image as given in Figure 2.4, in which the target pixels are white (1) and the background pixels are black (0). The pixels in this image corresponding to the elliptical contour and the spots of the secondary passive target are detected by an edge detection method.

In the previous study [2], a frame grabber device and a digital camera were utilized in the experimental setup. However, the developed software was not designed to access the frame buffer of the grabber device directly. Instead, the image of the scene containing the secondary passive target was captured using the utility software of the frame grabber device and saved as digital image files (Figure 2.5). These digital image files were utilized to measure the accuracy of the 3D reconstruction method in the previous study [2]. The previous studies [3, 4] also used some digital images of scenes containing the secondary passive target, to test the binarization, edge detection and ellipse (image of the circular secondary passive target) detection methods developed. The same digital image files are also utilized to compare and measure the performance of the enhanced or developed methods in this study and the methods developed in [2–4].

Figure 2.5. A Sample Digital Image File

As stated in Chapter 1, one of the objectives of this study is to end up with a real time working 3D reconstruction method such that the method should reconstruct 3D parameters of a moving secondary passive target at maximum sampling rate of the imaging device. Thus, this study also utilizes direct access to the frame grabber to get and to process grayscale (8-bit) frames corresponding to the 3D scene containing the moving secondary target (Figure 2.6). The proposed methods in this study are designed and forced to finish the 3D reconstruction process for one frame at a time before the following frame is formed by the grabber device.

Figure 2.6. A Frame Grabber Image.

2.3. Segmentation

Segmentation is the task of dividing an image into regions so that all the points of a planar homogenous surface of an object appear in the same region in the segmented image [19]. Segmentation methods can be divided into two groups: (i) histogram based segmentation and (ii) spatial coherence based segmentation. However, there is an agreement in the literature [19-23] that a totally correct segmentation is impossible without an intelligence or a priori information about the scene.

In this study, the segmentation is applied twice for a single image; first one is to the whole image (Figure 2.7) and second to the cropped image (Figure 2.8). The cropped image is the output of the target detection part, a small rectangular sub-region of the whole image containing the target only.
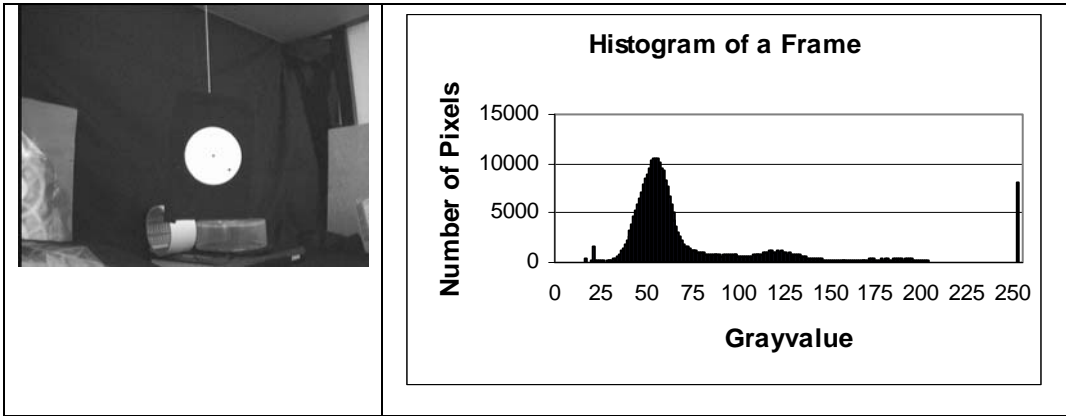
Figure 2.7. A frame and Its Histogram.
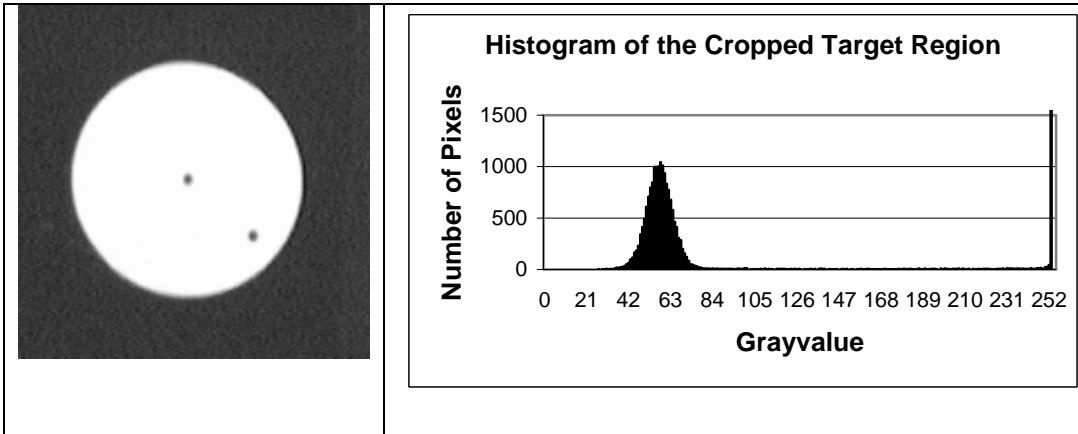
Figure 2.8. The Cropped Target Region and Histogram.

In the cropped image, the a priori information used is that the secondary passive target is a matte surface containing a white circle with a sufficiently large black background. This guarantees a high contrast edge for the circular boundary. Therefore, there exists a threshold value that can be used to segment the target from its background successfully and a histogram based threshold determination method solves the problem.

18

For the whole image (Figure 2.7), a grayscale (8-bit) image matrix obtained from a digital image file or from buffer of the frame grabber that is assumed to consist of a dark (low intensity) background and some bright (high intensity) foreground objects. This assumption implies that there exists a best threshold value between the minimum possible intensity value (0) and the maximum possible intensity value (255) for each image that correctly segments the background pixel and foreground pixels by the following binarization rule:

$$B(i, j) = \begin{cases} 1, & I(i, j) \geq t^* \\ 0, & I(i, j) < t^* \end{cases} \tag{2.1}$$

where B is the binary (black or white) value of a pixel in the image located at i^{th} row and j^{th} column, I is the grayscale value of this point, t^* is the threshold value.

Although this histogram-based segmentation of background and foreground objects seems to be a very simple problem to handle, it does not have a unique solution because of the choice of the threshold value t^*. The solution depends on the illumination conditions of the environment as well as the complexity of the scene.

The uncontrolled illumination conditions make it impossible to segment all the grayscale images into their foreground objects and background using a fixed threshold value. The problem here is to determine an optimum threshold value for each grayscale image that minimizes the segmentation error. In the earlier studies [3, 4], this problem was solved by the implementation of Kittler and Illingworth's minimum error thresholding method [21]. This method utilizes the histogram given in Figure 2.7 of the image to determine the gray value distributions of pixels of the object and the background, and an optimal, minimum error threshold value is determined using the statistical decision theory.

2.3.1. Kittler and Illingworth's Minimum Error Thresholding Method

The minimum error thresholding method proposed by Kittler and Illingworth [21] is a clustering-based thresholding technique. In this method, the histogram $h(g)$ of an image is assumed to be a compound of two Gaussian distributions, (μ_1, σ_1^2) and (μ_2, σ_2^2), with respective proportions P_1 and P_2. This histogram can be expressed as

$$h(g) = \sum_{i=1}^{2} P_i p(g \mid i),\tag{2.2}$$

where

$$p(g \mid i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(g - \mu_i)^2}{2\sigma_i^2}\right).\tag{2.3}$$

The optimum, minimum error threshold value t^* is selected to minimize the following criterion function $J(t)$.

$$J(t) = P_1(t)\ln\left(\frac{\sigma_1(t)}{P_1(t)}\right)^2 + P_2(t)\ln\left(\frac{\sigma_2(t)}{P_2(t)}\right)^2 \tag{2.4}$$

where

$$P_i(t) = \sum_{g=a}^{b} h(g) \tag{2.5}$$

$$\mu_i(t) = \frac{\sum_{g=a}^{b} h(g)g}{P_i(t)} \tag{2.6}$$

$$\sigma_i^2(t) = \frac{\sum_{g=a}^{b} [g - \mu_i(t)]^2 h(g)}{P_i(t)} \qquad (2.7)$$

$$a = \begin{cases} 0, & i = 1 \\ t+1, & i = 2 \end{cases} \qquad (2.8)$$

$$b = \begin{cases} t, & i = 1 \\ t_{max}, & i = 2 \end{cases} \qquad (2.9)$$

$$t \in \{0, 1, \ldots, t_{max}\} \qquad (2.10)$$

In the previous studies [3, 4], one of the segmentation approaches was the binarization of the entire grayscale image using the threshold value determined by the minimum error thresholding method and utilizing Equation (2.1). This implementation was named as GT (global thresholding), because a single threshold value determined from the histogram of the whole image is used to binarize the whole image. The GT method was successful in segmenting the secondary passive target form its background in grayscale images with almost uniform illumination.

If the illumination of the environment is uneven, a successful segmentation of the image into its background and foreground objects may not be possible by GT using a single threshold value. For instance, the background pixels in the upper left corner in the image may have greater intensity values than the foreground objects in the lower right corner. Therefore, a single fixed threshold value used in all over the image plane cannot segment the image successfully. This problem was solved in [3, 4] by applying the minimum error thresholding method locally. Two distinct local thresholding (LT) methods were developed: LT1 and LT2. In LT1, the image was partitioned into a set of non–overlapping, equally sized, square sub–images. For each quarter of the each sub-image, an optimum, minimum error threshold value was determined using the histogram of this quarter. Then, each sub–image was segmented using the average value of the four threshold values determined for its quarters. In LT2, a set of partially overlapping sub–images were used to enhance the smoothness of the binary global image at the sub-image borders.

The results of the GT, LT1, and LT2 were presented and discussed in [3, 4], and although all three implementations successfully separated the target pixels from the background, GT was selected as the most successful one. To evaluate the performances of these three approaches, means and variances of the two clusters (black pixels and white pixels) are calculated using grayscale values of each pixel inside a fixed region containing the target. The method that minimizes the sum of the variances of two clusters was selected as the most successful one.

2.3.2. Implementation Considerations

In this study, Kittler and Illingworth's [21] minimum error thresholding method is implemented as suggested in [3, 4] and tested. The segmentation performance of the method is found to be sufficient to separate the target from its background even in very poor lighting conditions. On the other hand, the implementation of the minimum error thresholding method described by Equations (2.4-2.10) reduces the processing speed. In the studies [3, 4], the means and variances, (μ_1, σ_1^2) and (μ_2, σ_2^2) of two clusters were calculated for each and every threshold value candidate, t, then the value of $J(t)$ for each t was evaluated. Repetitive summations in Equations (2.5-2.7) while calculating the means and variances, (μ_1, σ_1^2) and (μ_2, σ_2^2), for all possible t increase the process time.

If Equations (2.5-2.7) are rearranged as follows

$$\sigma_i^2(t) = \frac{\left[\left(\sum_{g=a}^{b} g^2 h(g)\right) - \left(2\mu_i(t)\sum_{g=a}^{b} gh(g)\right) + \left(\mu_i^2(t)\sum_{g=a}^{b} h(g)\right)\right]}{P_i(t)} \qquad (2.11)$$

$$\sigma_i^2(t) = \frac{\left(\sum_{g=a}^{b} g^2 h(g)\right)}{P_i(t)} - \mu_i^2(t) \qquad (2.12)$$

$$P_i(t+1) = P_i(t) + \begin{cases} h(t+1), & i=1 \\ -h(t+1), & i=2 \end{cases} \qquad (2.13)$$

$$\mu_i(t+1) = \frac{\left[\mu_i(t)P_i(t) + \begin{cases} (t+1)h(t+1), & i=1 \\ -(t+1)h(t+1), & i=2 \end{cases} \right]}{P_i(t+1)} \qquad (2.14)$$

$$\sigma_i^2(t+1) = \frac{\left(\left[\sigma_i^2(t) + \mu_i^2(t) \right]P_i(t) + \begin{cases} (t+1)^2 h(t+1), & i=1 \\ -(t+1)^2 h(t+1), & i=2 \end{cases} \right)}{P_i(t+1)} - \mu_i^2(t+1) \quad (2.15)$$

to calculate the means and variances of two clusters for $t + 1$ using the means and variances of the two clusters calculated for t, a good number of repetitive summations are avoided.

For a further speed enhancement, it is decided to construct the histogram matrix during image grabbing.

Moreover, the first and second moment vectors, $F(t)$ are $S(t)$ are evaluated before the threshold determination to avoid repetition of the multiplications.

$$F(t) = t \cdot h(t) \qquad (2.16)$$

$$S(t) = t \cdot F(t) \qquad (2.17)$$

In this study, two different sets of expressions for minimum error thresholding method are implemented and tested: the first implementation utilizes the original expressions as used in the previous studies [3, 4] and the second one uses the modified expressions.

23

2.3.3. Comparison of Thresholding Algorithms

The original and modified versions of minimum error thresholding method give exactly the same threshold values for the same test images as listed in Table 2.1. On the other hand, the modified expressions are processed approximately 45 times faster than original expressions as seen in the same Table. The resulting segmented images for a sample digital image and a frame are presented in Figure 2.9.

Table 2.1. Results of Minimum Error Thresholding Method

Image File	Threshold Value	Processing Time [ms]	
		Original Method	Modified Method
sample01	61	1.375	0.030
sample02	57	1.375	0.028
sample03	121	1.407	0.038
sample04	93	1.375	0.028
sample05	73	1.375	0.031
sample06	55	1.391	0.034
sample07	23	1.344	0.020
sample08	69	1.375	0.030
sample09	57	1.375	0.025
sample10	57	1.531	0.036
Average		**1.392**	**0.030**

Original Grayscale Image	Resulting Binary Image
sample01	sample01_bin
frame01	frame01_bin
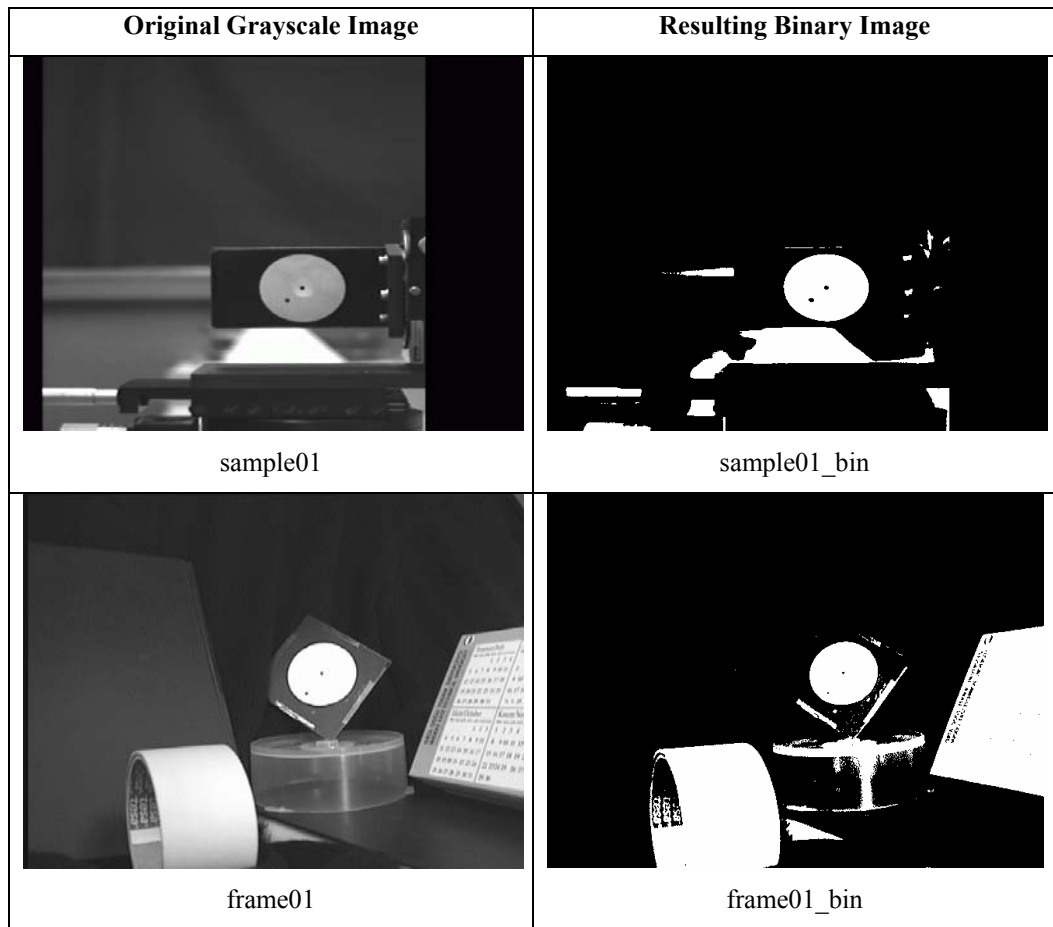
Figure 2.9. Examples of Segmentation

2.4. Edge Detection

Edge detection is a complementary operation to the segmentation, in which boundary contours of segmented regions are determined [22]. Large intensity changes are expected in regions corresponding to boundary crossings due the fact that the pixels of objects separated by these boundaries appear to have distinctly different intensity values. Most edge detectors utilize finite difference directional derivatives to detect these intensity changes in an image. Those locations in an image, where the magnitude of 2D first derivative vector (gradient) is larger than a preset value, are marked as edge pixels. The direction of this 2D first derivative vector becomes normal to its associated edge.

It is possible to apply edge detectors to grayscale or binary images. As in previous studies [3, 4], the edge detection method used in this study marks the contours of the regions in the binary image.

2.4.1. Sobel Filter

Sobel filter is one of the most commonly used 2D finite difference gradient operator [23]. By convolving the image data $\mathbf{B}(i, j)$ with horizontal and vertical Sobel filters

$$\mathbf{S_h} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{S_v} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{2.18}$$

two 2D arrays composed of horizontal

$$\mathbf{G_x}(x, y) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} \mathbf{B}(x+i, y+j)\mathbf{S_h}(i+2, j+2) \tag{2.19}$$

and vertical

$$\mathbf{G_y}(x, y) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} \mathbf{B}(x+i, y+j)\mathbf{S_v}(i+2, j+2) \tag{2.20}$$

components of the gradient vectors are obtained.

In the previous studies [3, 4], the binary image was convolved with Sobel filters, horizontal and vertical gradient component matrices, G_x and G_y, were constructed, and the pixels with magnitude of the gradient

$$\mathbf{E}(x, y) = \begin{cases} 1, & \mathbf{G_x^2}(x, y) + \mathbf{G_y^2}(x, y) > 0 \\ 0, & \mathbf{G_x^2}(x, y) + \mathbf{G_y^2}(x, y) = 0 \end{cases} \tag{2.21}$$

greater than zero were marked as edge pixels. Also the horizontal and vertical component of gradient vectors were constructed and stored to be used in the ellipse detection method. The resulting edge detected image for a sample binary image is presented in Figure 2.10.

Equation (2.21) yields edges thicker than one pixel, since the magnitude of the gradient vector is non-zero for all of the pixels having an inverted neighbor within its 8-neigborhood. Obviously, thicker contours so obtained from this edge detection method affect the outcomes of the ellipse detection method adversely, consequently yielding some false detections. Therefore, an edge thinning operation is required to avoid this problem. Moreover, if a binary image has already some one pixel thick objects, the Sobel edge detector gives edges on both sides of such objects as clearly seen in zoomed picture in Figure 2.10.

Binary Image	Edge Detected Image
sample03_bin	sample03_edge_Sobel
sample03_bin(zoomed)	sample03_edge_Sobel (zoomed)
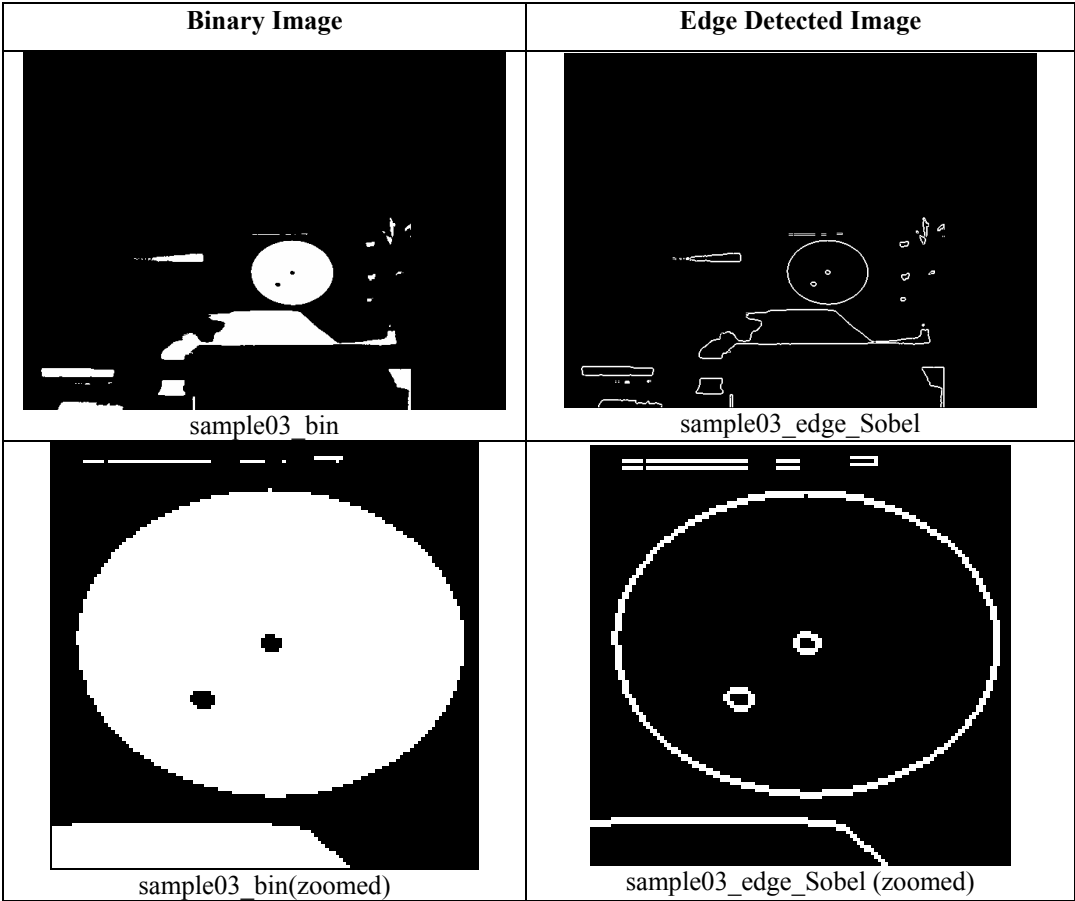
Figure 2.10. Sobel Edge Detector Results

2.4.2. Edge Thinning

In the previous studies [3, 4], a simple edge thinning operation, proposed in study [24], was implemented to thin the edges obtained by the Sobel edge detector. The method is based on the idea that on one pixel thick edges, there should not exist any three edge pixels combination that they are all in 8-pixels neighborhood of each other; and in order to obtain one pixel thick edges these three pixel combinations should be detected and one of the three pixels should be deleted. Although the process was simple, detecting all neighboring three pixels combinations for each edge pixel in the edge detected image was an exhaustive search. However, this method is also included into the current study to compare its performance with alternative solutions of the problem. Figure 2.11 shows the resulting edge thinned image. Although the thick edge problem is resolved, some deformations in the contours of internal spots of the target are noticeable. Therefore, this study utilizes the Laplacian kernel, as an alternative approach, to resolve this problem.
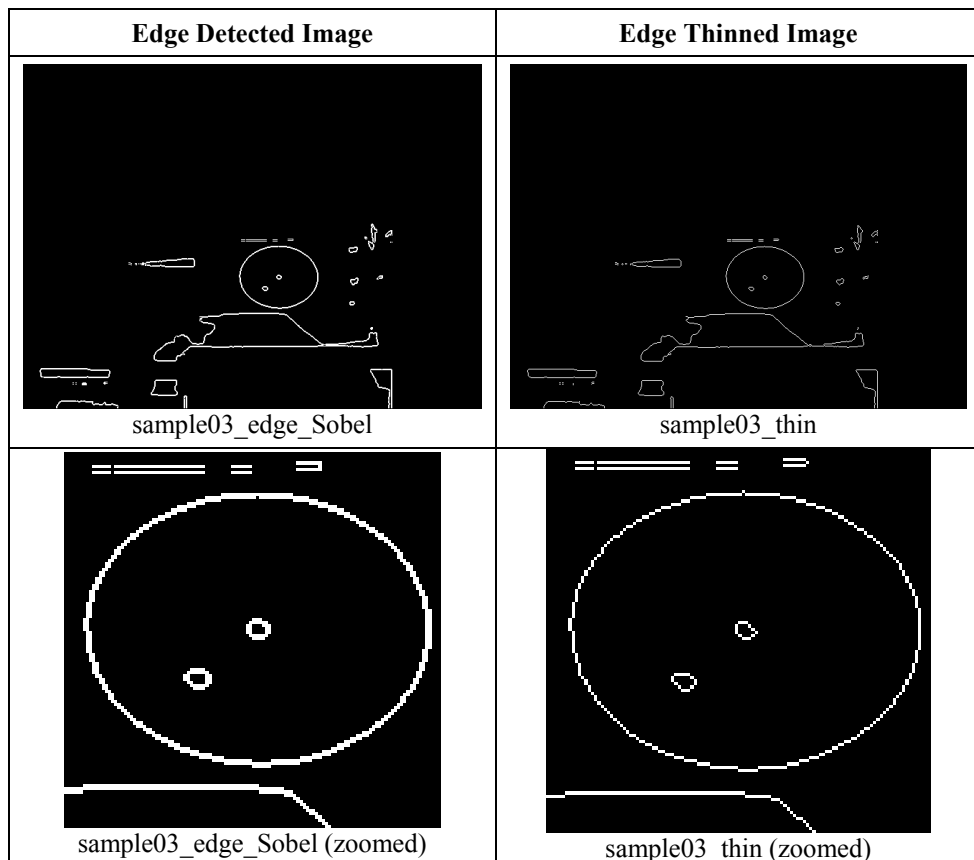
Figure 2.11. Edge Thinning Results

2.4.3. Laplacian Kernel

Laplacian kernel is actually a 2D second order finite difference operator. Edge pixels in an image are the points where the first derivative reaches a maximum or a minimum, in other words, the edge locations are zero crossings of the second derivative.

The convolution of a binary image data $\mathbf{B}(i, j)$ with the Laplacian kernel

$$\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.22}$$

can be expressed as

$$\mathbf{S_D}(x, y) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} \mathbf{B}(x+i, y+j)\mathbf{L}(i+2, j+2) \tag{2.23}$$

where $\mathbf{S_D}(x, y)$ is identical to sum of second derivatives in horizontal and vertical directions at the pixel in the binary image located at x^{th} row and y^{th} column. The values of $\mathbf{S_D}(x, y)$ are zero at constant or linearly changing regions and gives non-zero values at edge pixels as

$$\mathbf{E}(x, y) = \begin{cases} 1, & \mathbf{S_D}(x, y) \neq 0 \\ 0, & \mathbf{S_D}(x, y) = 0 \end{cases} \tag{2.24}$$

where $E(x, y)$ is the edge pixel data.

In the implementation of this Laplacian kernel, it is modified as

$$\mathbf{E}(x, y) = \begin{cases} 1, & \mathbf{S_D}(x, y) < 0 \\ 0, & \mathbf{S_D}(x, y) \geq 0 \end{cases} \tag{2.25}$$

to completely remove the thick edge problem when detecting edge pixels in a binary image.

Equations (2.23) and (2.25) imply that only white pixels may be the edge pixels if they have at least one black neighbor within their four pixel neighborhood. To reduce the process time for this edge detection method, only white pixels in the binary image are convolved with Laplace kernel. Actually, this convolution is identical to summation of the four neighboring pixels.

By using this modified Laplace kernel, the edges obtained do not require any additional edge thinning operation and the thick edge problem is resolved. However, Laplace kernel does not compute gradient vectors at the edge pixels required by the ellipse detection method proposed in the previous studies [3, 4]. So the gradient vectors are calculated using the Equations (2.19) and (2.20) at the edge pixels detected by the Laplace kernel. The results given in Figure 2.12 show that the Laplace edge detector does not cause any edge contour deformation in internal spots.

Binary Image	Edge Detected Image
 sample03_bin	 sample03_edge_Laplace
 sample03_bin(zoomed)	 sample03_edge_Laplace (zoomed)
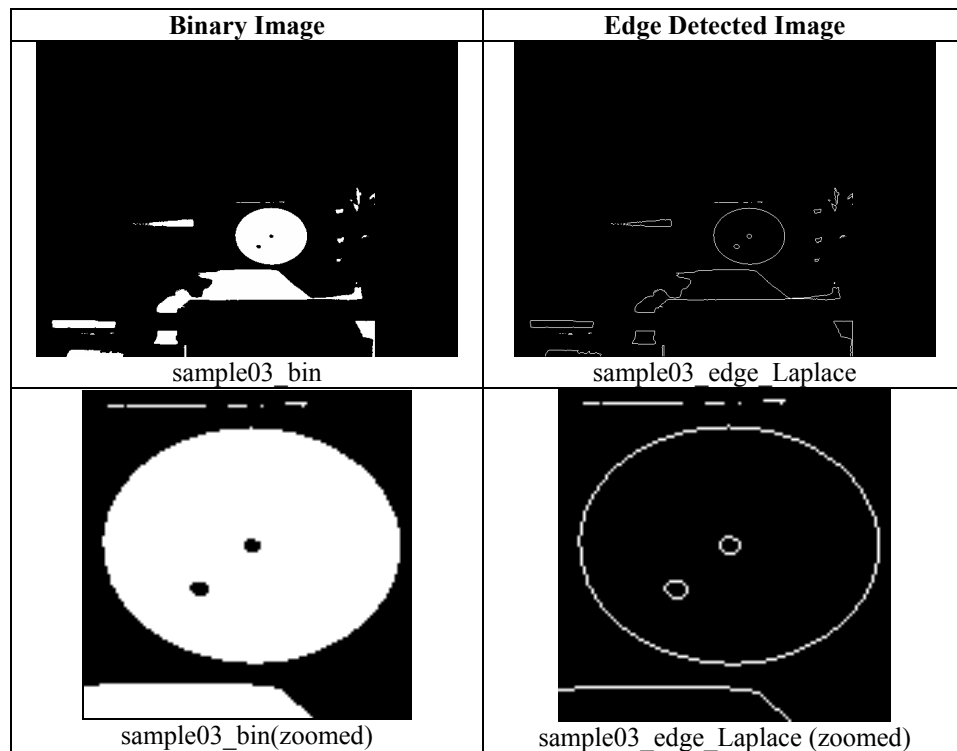
Figure 2.12. Laplace Edge Detector Results

30

2.4.4. Comparison of Edge Detection Algorithms

The processing times for two approaches implemented for edge detection are presented in Table 2.2. It is obvious that the implementation that utilizes Laplace operator to detect edges and Sobel operator to determine edge orientations at detected edge locations is approximately 13 times faster than the Sobel filter followed by an edge thinning operation. Although the resulting edge profiles seem similar, the resulting edge detected images for Sobel filter followed by an edge thinning operation the contours are deformed and some extra edge points are created as seen in Figure 2.13. When the speed of processing and the resulting contours are compared, it is obvious that the Laplace operator approach has a better performance.

Table 2.2. Processing Times of Edge Detectors

Binary Image	Processing Time [ms]	
	Sobel +Thinning	Laplace
sample01_bin	3.43	0.28
sample02_bin	3.43	0.30
sample03_bin	3.28	0.22
sample04_bin	3.28	0.16
sample05_bin	3.28	0.27
sample06_bin	3.43	0.31
sample07_bin	3.28	0.20
sample08_bin	3.28	0.25
sample09_bin	3.28	0.17
sample10_bin	3.28	0.24
Average	**3.33**	**0.24**
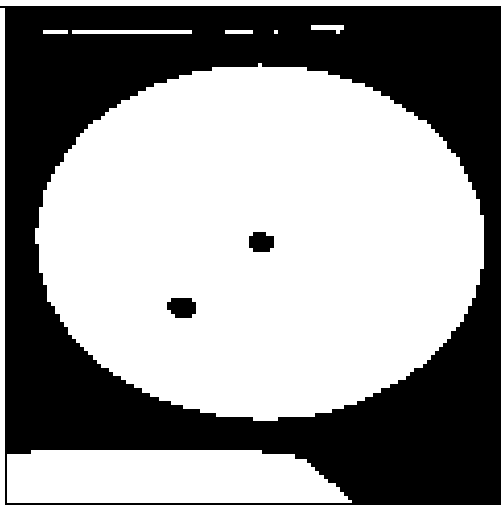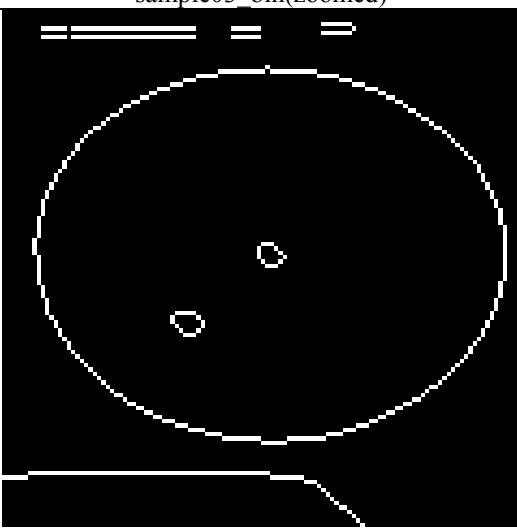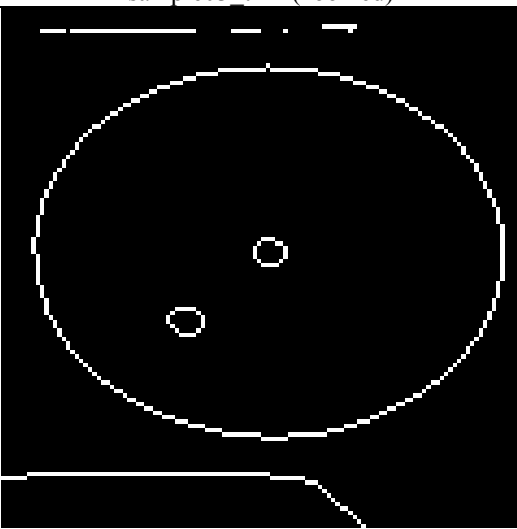
Binary Image	 sample03_bin(zoomed)
Sobel Edge Detected and thinned Image	 sample03_thin (zoomed)
Laplace Edge Detected Image	 sample03_edge_Laplace (zoomed)

Figure 2.13. Comparison of Edge Detection Results

2.5. Connectivity

Once the edge pixels are detected as borders of the segmented regions, the ellipse detection method developed in this study decides on the specific group of edge pixels that correspond to an ellipse, which will ultimately be identified as the image of the circular target. Thus, the edge pixels which are the borders of the same region or object should be classified into the same group to simplify and speed up the ellipse detection process.

In the previous studies [3, 4], a connectivity analysis was used to cluster pixels belonging to the border of the same object into one group. But, this connectivity analysis was not well-implemented. The results showed that the connectivity analysis did not satisfy some design requirements. For example, the number of pixels in a connected path was limited by a fixed maximum number, sometimes causing boundaries for disconnected regions. Also the algorithm did have some bugs such that the same pixel may appear in more than one connected pixel groups and some pixels did not appear in any of the connected groups at all.

In this study, the connectivity algorithm designed in the previous studies [3, 4] is re-arranged, its bugs and limitations of are removed, and it is named as CN0 to compare its results with the newly designed one.

2.5.1. 8-Neighbors Approach

The modified Laplace kernel classifies all white pixels having an at least one black pixel within its 4-neighborhood as edge pixels seen in Figure 2.14. Since the connectivity method is desired to result a single connected path for the edge profile of the binary region shown in Figure 2.14, an 8-neighborhood connectivity approach should be used.
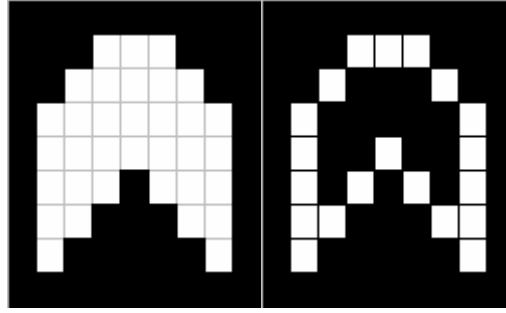
Figure 2.14. A Binary Region and its Edge Pixels Detected by Laplace Kernel

The connectivity analysis method, named as CN1, developed and implemented in this study, utilizes conductivity like logic to identify the path number of each pixel. First of all, the output structure of the edge detection method is organized as a pixel array to speed up the connectivity analysis. The first pixel in the list is assigned to the first group. Then the neighbors of this pixel are transferred to a position just after the first pixel in the list and assigned to the first group. This process is repeated for each pixel in the list until an unassigned pixel is reached. This pixel is assigned to the second group and steps applied for the first group is repeated. The overall process is continued until the end of edge pixel list.

The main advantage of this type of connectivity is its reduced computation time due to fact that it is completed in only one pass. Also, the list obtained just after the connectivity does not require any preprocessing for the ellipse detection algorithm to be used as the next step.

2.5.3. Results of the Connectivity Analysis

The process times of two connectivity algorithms, improved version of the algorithm developed in previous studies [3, 4] (CN0) and the algorithm developed in this study (CN1) are presented in Table 2.3. Some resulting connected images can be seen in Figure 2.15, in which each connected path is plotted by a different color. The resulting paths for CN0 and CN1 are the same and both algorithms correctly identify the connected paths.

34

	CN0	CN1
sample03		
sample04		
sample06		

Figure 2.15. Connectivity Analysis Results

But, due to the fact that the computational order of the algorithm used in the previous study is n^2, the process time is highly affected by the number of edge pixels in the edge detected image. As seen in Table 2.3, the process time of CN0 varies from 2.34 ms to 189 ms, whereas for CN1 as an order n algorithm the process time varies between 2.18 ms and 3.9 ms for the sample images. The CN1 algorithm is approximately 20 times faster than CN0.

Table 2.3. Processing Times of Connectivity Analysis Methods

Edge Detected Image	Processing Time [ms]	
	CN0	CN1
sample01_edge_Laplace	139.53	3.44
sample02_edge_Laplace	141.09	3.59
sample03_edge_Laplace	21.56	2.50
sample04_edg_Laplace	2.34	2.18
sample05_edge_Laplace	19.85	2.65
sample06_edge_Laplace	189.06	3.90
sample07_edge_Laplace	14.69	2.50
sample08_edge_Laplace	15.00	2.50
sample09_edge_Laplace	5.15	2.35
sample10_edge_Laplace	33.75	2.66
Average	**58.20**	**2.83**

2.6. Ellipse Detection

In image processing studies, ellipse detection is one of the most commonly encountered problems. In this study, the image of a circular target positioned in 3D space is an ellipse, if it is not partially occluded. Therefore, those pixels corresponding to the border of the image of the circular target should be detected and used by the 3D reconstruction method.

Various methods were developed to recognize ellipses in an image. There are mainly two approaches: (i) clustering and (ii) ellipse fitting. Clustering approaches utilize the mapping of pixel data to a parameter space according to a predetermined relationship. On the other hand, ellipse fitting approaches utilize some direct ellipse fitting methods to the data. In the previous studies [3, 4], Bennett's algorithm [25], which is a clustering approach and Hough transform based method, was implemented because it does not require any object based pre-segmentation, it has a 2D accumulation space, and it has a lower complexity.

2.6.1. Bennett's Approach

Bennett's approach [25] parameterizes a family of ellipses, which are tangent to two lines at two points as shown in Figure 2.16. Using the equations of these tangent lines and the equation of the line connecting these two points, the quadratic equation for a family conics can be written as

$$C(x, y) = L^2(x, y) - \lambda I_1(x, y) I_2(x, y) = 0 \qquad (2.26)$$

where, λ is a positive real coefficient, $L(x, y) = 0$ is the line connecting P_1 and P_2, and $I_1(x, y) = 0$ & $I_2(x, y) = 0$ are the tangent lines at points P_1 and P_2, respectively.
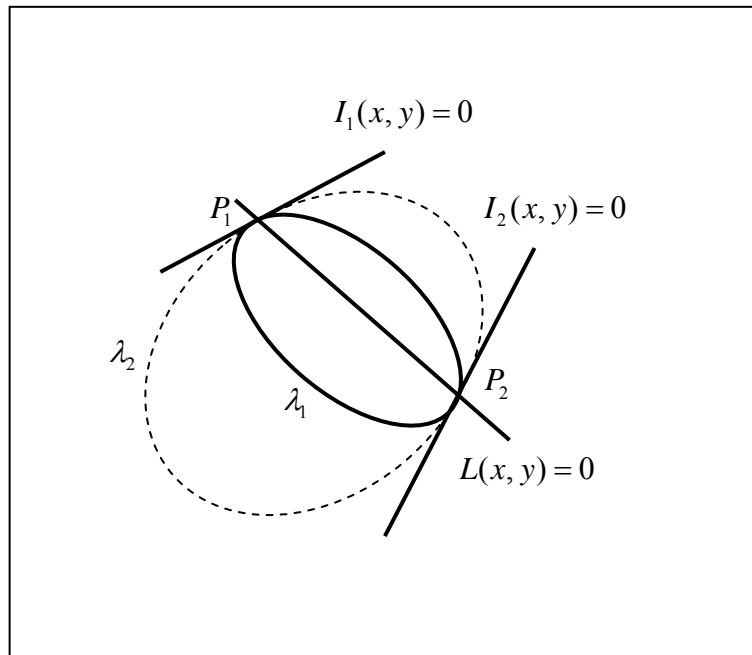
Figure 2.16. Ellipses Parameterized by Bennett's Approach

On the other hand, the equation of a conic can also be written in the form

$$\mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x} + 2 \cdot \mathbf{k}^T \cdot \mathbf{x} + c = 0 \qquad (2.27)$$

37

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \ \mathbf{k} = \begin{bmatrix} f \\ g \end{bmatrix} \ \text{and} \ \mathbf{A} = \begin{bmatrix} a & h \\ h & b \end{bmatrix} \quad\quad (2.28)$$

where a, b, c, f, g, and h are linear functions of λ and depend on the coordinates of points P_1 and P_2, and slopes of the tangent lines to edge at these points. The matrix $\mathbf{A}$ should be positive definite if the conic equation represents an ellipse. A λ range, in which matrix $\mathbf{A}$ is positive definite, is calculated for each pair of edge points P_1 and P_2. Then, centers of the ellipse corresponding to each λ in the calculated range are accumulated in a matrix. The point which has the maximum score in the accumulation matrix is declared as the ellipse center. The order of the algorithm is n^2 since the process is repeated for each pair of points.

In the previous studies [3, 4], Bennett's algorithm [25] was implemented with some modification for an increased speed and robustness. First problem encountered in those studies was the huge computation time due to the fact that the order of the algorithm is n^2. By the implementation of connectivity analysis, edge pixels were segmented to connected groups. Although the total number of points remained unchanged, the number of all possible point pairs was reduced, thus the computation time was decreased. The other problem encountered was deciding on the location of the ellipse center thus the ellipse itself by inspecting the accumulation space, as pictured in Figure 2.17. Even for the synthetically generated elliptical images, there exists a possibility to have some local maximum points around the exact ellipse center some of these false peaks may have greater scores than the exact ellipse center due to the discrete pixel locations and edge orientations. Moreover, there is a great possibility of having much higher peaks for long but non-elliptical connected paths. Quality factor and symmetry criteria were added to decision making process in the ellipse detection method to detect the correct peak that represents the real ellipse center. The quality factor of a peak in accumulation space was determined as the score of the peak divided by the square

38

of the distance of the closest point to the peak whose score was half of the peak. The symmetry was measured as the distance at which the symmetry of the peak was vanished. The peaks sharper than an experimentally determined value and peaks having symmetry values less than a fixed value were evaluated as noise and eliminated. Peaks smoother than a constant value, corresponding some low quality factors, were also eliminated because of the fact that they could not represent ellipse centers.
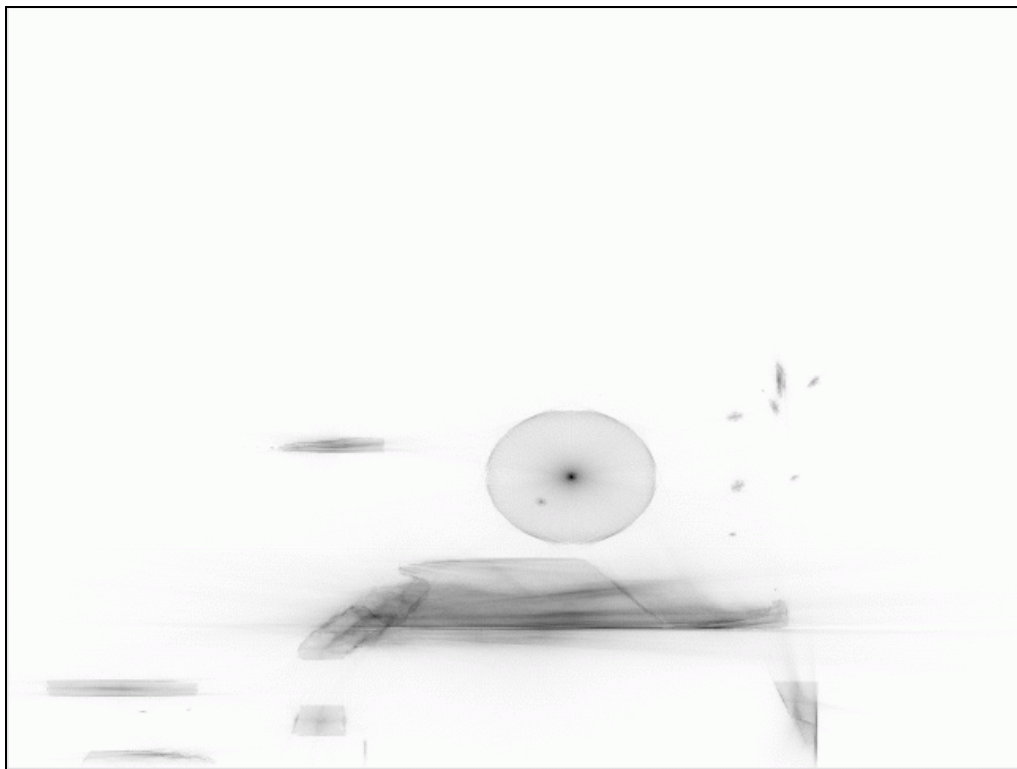
Figure 2.17. Accumulation Space for Sample03

In the previous studies [3, 4], even with some subjective modifications discussed above, the results of Bennett's algorithm were unsatisfactory and it could detect ellipses in only 30 % of test images and the computation time did not fall below 15 seconds. The main reason for the failure of the implemented Bennett's approach is found to be the faulty results of the improperly implemented connectivity analysis. Also the edge thinning operations deflected the ellipse contour unexpectedly. The use of a single accumulation space for all connected paths was another problem,

since the accumulation results of the paths deteriorates the quality factor and symmetry measurement of the peak corresponding to other paths. Moreover, determining the local maximum points in the accumulation space consumed an extra processing time.

In this study, Bennett's algorithm [25] is re-implemented without the modifications proposed in previous studies and a synthetic image library, composed of images containing a single ellipse or a rectangle of different size, orientation and aspect ratio, is formed to test and determine the limitations of this method. The characteristics of the images in the synthetic image library are presented in Appendix B. Experiments are conducted to determine characteristics of the accumulation space for synthetic elliptical and rectangular images to design a better decision method. The number of edge pixels in synthetic images can be seen in Figure 2.18.
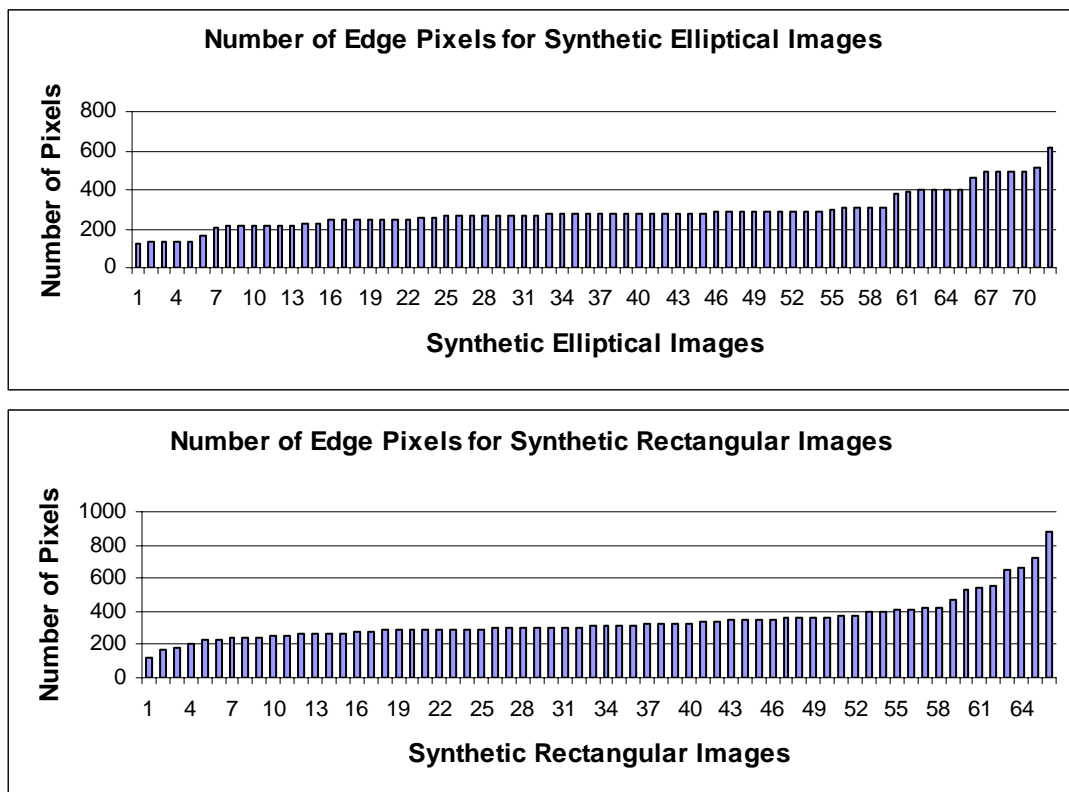
Figure 2.18. Number of Edge Pixels in Synthetic Image Library

The absolute maximum of the accumulation space for synthetic elliptical images appears at the exact ellipse centers. On the other hand, the absolute peak of accumulation space for rectangular images can either be at the centers of the rectangles or one of the edge centers. In Figure 2.19, accumulation spaces for an elliptical and a rectangular image in the synthetic image library are presented.
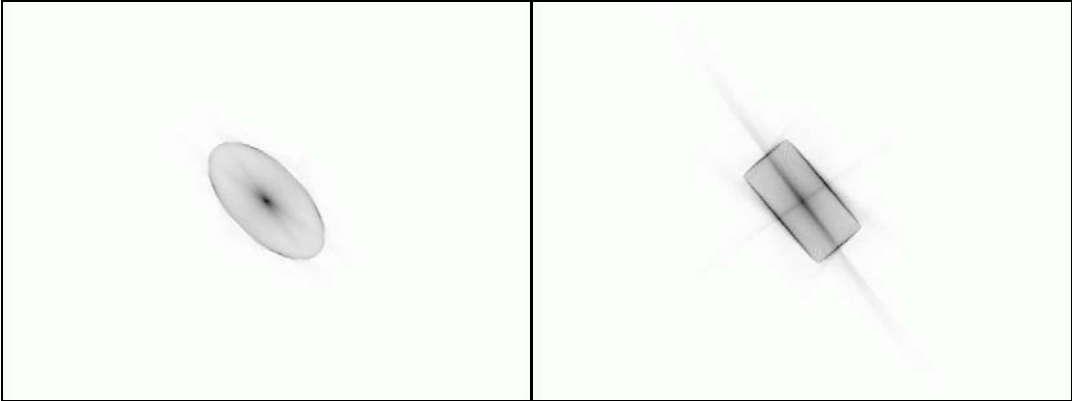
Figure 2.19. Accumulation Space for an Elliptical and a Rectangular Image in the Synthetic Image Library

The symmetry measurement of the absolute peak of the accumulation space is not affected by the shape, size, orientation and aspect ratio of the ellipses or rectangles as shown in Figure 2.20. The symmetry of the absolute peak in the accumulation space for ellipses is vanished at the closest image border, except two images. In these two images the location of absolute maximum in the accumulation space is deviated from the exact ellipse center due to numerical errors. For rectangles, the symmetry does not vanish till to the closest edge, only if the absolute maximum is at rectangle center. The only decision rule can be stated with these facts is that if the symmetry of the absolute peak in an accumulation space for a connected path vanishes at a point closer than the closest image border, then the connected path can not be an ellipse or a rectangle.
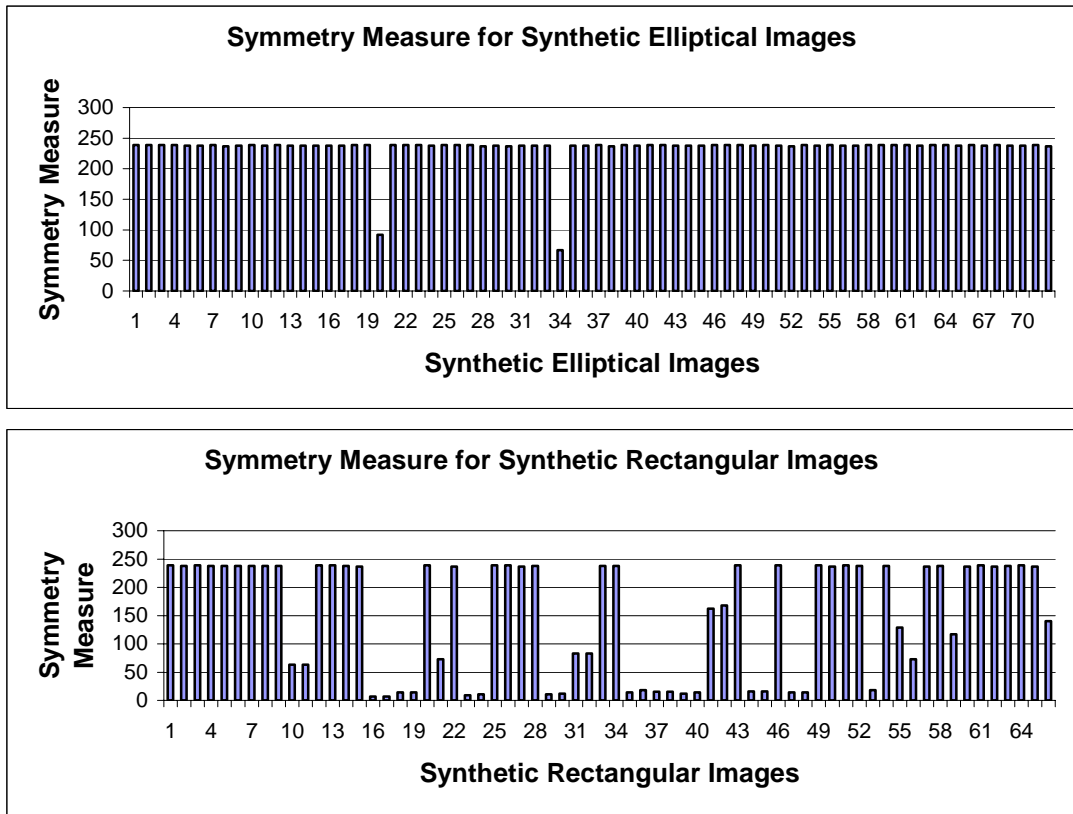
Figure 2.20. Symmetry Measure in Synthetic Images

The score of the absolute maximum of the accumulation space (Figure 2.21) is not proportional with the number of points in the connected path and is not related with the shape of the connected path. The score of the peak in the accumulation space for a smaller rectangle can be higher than that of a larger ellipse. However, if those eight elliptical images with the highest accumulation scores are neglected, the maximum accumulation scores for elliptical images are observed between 550 and 840. The connected paths with accumulation scores out of this range can be classified as non-ellipses. With this decision rule, 54 rectangles out of 66 are eliminated while only the neglected elliptical images are misclassified.
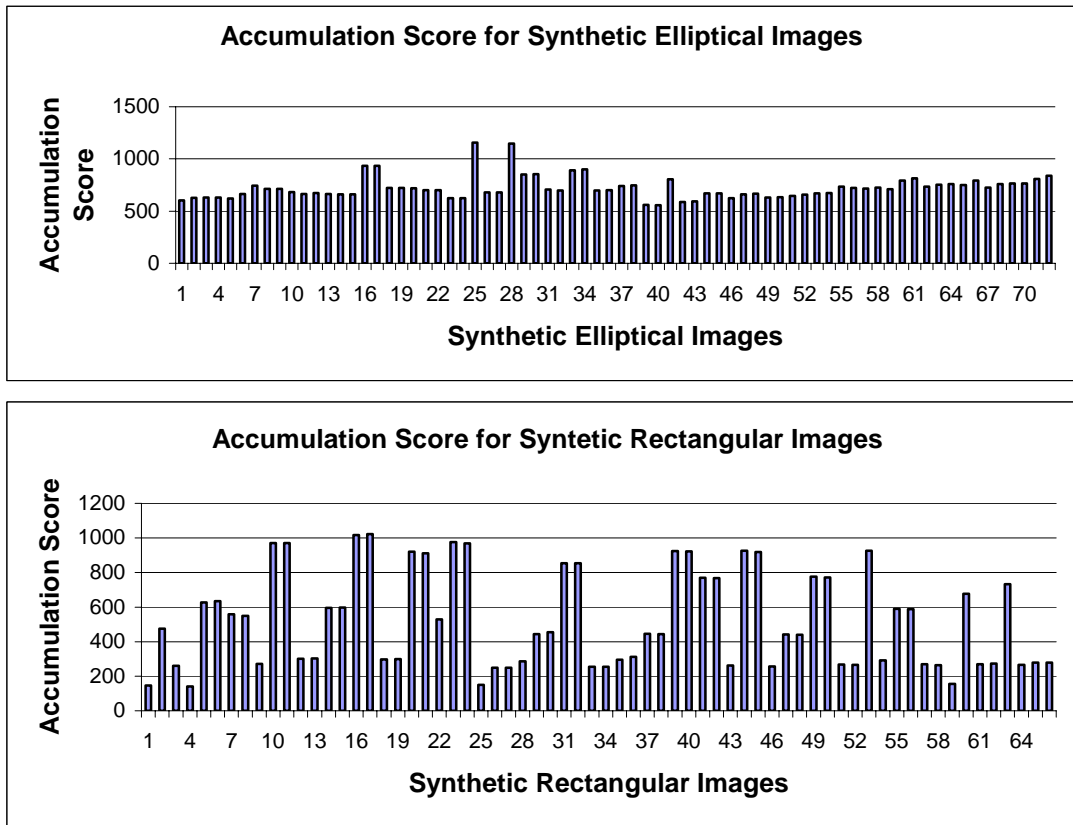
Figure 2.21. Maximum Accumulation Scores in Synthetic Images

In the elliptical images, the relation between the maximum scores of the accumulation space and the radius of the circle on which the average accumulation score is half of the peak's score is extracted and a new quality factor is designed to give almost constant values for the peaks of the accumulation scores for elliptical images

Considering the above facts, the new quality factor is defined as the accumulation score of the peak divided by square root of the radius of the circle on which the average score is half of the peak's value. In Figure 2.22, the accumulation space of a synthetic elliptical image is presented. In this figure, the average of the accumulation scores of the points on the blue circle is half of the score of absolute peak, which is the center of this circle. The new quality factor is calculated as peak's score divided by square root of the radius of this blue circle.
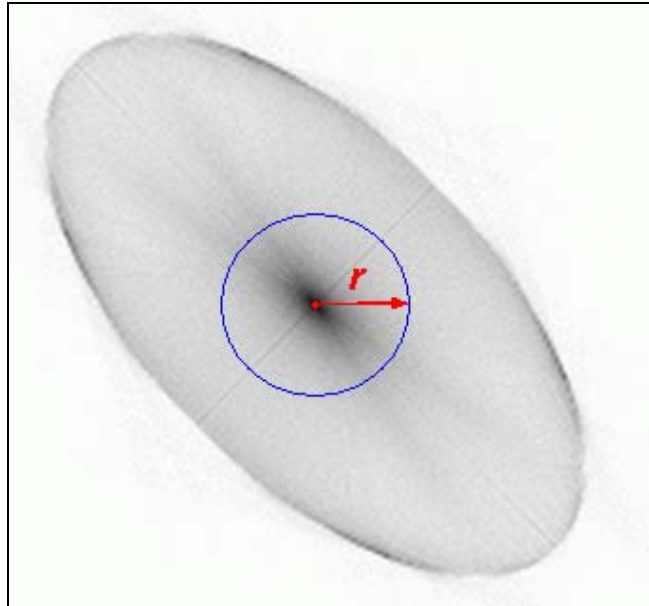
Figure 2.22. Quality Factor Visualization

In Figure 2.23, the quality factors of the absolute peaks of the accumulation spaces for the synthetic elliptical and rectangular images calculated with this definition are presented. Figure 2.23 implies that the quality factor of an ellipse within the size range shown in Figure 2.18 is almost constant, except for eight highest quality elliptical images; note that these images are also the outliers for accumulation score.

The distributions of the quality factors for synthetic elliptical and rectangular images (Figure 2.23) show that the elliptical and rectangular images can not be classified using a single fixed quality factor threshold. But if the eight elliptical images with highest quality factors are neglected, then one can state that the peaks of the accumulation space for elliptical images have quality factors within the range 255 to 375, and there are only eight rectangular images which have the quality factor within this range.
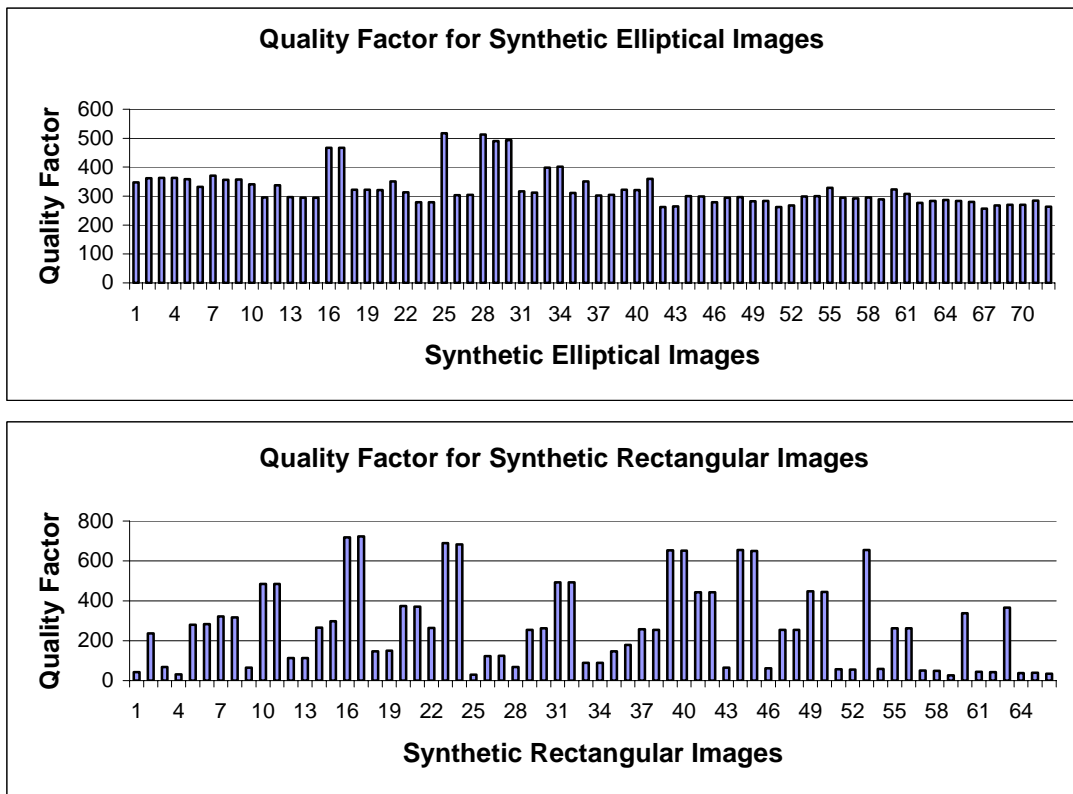
Figure 2.23. Quality Factor in Synthetic Images

The very first inference of the results of the experiments is that by using distinct accumulation spaces for each connected path the inspection of accumulation space is simplified and the accumulation results of connected paths will not affect each other.

The maximum accumulation scores, symmetry measurement and quality factor results obtained for the elliptical and rectangular images in synthetic image library imply the following ellipse detection algorithm:

1. Eliminate the connected paths for which the symmetry of the absolute peak of the accumulation space for the connected path vanishes at a closer point than the closest image border.

2. Eliminate the connected paths whose quality factor is not between 255 and 375.

3. Eliminate the connected paths whose accumulation score is not between 550-and 840.

4. Mark the connected path with the maximum accumulation score from remaining paths as proposed by Bennett [22].

The decision rules stated above result in 7 misclassifications for 66 rectangular images and 9 misclassifications for 72 elliptical images. The detection performance of this ellipse detection method is questionable, since rectangle is one of the most commonly encountered shapes in almost every environment. Also, this method and the decision rule will automatically eliminate all partial elliptical images, because of the fact that the resulting accumulation space would not be symmetric. However the method succeeded to detect the full ellipses in the all eight sample real images. One of the ellipse detected images is presented in Figure 2.24. In the figure, the numbers at the center of the ellipse are the accumulation score, quality factor and symmetry measurement, from top to bottom.
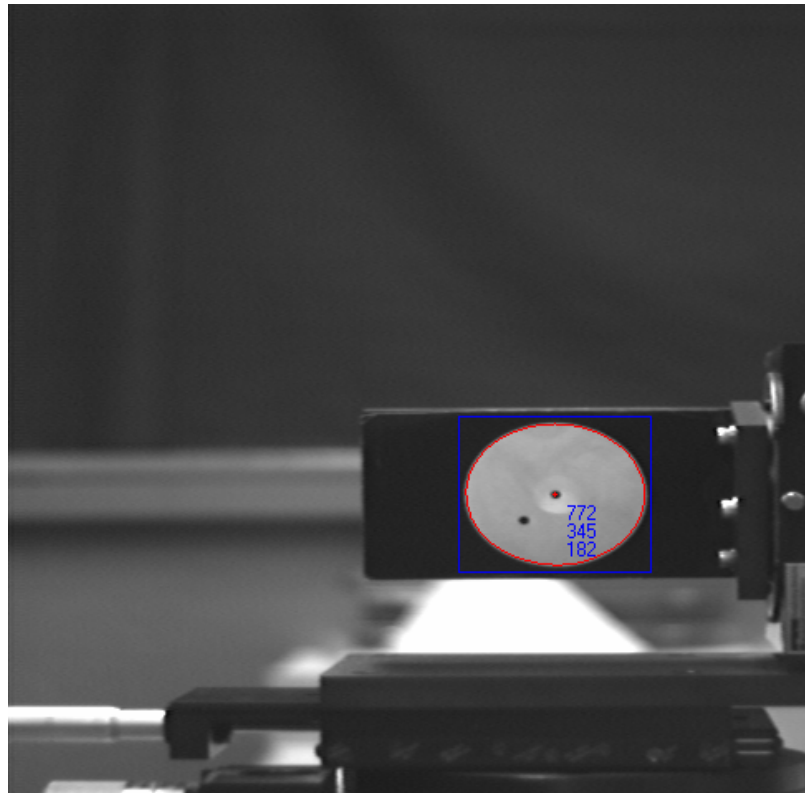
Figure 2.24. Detected Ellipse in Image Sample03.

The processing times for the ellipse detection method by Bennett's algorithm for sample real images are presented in Table 2.4. The results show that the speed of the method is highly dependent to the number of edge pixels and number of paths in an image.

Re-implementation of the method and the modifications discussed above increased the average processing time for this method approximately 10-15 times if compared with the speed of the implementation in previous studies [3, 4]. However the reached average processing time, 1473 ms or even the minimum processing time, 243 ms achieved for sample images is not an acceptable value for real time working.

Table 2.4. Processing Times for Ellipse Detection Method by Bennett's Algorithm

Image	Processing Time [ms]
sample01	1406
sample02	3156
sample03	1672
sample04	250
sample05	266
sample06	1375
sample07	907
sample08	2906
sample09	214
sample10	2578
Average	**1473**

2.6.2. Direct Ellipse Fitting

Direct ellipse fitting techniques concentrate on finding a set of parameters that minimize the distance between data points and the ellipse. Most studies conducted on ellipse fitting uses the following conic equation:

$$F(\mathbf{a},\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = ax^2 + bxy + cy^2 + dx + ey + f = 0 \qquad (2.29)$$

where $\mathbf{a} = \begin{bmatrix} a & b & c & d & e & f \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x^2 & xy & y^2 & x & y & 1 \end{bmatrix}^T$.

$F(\mathbf{a},\mathbf{x_i})$ is the algebraic distance of a point (x_i, y_i) to the conic $F(\mathbf{a},\mathbf{x}) = 0$. If the discriminant $b^2 - 4ac$ is negative, then the conic Equation (2.29) represents an ellipse. The parameter vector $\mathbf{a}$ that minimizes the sum of square of algebraic distances

$$D_A(\mathbf{a}) = \sum_{i=1}^{N} F(\mathbf{a},\mathbf{x_i})^2 \qquad (2.30)$$

for N data points $\mathbf{x_i}$, is selected as the parameter vector for such a conic. In order to avoid the trivial solution $\mathbf{a} = \mathbf{0}$ and to avoid the solutions which are the multiples of another solution, the parameter vector $\mathbf{a}$ is constrained.

In this study, the ellipse fitting method developed in [26] is implemented. In the study [26], the ellipse fitting problem is reduced to minimizing the equation,

$$E = \|\mathbf{D} \cdot \mathbf{a}\|^2 \qquad (2.31)$$

subjected to the constraint $4ac - b^2 = 1$,

$$\mathbf{a}^T \cdot \mathbf{C} \cdot \mathbf{a} = 1 \qquad (2.32)$$

48

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{x_1} & \mathbf{x_2} & \dots & \mathbf{x_N} \end{bmatrix}^{\mathrm{T}} \qquad (2.33)$$

and

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (2.34)$$

Introducing the Lagrange multiplier λ and differentiating, following system of simultaneous equations are obtained.

$$2 \cdot \mathbf{D}^{\mathrm{T}} \cdot \mathbf{D} \cdot \mathbf{a} - 2 \cdot \lambda \cdot \mathbf{C} \cdot \mathbf{a} = 0$$
$$\mathbf{a}^{\mathrm{T}} \cdot \mathbf{C} \cdot \mathbf{a} = 1 \qquad (2.35)$$

This system of equations can be rewritten in the following form.

$$\mathbf{S} \cdot \mathbf{a} = \lambda \cdot \mathbf{C} \cdot \mathbf{a} \qquad (2.36)$$

$$\mathbf{a}^{\mathrm{T}} \cdot \mathbf{C} \cdot \mathbf{a} = 1 \qquad (2.37)$$

where

$$\mathbf{S} = \mathbf{D}^{\mathrm{T}} \cdot \mathbf{D} \qquad (2.38)$$

is referred to as the scatter matrix. The system is solved by considering the generalized eigenvectors of Equation (2.36). If a generalized eigenvalue and eigenvector couple $(\lambda_i, \mathbf{u_i})$ solves Equation (2.36) then parameter vector $\mathbf{a} = \mu_i \cdot \mathbf{u_i}$ solves Equation (2.37) as

$$\mathbf{S} \cdot \mathbf{u_i} = \lambda_i \cdot \mathbf{C} \cdot \mathbf{u_i} \qquad (2.39)$$

$$\mu_i^2 \cdot \mathbf{u_i}^T \cdot \mathbf{C} \cdot \mathbf{u_i} = 1 \qquad (2.40)$$

and if Equation (2.39) is pre-multiplied by $\mathbf{u_i}^T$

$$\mathbf{u_i}^T \cdot \mathbf{S} \cdot \mathbf{u_i} = \lambda_i \cdot \mathbf{u_i}^T \cdot \mathbf{C} \cdot \mathbf{u_i} \qquad (2.41)$$

Since the scatter matrix $\mathbf{S}$ is positive definite, then the left hand side of Equation (2.41), $\mathbf{u_i}^T \cdot \mathbf{S} \cdot \mathbf{u_i}$ is positive. Furthermore, $\mathbf{u_i}^T \cdot \mathbf{C} \cdot \mathbf{u_i}$ should be positive to satisfy Equation (2.40). Hence, in order to satisfy Equation (2.41), λ_i should be positive. All parameter vectors $\mathbf{a} = \mu_m \cdot \mathbf{u_m}$ minimize Equation (2.31), if the pair $(\lambda_m, \mu_m \cdot \mathbf{u_m})$ satisfies the Equations (2.36) and (2.37), and $\lambda_m > 0$.

In study [26], it is proven that for a real and positive definite matrix $\mathbf{S}$, the generalized eigenvalues of the equation $\mathbf{S} \cdot \mathbf{u} = \lambda \cdot \mathbf{C} \cdot \mathbf{u}$ has the same sign with the eigenvalues of matrix $\mathbf{C}$.

Since the eigenvalues of $\mathbf{C}$ are $\{-2 \quad -1 \quad 2 \quad 0 \quad 0 \quad 0\}$ and scatter matrix $\mathbf{S}$ is positive definite, only one of the six generalized eigenvalues of Equation (2.36) is positive. Therefore, the solution of the minimization problem is $\mathbf{a} = \mu_m \cdot \mathbf{u_m}$ and it is unique. The pair $(\lambda_m, \mu_m \cdot \mathbf{u_m})$ satisfies Equations (2.36) and (2.37), where λ_m is the only positive generalized eigenvalue.

In this study, direct ellipse fitting method [26] is implemented to achieve real time processing speeds, since its computational load is much less than Bennett's approach. The experiments with synthetic images containing a single ellipse show that the ellipse fitting error defined as the sum of algebraic distances given by Equation (2.31) is proportional with number of data points N on a connected path,

which is equivalent to the number of edge pixel for these images. Thus the relative error criterion is defined as

$$E_{rel} = \frac{\|\mathbf{D} \cdot \mathbf{a}\|^2}{N} \qquad (2.42)$$

and it is used to compare various fitting alternatives and to choose the best ellipse fit among all connected paths in an image.

For the direct ellipse fitting method [26], resulting relative errors for elliptical and rectangular images in the synthetic image library are presented in Figure 2.25. As seen in the figure, the maximum relative error value for the synthetic elliptical images is smaller than the minimum relative error value of rectangular images. This implies that the threshold value 0.02 for relative error criterion should classify all ellipses and rectangles in the synthetic image library. Indeed, in the tests conducted with this threshold value, all ellipses and rectangles are classified correctly.
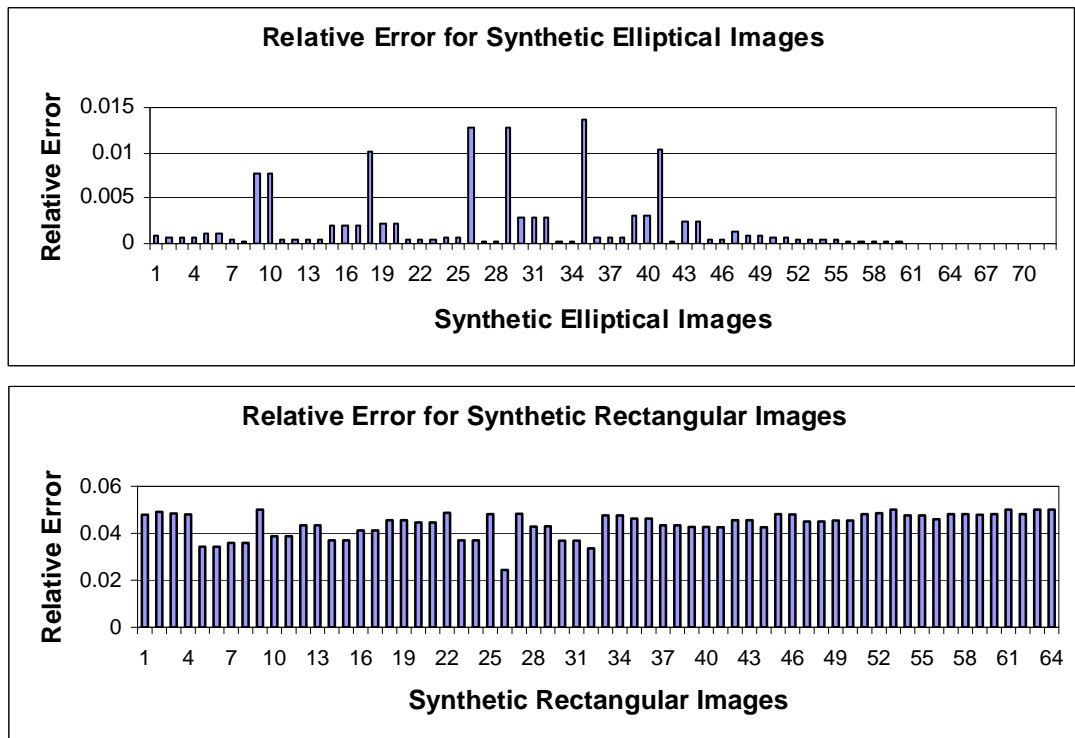
Figure 2.25. Relative Ellipse Fitting Errors for Synthetic Images.

The direct ellipse fitting method is able to detect the full ellipses (targets) in all eight sample digital images. One of the ellipse detect sample image is presented in Figure 2.26. The number at the ellipse center in the figure is the relative error multiplied by 1000. As opposed to Bennett's algorithm, it can also detect partial ellipses at the image boundaries, Figure 2.27, since the image boundaries are not evaluated as edges, so the connected path for the ellipse does not include a linear part as in case of occlusion by another object. However, in Figure 2.28, the connected path corresponding to target includes some linear parts causing huge relative fitting errors. The direct ellipse fitting method cannot detect the target in this image.
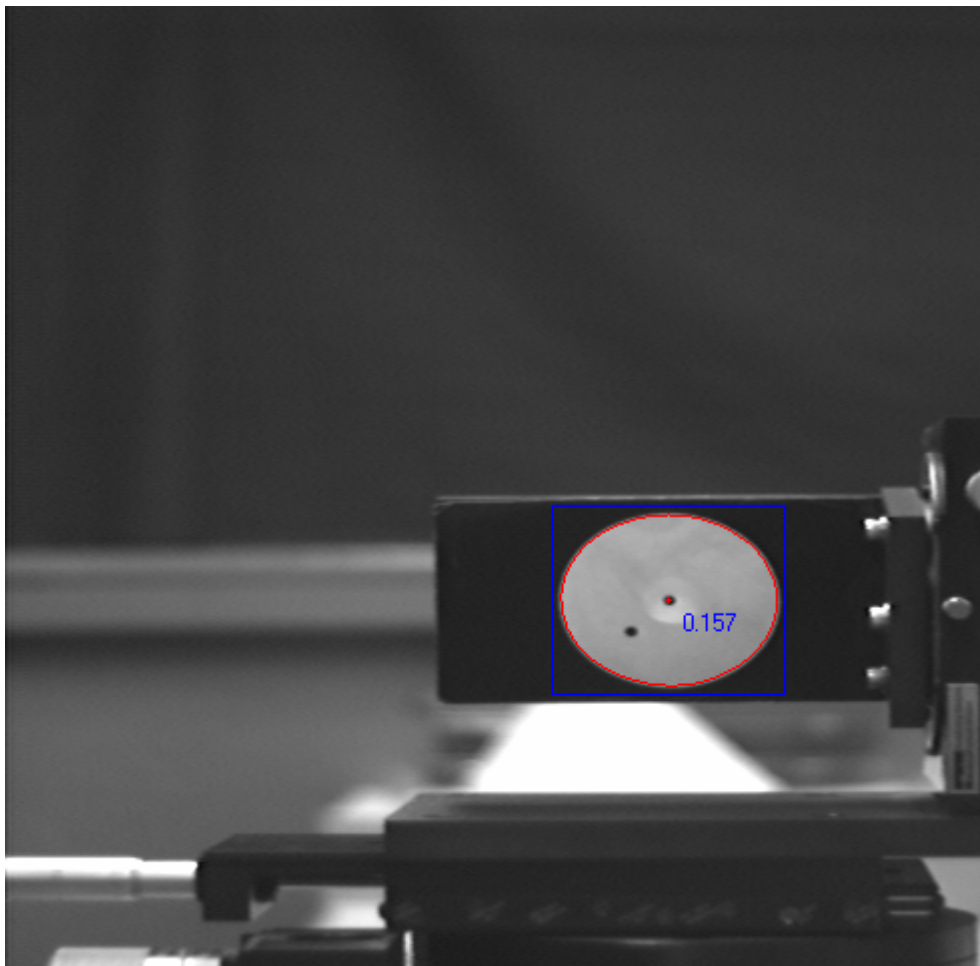
Figure 2.26. Detected Ellipse in image sample03.
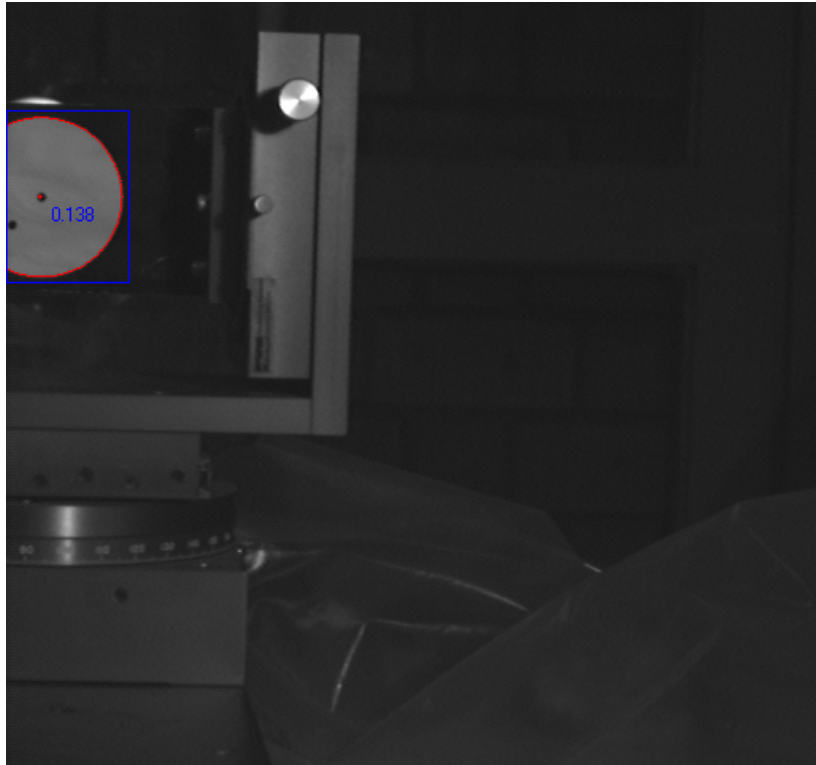
Figure 2.27. Detected Ellipse in Image sample10.

Figure 2.28. Edge Detection Results in Image sample08.

In Table 2.4, the processing times of direct ellipse fitting method is presented. As seen in the table, the maximum processing time is 47 ms and the average processing time is 25 ms. The speed of this method is 60 times faster than the Bennett's algorithm and suitable for real time target detection.

Table 2.5. Processing Times for Direct Ellipse Fitting Method

Image	Processing Time [ms]
sample01	47
sample02	46
sample03	31
sample04	8
sample05	16
sample06	31
sample07	31
sample08	15
sample09	15
sample10	16
Average	**25.6**

2.6.3. Comparison of Ellipse Detection Methods

The Bennett's algorithm [25] and its decision rule implemented in this study and the direct ellipse fitting method [26] are tested with synthetic images and real time frames to compare the detection performances of the methods for extreme cases. In Figure 2.29, the results of the Bennett's approach are presented. The numbers presented in the figure are accumulation score, quality factor and symmetry measure of the absolute peak of accumulation space, from top to bottom. The results show that Bennett's algorithm can not distinguish the ellipse if the shapes in Figure 2.29 appear in the same image. On the other hand, the resulting relative ellipse fitting errors, given in Figure 2.30 show that the direct ellipse fitting method is successful to detect the ellipse even in a worse case.
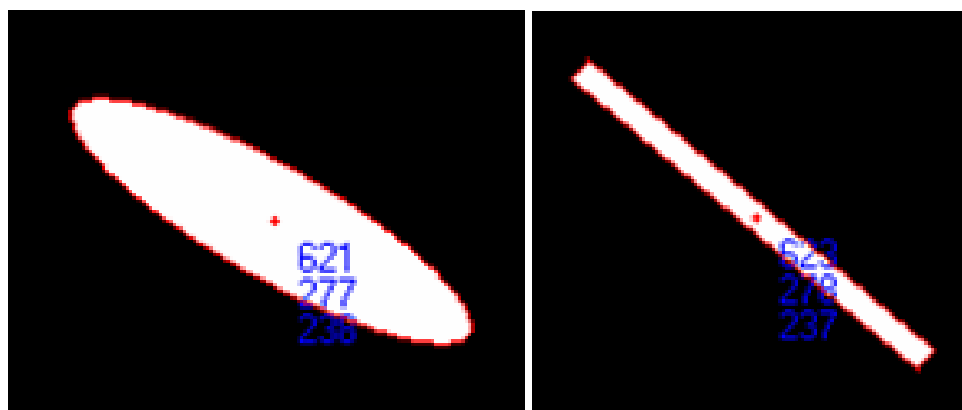
Figure 2.29. The Resulting Accumulation Scores, Quality factors and Symmetry Measurements for a Synthetic Ellipse and Rectangle.
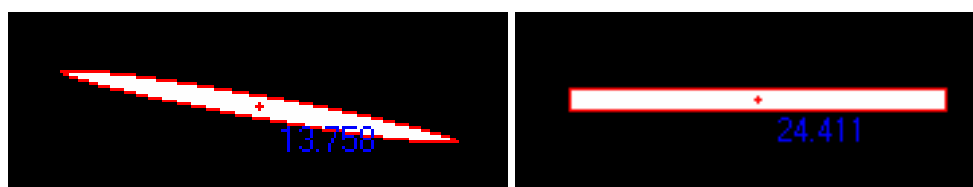
Figure 2.30. The Resulting Relative Ellipse Fitting Errors (multiplied by 1000) for a Synthetic Ellipse and Rectangle.

The detection results of the Bennett's algorithm given in Figure 2.31 and the direct ellipse fitting method given in Figure 2.32, for the same synthetic image containing an ellipse and a rectangle imply that Bennett's approach may fail to detect ellipses, if a high aspect ratio rectangular connected path appears in an image. On the other hand, this extreme case did not appear in experiments with real frames and both Bennett's algorithm and direct ellipse fitting method are successful to detect the target. Two sample ellipse detected frames are presented in Figure 2.33 and Figure 2.34.
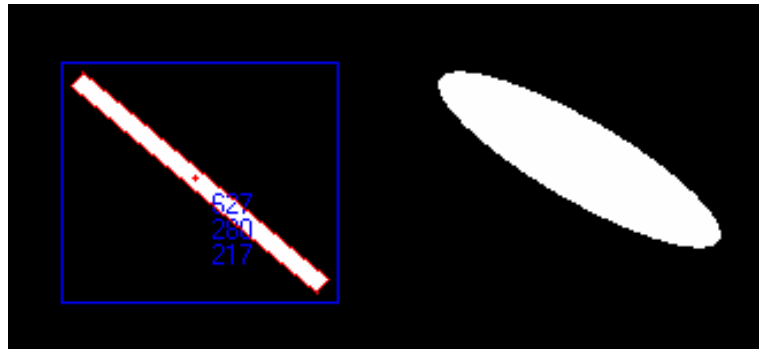
Figure 2.31. Ellipse Detected by Bennett's Approach.
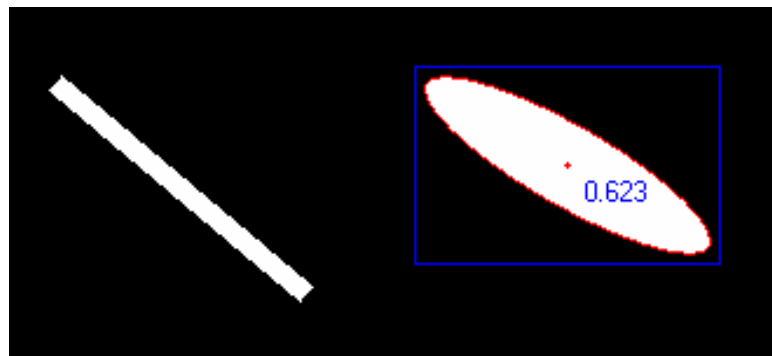
Figure 2.32. Ellipse Detected by Direct Ellipse Fitting Method.
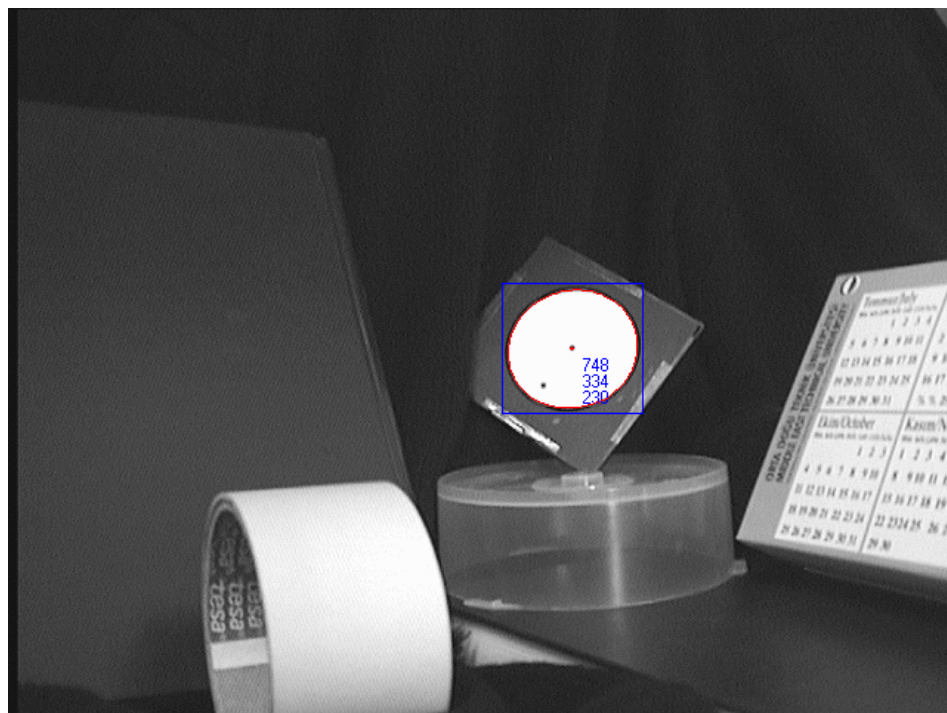
Figure 2.33. Ellipse Detected by Bennett's Approach.
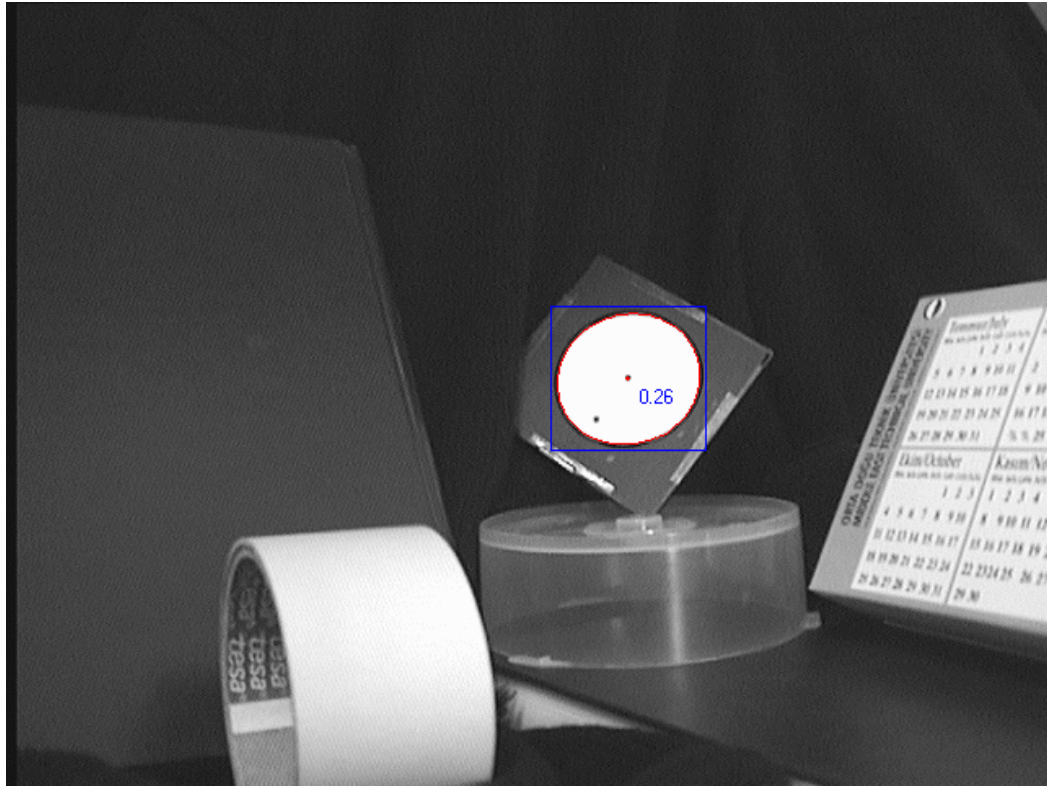
Figure 2.34. Ellipse Detected by Direct Ellipse Fitting Method.

The processing times of Bennett's algorithm and direct ellipse fitting method for sample real images are presented in Table 2.6. As seen from the table, the direct ellipse fitting method is approximately 60 times faster than the Bennett's algorithm. Since the computational order of Bennett's algorithm is n^2, its processing time is highly affected by number of pixels in a connected path. The processing times for Bennett's algorithm are proper for real time 3D reconstruction. Moreover, the results, presented in Table 2.6, show that the speed of direct ellipse fitting method is almost above the limit of 25 Hz, which is maximum frame rate of the grabber.

Table 2.6. Processing Times for Ellipse Detection Methods

Image	Processing Time [ms]	
	Bennett	**Ellipse Fitting**
sample01	1406	47
sample02	3156	46
sample03	1672	31
sample04	250	8
sample05	266	16
sample06	1375	31
sample07	907	31
sample08	2906	15
sample09	214	15
sample10	2578	16
Average	**1473**	**25.6**

2.7. Summary

In this Chapter, the processes starting from the image formation to the ellipse detection are presented, discussed, and the results of the developed and implemented methods in this study are compared with the results of the previous studies [1-4]. Although the overall process sequence is kept unchanged, all methods used in previous studies are investigated and optimized in terms of robustness and speed, and some new methods are added.

The virtual positioning and perspective projection, and digital image files are used to test the developed methods in previous studies [1-4]. But this study also enables a direct access to the frame buffer of the grabber and use real time frames to test the overall reconstruction method.

The previous studies [3, 4] used a minimum error thresholding method to determine the threshold value, which was then used to binarize the grayscale images. In this study, the experiments with digital image files and real time grayscale frames show that the threshold value determined with this method is robust to segment the

foreground objects from the background. However the previous implementation of the method contained some repetitive calculations and therefore was not optimum when its processing speed is considered. The equations and the algorithm are rearranged to avoid these repetitions and the speed of the method is enhanced by a factor of 45.

In previous studies [3-4], edge gradient vectors in a binary image were determined by convolving the image with horizontal and vertical Sobel filters. The pixels with non-zero gradient magnitude were marked as edge pixels. The problem was that the resulting edge contours were thicker than one pixel, and an additional edge thinning operation was needed to solve this problem. As a result of this edge thinning operation, the edge contours were deformed, so the ellipse detection process was affected adversely. In this study, a modified version of Laplacian kernel is implemented as an edge detector and Sobel filters are kept for gradient vector calculation for only edge pixels detected by the Laplacian edge detector. The edge thinning operation is avoided since the resulting edge contours have only one pixel thickness. The results of the experiments conducted with this new algorithm show that the obtained edge contours are not deformed, no extra edges are created, and the new algorithm works 13 times faster than its predecessor.

The connectivity algorithm designed in previous studies [3, 4] did not work properly because of the implementation errors. Some edge pixels were appearing in more than one connected path. In this study, the implementation errors of the previously designed method are corrected and the desired connectivity results are obtained. However the speed of the algorithm is above the limits of real time working and it can not be enhanced, since its computational order is n^2. A new connectivity approach, inspired from the conductivity concept is developed. The new connectivity algorithm results in exactly the same connected paths but it is 20 times faster.

To detect ellipses, the Bennett's approach [25] was implemented in previous studies. Quality factor and symmetry measurement were designed to evaluate the

resulting accumulation space and to decide on the ellipse center. The problem in the implementation was that although the edge pixels are grouped into connected paths, their accumulation scores were collected in a single accumulation space. The accumulation scores for a path adversely affected the quality factors and symmetry measurement of a peak corresponding to another path. In this study, the results of the experiments conducted with synthetic images containing a single ellipse show that, (i) the absolute maximum point of the accumulation space coincides with the exact ellipse center, (ii) previously designed quality factor, which is used to calculate the sharpness of the peak of accumulation space, is unexpectedly changing with ellipse size, and (iii) the symmetry of the peak does not vanish till to the closest image border. The Bennett's algorithm is re-implemented to use distinct accumulation spaces for each connected path, and the quality factor is redesigned to give almost constant values for different sized ellipses. For each connected path, the accumulation score, the quality factor and symmetry measurement are calculated for the absolute maxima of accumulation space. The decision rule is redesigned to minimize the classification errors in the tests conducted with synthetic image library consist of elliptical and rectangular images. Although the detection rate is enhanced, a decision boundary that correctly identifies all the ellipses and rectangles does not exist. Some rectangles are falsely detected as ellipses and some ellipses are classified as non-ellipse. In addition to detection problem the speed of the algorithm does not allow real time working.

The direct ellipse fitting method [26] is discussed and implemented as an alternative ellipse detection method to Bennett's algorithm. The average sum of algebraic distance of the points to the ellipse equation determined by fitting is defined as the relative error criterion. The results of the experiments for elliptical and rectangular images in the synthetic image library show that the direct ellipse fitting method is capable of detecting all ellipses in the library while eliminating all rectangular images. Moreover, the direct ellipse fitting method is 60 times faster than the Bennett's algorithm and the speed is just above the limit of 25 Hz, which is the maximum frame rate used by the frame grabber device.

60

The image processing algorithms discussed in this chapter are combined to form the overall target detection method and tested with 640x480 pixel real time frames. The target detection method, which uses the slowest combination of the methods, has a speed of 0.5 Hz to 5 Hz, changing with the complexity of the scene. On the other hand, the combination of fastest methods has as a detection speed between 10 Hz and 25 Hz, changing with complexity of the scene and limited by the frame grabber's maximum sampling rate.

CHAPTER 3

TARGET TRACKING

3.1 Overview

In this chapter, the aim of the target tracking method and the proposed method for target tracking will be explained. The limitations and experimental results for the target tracking method will be presented and discussed.

Although the working speed of the target detection method proposed in Chapter 2 reaches to 25 Hz for some of the 640x480 real time frames, it is not a fixed value. The complexity of the scene, in other words, the number of edge pixels and number of connected paths in the edge detected frame determines the target detection speed.

One of the solutions to fix the target detection speed at the maximum frame grabbing speed, 25 Hz, can be upgrading the computer hardware. However, any upgrade that increases the resolution of the frames or frame grabbing speed may not permit the working at the frame grabbing speed.

In this study, a target tracking method is designed and used to overcome this problem. The proposed target tracking method is based on estimating the target region for a frame using the data obtained from the past frames. Processing a smaller region in a frame that includes the target will reduce the complexity of the frame computations and will eventually increase the target detection speed.

Kalman filtering is the most commonly used target tracking method. It is known with its high capability on noise rejection and ability to adapt itself according to estimation errors, but requires a good modeling of the dynamics of the observed system. Another method commonly used in target tracking studies is the polynomial

fit to the measured data and extrapolation. Generally, it is preferred because of its low computational load.

3.2. Tracking Method

The target tracking method used in this study first estimates the coordinates of the smallest rectangular region that includes the target in the frame using the position information of the target in previous frames. Only this portion of the frame will be searched for the target with the proposed target detection method. Thus, the estimation accuracy is not crucial as long as the target appears inside the estimated region. Considering these facts, a polynomial fitting method is preferred, because its computational load is low and it does not require any priori information about the target dynamics.

The idea of using a secondary passive target to reconstruct the 3D parameters of the camera or the target implies a low possibility of having an agile target motion. However, in the real time frames, the target can start moving, change direction, or stop moving at any time, meaning that the characteristics of the motion may change. If the information on past data points are still kept and used to estimate the position of the target after a change in motion characteristics, the estimation errors will increase. Thus, the tracking method should track estimation and polynomial fitting errors and change the number of data points used and polynomial degree when necessary.

The proposed target tracking method in this study is an extrapolation by polynomial fitting, with changing polynomial degree and number of data points.

3.2.1. Least Square Polynomial Fitting

The least square polynomial fitting is well known problem and the polynomial of degree **m** can be represented as

$$P_m(t) = c_0 + c_1 t + c_2 t^2 + \ldots + c_m t^m = \mathbf{c}^T \cdot \mathbf{t} \qquad (3.1)$$

63

where $P_m(t)$ is the mth degree polynomial of t, $\mathbf{c} = \begin{bmatrix} c_0 & c_1 & \cdots & c_m \end{bmatrix}^T$, and $\mathbf{t} = \begin{bmatrix} 1 & t & t^2 & t^m \end{bmatrix}^T$ with t being the time.

If (t_i, x_i) are the ith data points among N measurements and if a polynomial of degree m is fitted to the last n data points, then the least square polynomial fitting can be reduced to the solution of the following system of equations

$$\mathbf{A} \cdot \mathbf{c} = \mathbf{x} \tag{3.2}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & t_N & t_N^2 & t_N^m \\ 1 & t_{N-1} & t_{N-1}^2 & t_{N-1}^m \\ \vdots & & & \\ 1 & t_{N-n+2} & t_{N-n+2}^2 & t_{N-n+2}^m \\ 1 & t_{N-n+1} & t_{N-n+1}^2 & t_{N-n+1}^m \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \\ c_m \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_N \\ x_{N-1} \\ \\ x_{N-n+2} \\ x_{N-n+1} \end{bmatrix}$$

Then the coefficient vector $\mathbf{c}$ is obtained by pseudo-inversion.

$$\mathbf{c} = \left(\mathbf{A}^T \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{x} \tag{3.3}$$

For the next time step t_{N+1}, the estimated value of the data $\tilde{x}_{N+1}$, can be evaluated as

$$\tilde{x}_{N+1} = P_m(t_{N+1}) = c_0 + c_1 t_{N+1} + c_2 t_{N+1}^2 + \ldots + c_m t_{N+1}^m \tag{3.4}$$

The correlation coefficient of the fit, $\mathbf{r}$ can be expressed as

$$r = \sqrt{r^2} \tag{3.5}$$

where r^2 is called coefficient of determination and can be expressed as

$$r^2 = \frac{\sum_{N-n+1}^{N} (x_i - \bar{x})^2 - \sum_{N-n+1}^{N} (x_i - P_m(t_i))^2}{\sum_{N-n+1}^{N} (x_i - \bar{x})^2} \tag{3.6}$$

and $\bar{x}$ is the mean of the last n data $\begin{bmatrix} x_{N-n+1} & x_{N-n+2} & \cdots & x_{N-1} & x_N \end{bmatrix}$ and it can be evaluated as

$$\bar{x} = \frac{\sum_{N-n+1}^{N} x_i}{n} \tag{3.7}$$

After the measurement, the estimation error, e can be evaluated as.

$$e = \left| x_{N+1} - \tilde{x}_{N+1} \right| \tag{3.8}$$

3.2.2 Tracked Points

The aim of introducing the target tracking method to this study is reducing the target detection load by reducing the area in which the target searched. Therefore, target tracking method should estimate the target region for the subsequent frame to be processed by utilizing the data obtained for past frames. However, it is possible to track either 2D image coordinates of the target obtained from the target detection method or 3D parameters of the target obtained from 3D reconstruction method. Both of these approaches are implemented.

3.2.2.1 Tracking 2D Corners

In this approach, it is aimed to estimate the target region by tracking the 2D image coordinates of the elliptical target detected in the previous frames. However it is

practical to track image coordinates of the smallest rectangle that fully includes the elliptical image of the target, since the goal of the target tracking algorithm is to estimate the coordinates of this rectangle for the subsequent frames. The rectangle that fully includes the target as shown in Figure 3.1, can be represented with the horizontal and vertical image coordinates of upper left corner, **A**, and bottom right corner, **B**. The 2D image coordinates of these points are stored and four distinct polynomials of desired degree are fitted to the desired number of newest values of these image coordinates. Then image coordinates of these points are estimated for the time which the subsequent frame is grabbed.
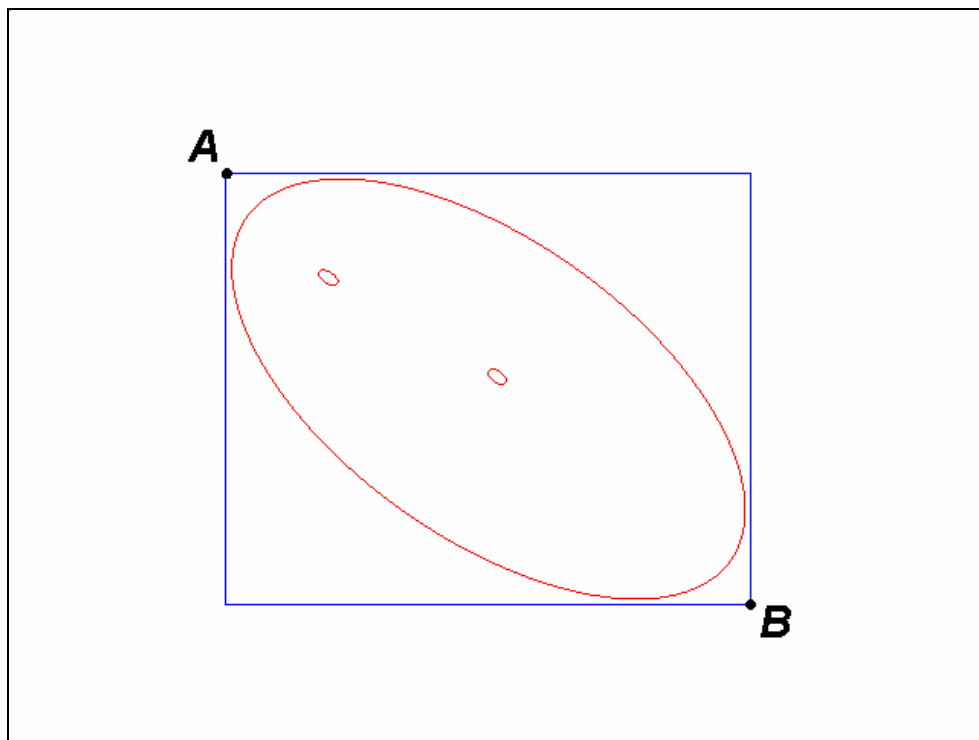
Figure 3.1. Tracking 2D Corners

3.2.2.2 Tracking 3D Parameters

An alternative approach for estimating the target region is using 3D configuration parameters of the target calculated by 3D reconstruction method for the previous frames. In this approach, the 3D parameters, as 3 rotations and 3 translations, of the target are stored and six distinct polynomials are fitted to estimate the values of

these parameters for the time at which the subsequent frame is grabbed. Once the estimated values of the 3 translations and 3 rotations of the target with respect to camera are obtained, the image coordinates of the corners (point **A** and **B** shown in Figure 3.1) of the enclosing rectangle are calculated with virtual positioning and perspective projection.

3.3 Tracking Strategy and Parameter Space.

The proposed target tracking method in this study is an extrapolation by polynomial fitting, with changing polynomial degree and number of data points. The positions of the corners of the smallest rectangular region includes the full target (Figure 3.1) in the subsequent frame are estimated by polynomial fitting to either 2D image coordinates of the target or to 3D target parameters obtained for previous frames. Change in motion characteristics is checked by evaluating the correlation coefficient of the fit and the degree of the polynomial and used number newest data points are reduced accordingly. Since the estimation accuracy is not crucial, as long as the target appears inside the estimated region, it is possible to enlarge the estimated region to guarantee the target is inside the region. Also fast shrinkage of the estimated region should be damped to avoid partially outside target situations which may appear due to estimation errors caused by measurement noise.

As stated before the polynomial fitting method is a low computational load and easy to implement method. However, the above discussed steps of the method offers a multi dimensional parameter space, during adaptation of the method to a case as in this study and reveals the following questions about selection of strategy and parameters:

- Which is the best tracking approach? 2D image coordinates or 3D target parameters?
- What is the best degree of the polynomial?
- What is the best number of newest data points used for polynomial fitting?
- What is the best the critical correlation coefficient limit below which the fit is accepted as unsatisfactory?

- What is the best amount of estimated region enlargement to guarantee the target is inside?

The answers of those questions are not independent of each other, and the "best" concept is subjective. But it is possible to define an objective function that measures the performance of a target tracking strategy and the set of parameters that gives the maximum performance may be searched. However defining an objective function for evaluating the performance of a tracking strategy is also subjective. Moreover, it is obvious that conducting experiments to obtain best set of parameters is an exhaustive search, if the dimension of the parameter space is taken into account.

This study does not aim to obtain and use the set of parameters that reveals the best tracking strategy; it only aims to find a set of parameters that satisfies the goal of the target tracking method, which is reduction of the search area of target detection method. Thus a stepwise parameter determination approach is followed to avoid an exhaustive search.

3.3.1 Critical Correlation Coefficient

The target tracking algorithm is designed to distinguish the changes in the motion characteristics. The correlation coefficient of the fit is calculated and checked for each time step, just after the estimation but before the measurement. If the correlation coefficient is below a critical limit, the number of last data points used for fitting is reduced to two, and then the position is re-estimated with the first degree polynomial fitted to the last two data points. After this action, the number of data points used and degree of polynomial is increased to the desired values one by one as the new data is obtained.

Two experiments are conducted to determine the critical correlation coefficient value below which the characteristic of the motion is accepted as changed. In the first experiment, the camera is rotated about y-axis with a constant rotational speed and the target is stationary. The motion of the target is estimated by fitting a third

degree polynomial to the newest 10 data obtained for previous frames. The correlation coefficients calculated for this experiment are presented Figure 3.2. In the second experiment, a small amplitude swinging motion of the target is tracked. Third degree of polynomial is fitted to the latest 10 data obtained for the previous frames. The resulting correlation coefficients are presented in Figure 3.3. The minimum correlation coefficient value occurred in these two experiments is selected as the critical value, since the characteristics of the motion of the target for both cases are assumed to be constant during the experiments.
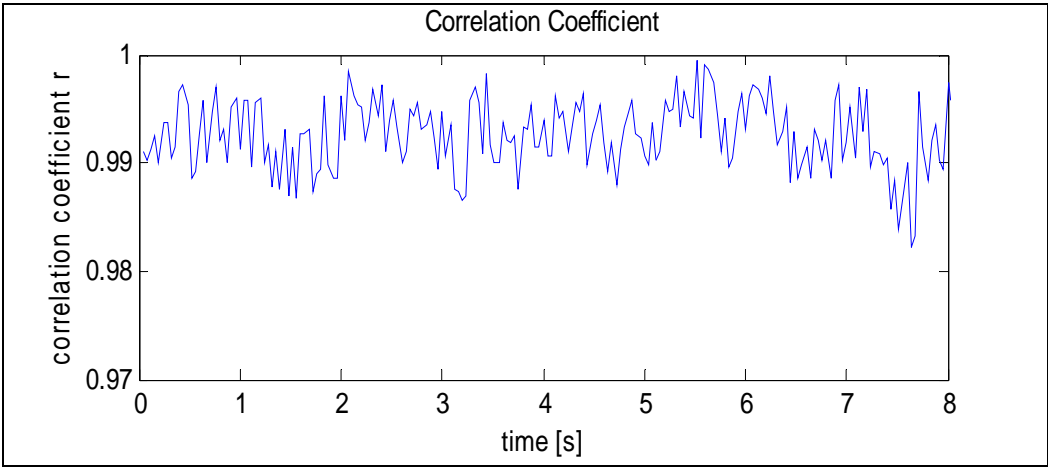
Figure 3.2. Correlation Coefficient Values for Fixed Target, Camera Rotation Case
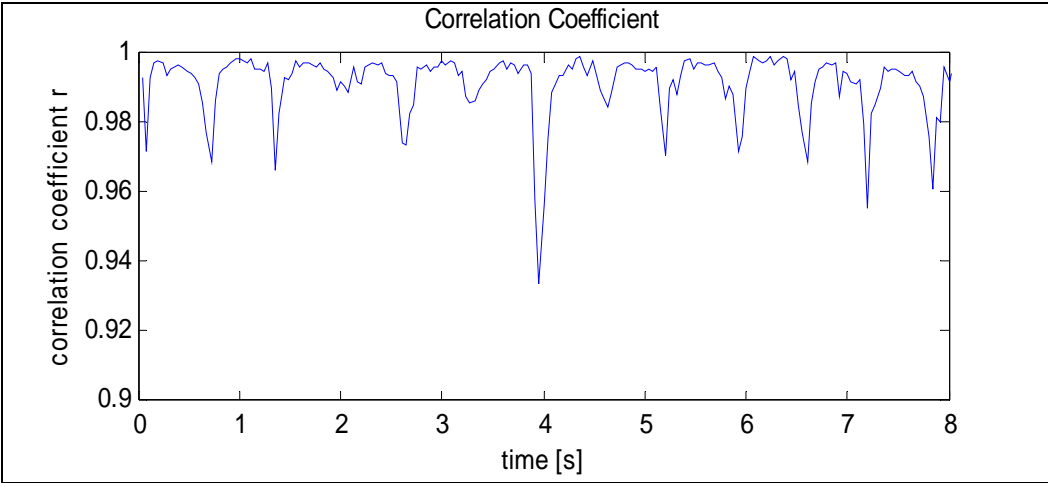
Figure 3.3. Correlation Coefficient Values for Swinging Motion of the Target

3.3.2. Number of Data Points and Degree of Polynomial

In the least square polynomial fitting, increasing the number of data points used reduces the effects of noise, as long as the characteristic of the motion of the target is not changed. However, if this characteristic changes drastically, using the data obtained from the previous frames of the motion obviously increases the estimation error and consequently the target may appear outside the estimated region.

Theoretically, it is possible to use a fixed higher degree polynomial even for a data that can be represented by a lower degree polynomial. But, the fitting a higher degree of polynomial to a noisy data may exaggerate estimation errors.

In order to test the target tracking algorithm and to determine the best choice for the degree of polynomial and number of last data points to be used for fitting, three test case scenarios are crated, two in virtual environment and a third one in real world. The first one is a sinusoidal motion of the target in x-direction crated by virtual positioning and perspective projection (Figure 3.4). The second one is a saw-tooth motion of the target in x-direction, also in virtual environment. The third case is the motion of a secondary target attached pendulum in real world, Figure 3.5.
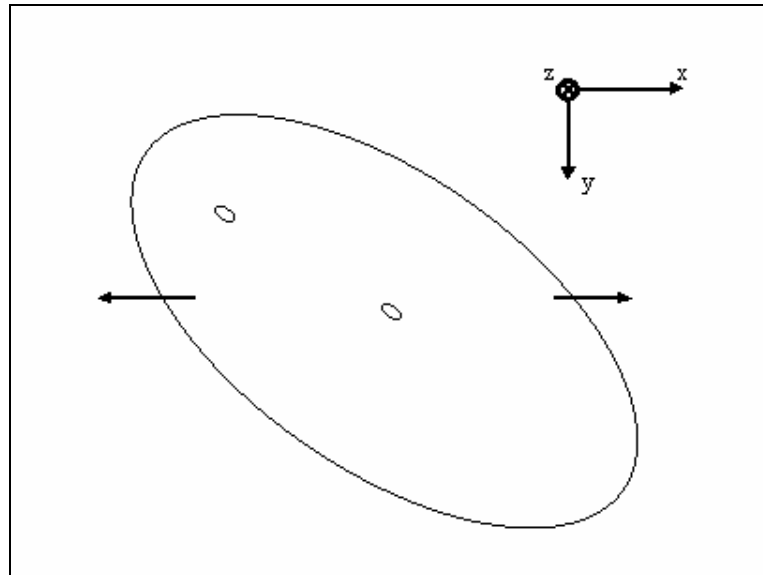
Figure 3.4. Direction of Motion of the Target Created by Virtual Perspective Projection

70
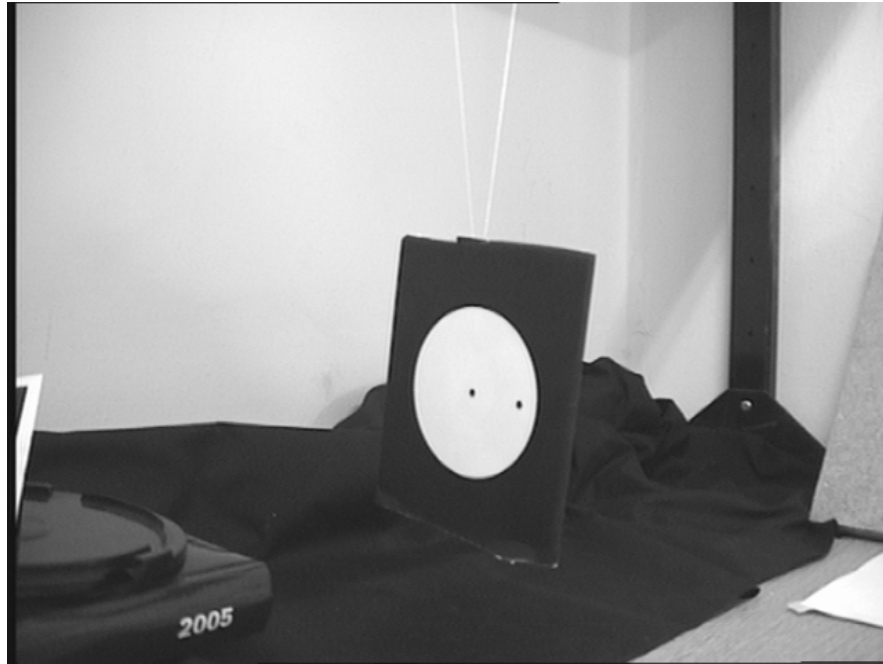
Figure 3.5. Secondary Target Attached Pendulum.

Several experiments are conducted to determine the best choice for the degree of polynomial and number of last data points to be used for fitting. The target tracking method is tested for third degree polynomials, the highest, and last 10 and last 5 data points, to track a moving target in synthetically generated frames and in real frames, as well. The critical correlation coefficient discussed and determined in the previous section is not altered in the experiments.

The results of the 2D and 3D target tracking approaches for three cases, sinusoidal and saw-tooth motion in virtual environment and swing motion of pendulum in real world, using last 5 and last 10 data points and zero, first, second and third degree polynomial fitting are presented in Table 3.1, 3.2 and 3.3. Investigation of maximum estimation errors, average absolute estimation errors and standard deviations of the absolute errors of tracking preferences presented in these tables reveals that the 2D tracking method with second degree polynomial fitted to the last 5 data gives the best average performance for three cases. On the other hand, 2D tracking with third degree polynomial fitted to the last 5 data exhibits the best performance for real data, pendulum case.

71

Table 3.1. Maximum Estimation Errors of Target Tracking Strategies

Tracking	2D								3D							
# of points	5				10				5				10			
Pol. Degree	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Sinusoidal	31	3	1	11	54	11	2	11	35	2	5	3	55	9	3	5
Saw Tooth	20	14	8	39	37	23	12	14	22	12	10	17	38	21	12	20
Pendulum	23	11	11	8	31	22	11	20	26	11	10	9	33	18	11	40
Average	24.7	9.3	6.7	19.3	40.7	18.7	8.3	15.0	27.7	8.3	8.3	9.7	42.0	16.0	8.7	21.7

Table 3.2. Average Absolute Estimation Errors of Target Tracking Strategies

Tracking	2D								3D							
# of points	5				10				5				10			
Pol. Degree	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Sinusoidal	15.8	1.6	0.4	2.4	34.6	4.6	0.8	0.8	21.8	0.9	1.5	1.4	30.1	4.2	1.3	1.6
Saw Tooth	15.8	0.8	1.1	4.4	30.0	1.8	1.6	2.9	18.3	1.7	2.2	2.0	30.7	1.7	3.7	3.0
Pendulum	12.3	4.8	2.6	1.9	18.0	7.4	5.3	5.3	11.5	4.3	3.2	3.5	18.2	5.5	5.1	4.8
Average	14.6	2.4	1.4	2.9	27.5	4.6	2.6	3.0	17.2	2.3	2.3	2.3	26.3	3.8	3.4	3.1

Table 3.3. Standard Deviation of Absolute Estimation Errors of Target Tracking Strategies

Tracking	2D								3D							
# of points	5				10				5				10			
Pol. Degree	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Sinusoidal	9.3	1.0	0.5	2.3	16.0	2.9	0.6	1.7	8.5	0.6	1.0	0.9	16.4	2.7	0.9	1.0
Saw Tooth	4.0	2.7	1.8	7.8	7.5	5.2	3.0	4.3	3.4	2.3	2.4	2.9	8.3	3.5	3.7	3.8
Pendulum	6.8	3.0	2.3	1.8	8.6	6.7	2.9	4.5	7.7	2.8	2.3	2.3	9.2	4.7	3.0	5.6
Average	6.7	2.2	1.5	4.0	10.7	4.9	2.2	3.5	6.5	1.9	1.9	2.0	11.3	3.6	2.5	3.5

3.3.3 Enlargement of Estimated Region

As stated before the accuracy of the estimation process is not so critical as long as the target is inside the estimated rectangle. Besides, the speed of the target detection method is almost equal to the maximum frame rate of the frame grabber even without the target tracking method. Thus, enlarging the estimated region determined by the target tracking method will increase the possibility of having the target fully

inside the estimated region in case of estimation errors. Moreover, the processing speed will not be affected as long as the size of the estimated region does not approach to the size of a full frame.

Three estimation region enlargement (Figure 3.6) approaches are utilized in this study:

1. The width and height of the estimated rectangular region are increased by a fixed amount of pixels.
2. The estimated rectangular region is enlarged by a fixed scale factor.
3. The estimated rectangular region is enlarged by a scale factor proportional with the magnitude of the 2D velocity of the target center.
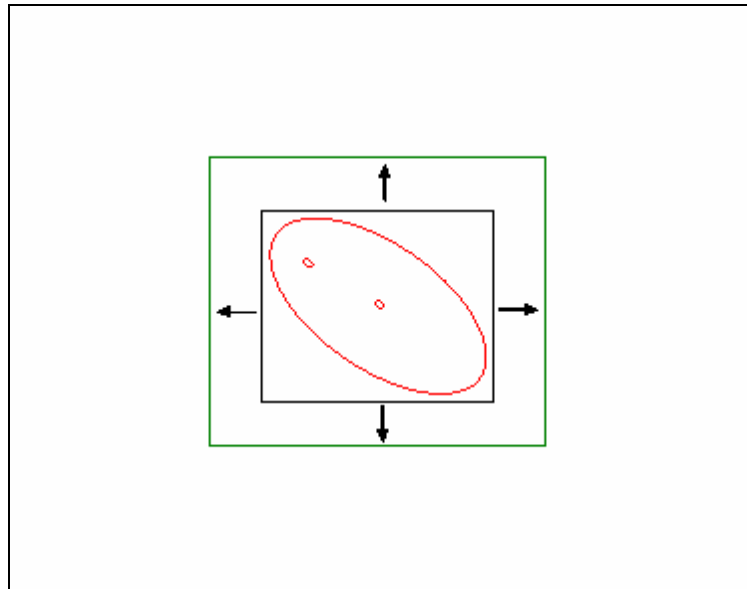
Figure 3.6. Estimated Region Enlargement.

These enlargements aim to reduce the possibility of having the target located partially or fully outside of the estimated rectangle for the subsequent frame in the sequence. The amount of enlargement factors are determined by investigating the maximum absolute estimation error presented in Table 3.1. If the estimated target region is totally enlarged, as summation of three approaches, more than twice of the maximum absolute estimation error, it is guaranteed to have the target inside the estimated region.

3.4. Results of Target Tracking

The performance of the target tracking method is simulated for the three cases, sinusoidal and saw-tooth motion of the target in x-direction created by virtual perspective projection, and swinging motion of the secondary passive target attached pendulum, using the best tracking parameters discussed in previous sections. Note that it is not claimed that the selected set of parameters are optimum by means of estimation error or any other criteria, however the parameters, critical correlation coefficient, degree of polynomial, the number newest data points used, and the enlargement factors are selected to meet the goals of the target tracking method.

For sinusoidal motion of the secondary passive target in x-direction in virtual environment case, tracking 2D corners approach using a second degree polynomial and last 5 data points exhibits the best performance if the maximum absolute error, mean absolute error and standard deviation of absolute error presented in Tables 3.1, 3.2, and 3.3 are considered. However, tracking 3D parameters approach with first degree polynomial fitted to the last 5 data shows a similar performance. In Figures 3.7 and 3.8, the results of the target tracking method for these strategies are presented.
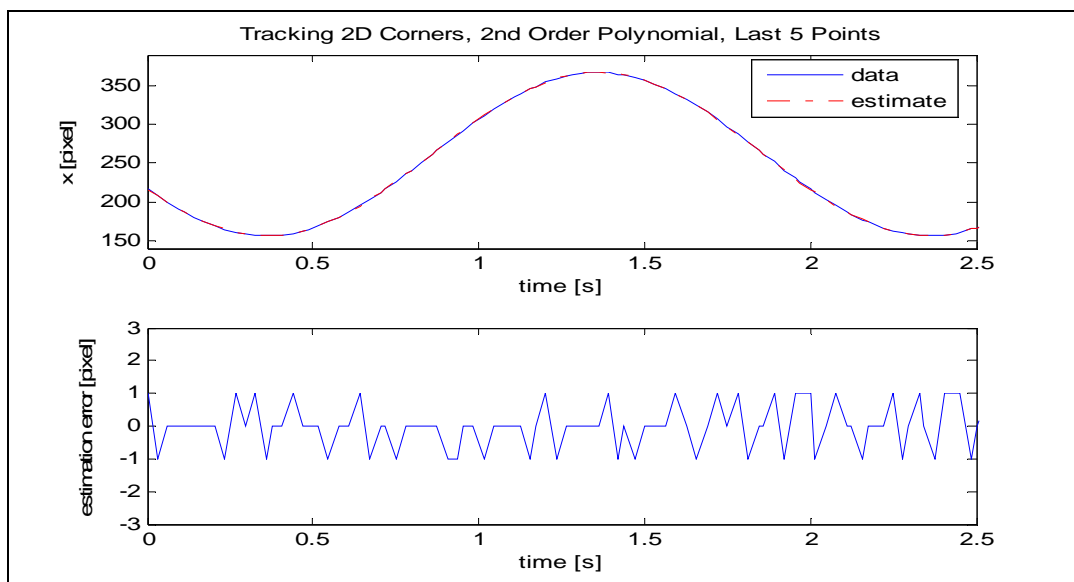
Figure 3.7. Sinusoidal Motion, Tracking 2D Corners.

74
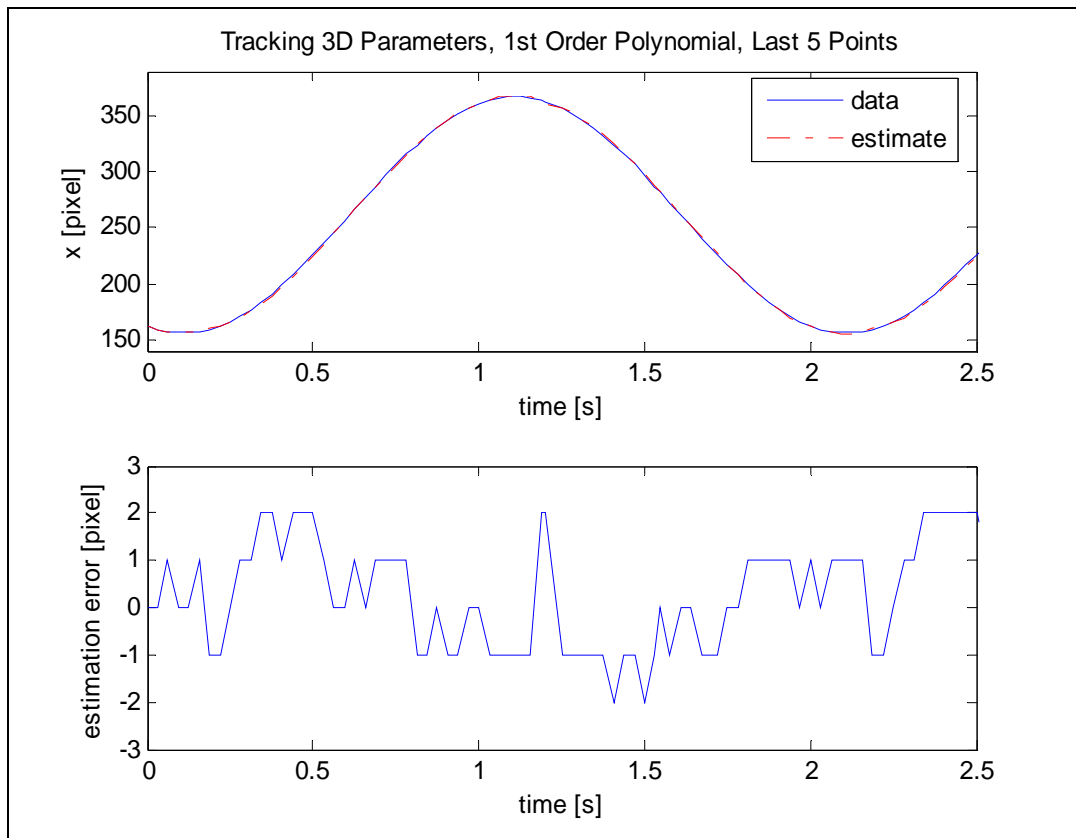
Figure 3.8. Sinusoidal Motion, Tracking 3D Parameters.

For saw-tooth motion of the secondary passive target in x-direction in virtual environment case, tracking 2D corners approach using a second degree polynomial and last 5 data points exhibits the best performance if the maximum absolute error, mean absolute error, and standard deviation of absolute error are considered (Figure 3.9). However tracking 3D parameters approach with second degree polynomial fitted to the last 5 data shows a similar performance (Figure 3.10) if the maximum absolute estimation errors are compared.
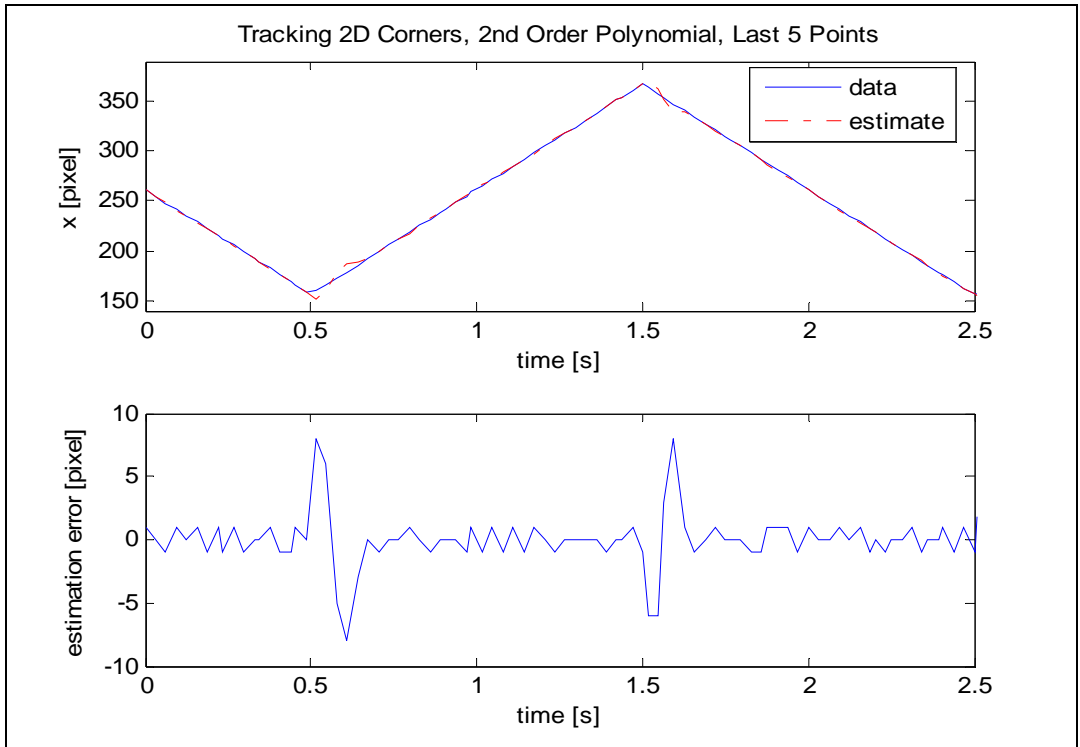
Figure 3.9. Saw-tooth Motion, Tracking 2D Corners.
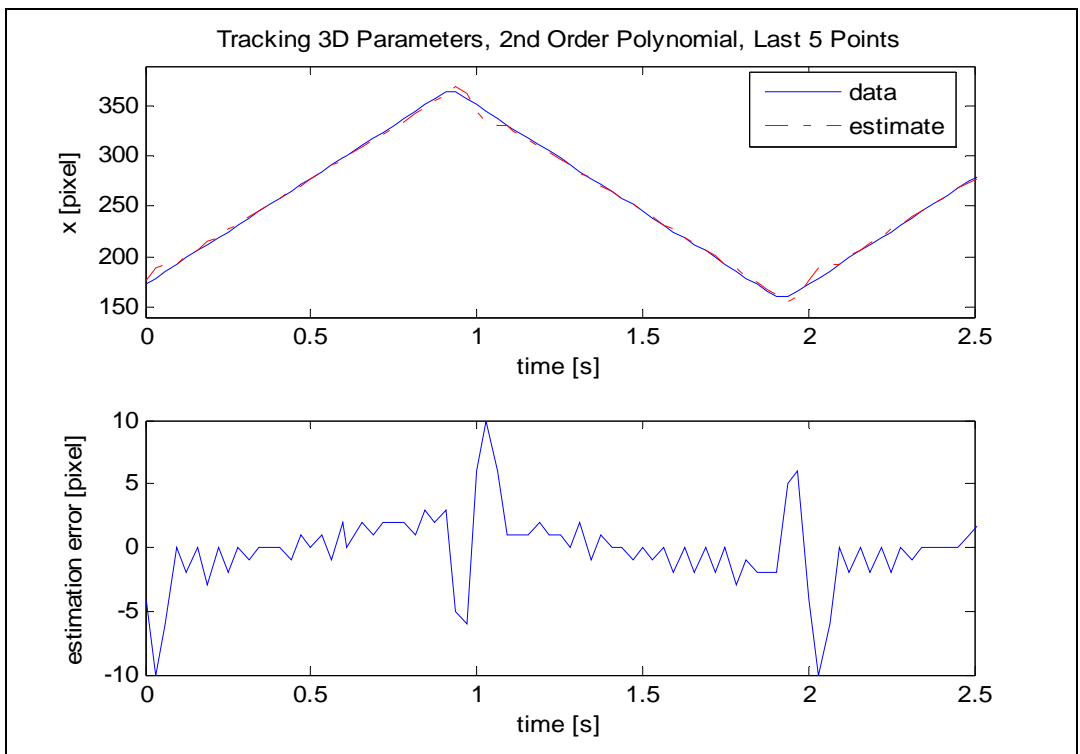
Figure 3.10. Saw-tooth Motion, Tracking 3D Parameters.

For swinging motion of the secondary passive target attached pendulum in real world case, tracking 2D corners approach using a third degree polynomial and last 5 data points exhibits the best performance if the maximum absolute error, mean absolute error, and standard deviation of absolute error are considered. However, tracking 3D parameters approach with third degree polynomial fitted to the last 5 data shows a similar performance.
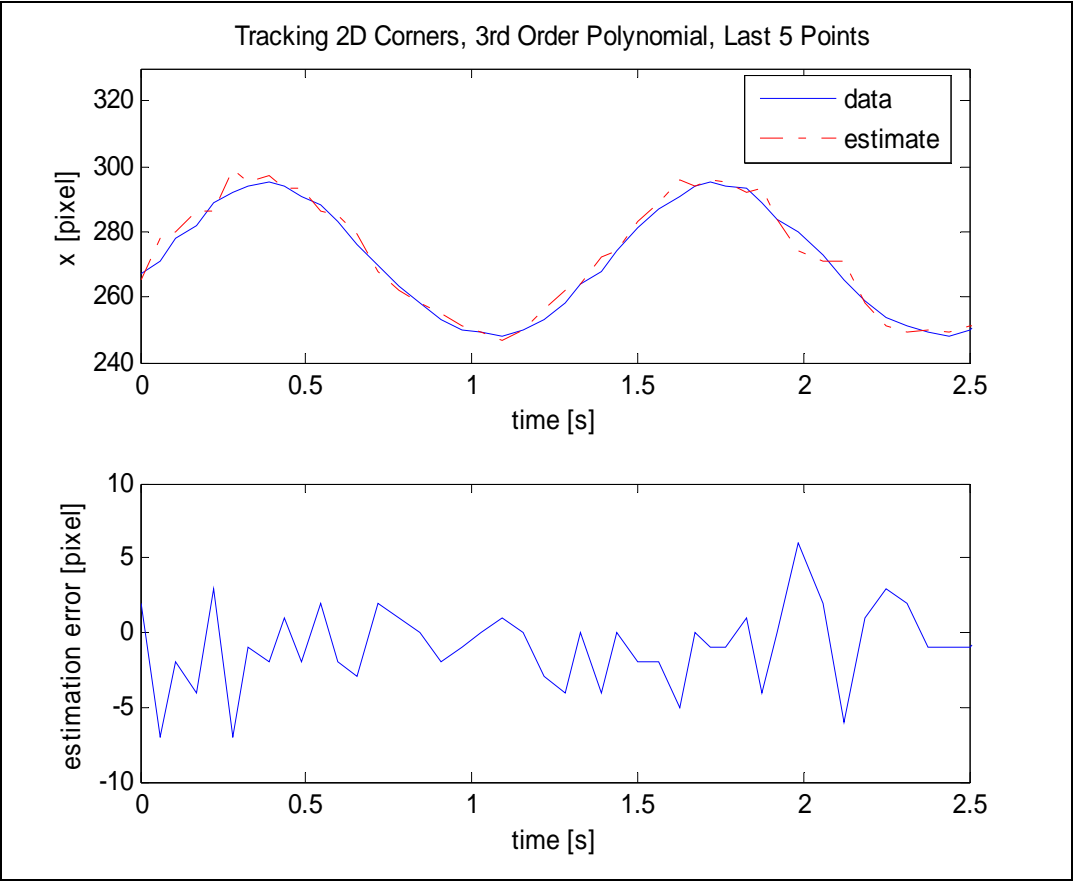
Figure 3.11. Swinging Pendulum, Tracking 2D Corners.
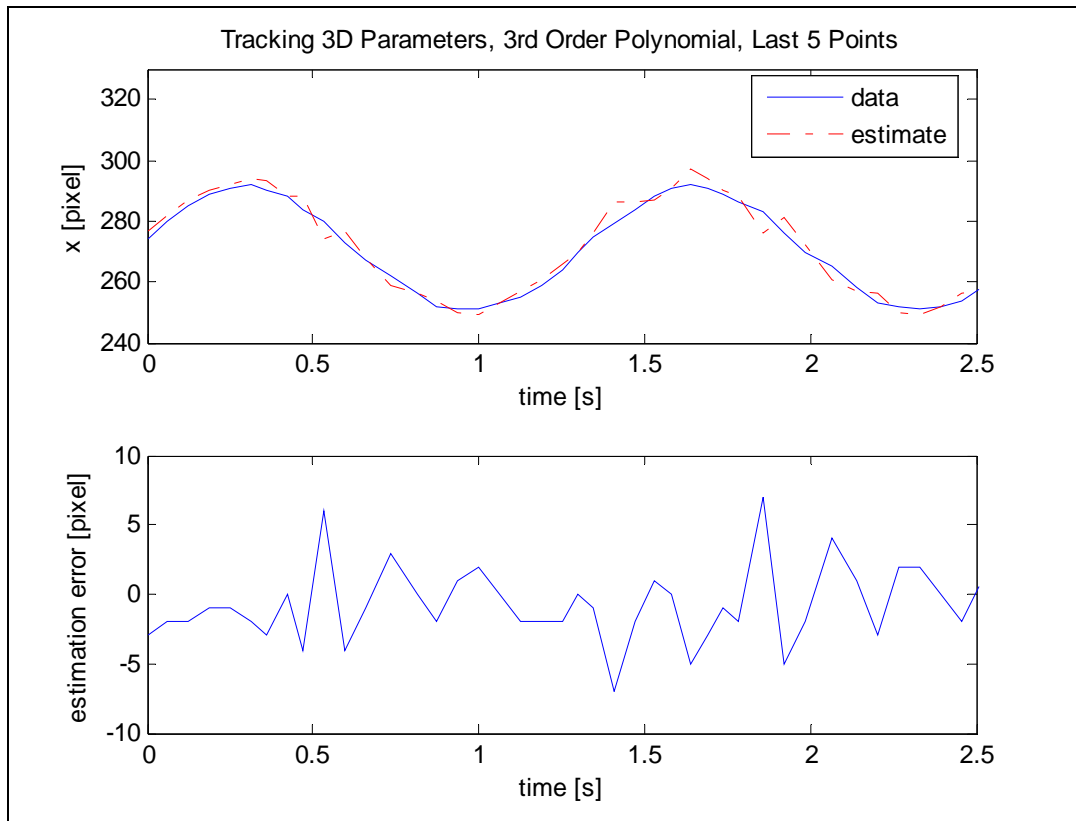
Figure 3.12. Swinging Pendulum, Tracking 3D Parameters.

The results of the experiments presented in Figures 3.7, 3.8, 3.9, 3.10, 3.11, and 3.12 show that the maximum estimation error of the tracking method does not exceed 10 pixels. This error value provides using small amount of enlargement factors, compared to size of the full image. Thus the search area of the target detection method is reduced drastically. One of the sample frames captured while tracking the swinging pendulum with 2D corners approach using 3rd degree polynomial fitted to the last 5 data is shown in Figure 3.13. The green rectangle in this figure is the enlarged estimated target region in which the target is searched. And the blue box shows the smallest rectangular region including the detected target.

78

Figure 3.13. A Sample Frame Captured During Target Tracking.

The major aim of introducing the target detection to this study is increasing the overall process speed to maximum sampling rate of the frame grabber, 25 Hz and fix there. In the experiments, it is observed that the CPU utilization reduces down to 50 %, while processing speed is constant at 25 Hz. This fact and the processing speed, 50 Hz, achieved for synthetic frames imply that the overall method, composed of target detection, 3D reconstruction and target tracking methods is capable of working at higher speeds.

CHAPTER 4

INTERNAL CAMERA CALIBRATION

4.1 Overview

This Chapter investigates those internal camera parameters that determine the image formation. Internal camera parameters are defined, distortion types and their causes are described, and distortion model and the proposed calibration procedures are explained in this Chapter.

Internal camera parameters, also called as "intrinsic parameters" of a camera, can be defined as the internal geometry of the camera that determines the image formation rules. These parameters are determined by the material properties, dimensions and orientations of the lens and image sensor with respect to camera casing.

The internal geometry of a camera is assumed to be unknown because of the uncertainties due to manufacturing and assembly errors and material imperfections, although its manufacturer declares these parameters. Improvements in manufacturing technologies and material science reduced these uncertainties. However, the topic of determination of internal camera calibration is still popular because of the fact that the image sensor dimensions are in micro scale and is still decreasing as result of developing technology.

As a result of assumptions and ignoring some uncertainties, the problem of determining internal camera geometry can be reduced to determination of the following camera parameters: image plane distance, principal point, lens distortion parameters and pixel dimensions.

Image plane distance is the distance of the image plane to the center of the projection or the principal point, which is taken as is the origin of the image plane coordinate system. Orientation uncertainties in the lens and image plane with respect to the camera are assumed to be negligible. Therefore, the problem of determination of positions and orientations of the lens and image plane with respect to camera casing is reduced to the determination of two parameters: image plane distance and the location of principal point.

Pixel dimensions can be defined as the width and height of the one pixel on the image sensor. On the other hand, the pixels structure is not contiguous in most types of image sensors, so the definition should be modified as horizontal and vertical distances between pixel centers. The determination of pixel dimensions experimentally is not possible since effects the uncertainty of the image plane distance on image formation is in the same order with those of pixel dimensions. But neglecting the uncertainties in pixels dimensions does not create problems, since these uncertainties are handled during determination of image plane distance.

Still the most commonly encountered problem in image processing and computer vision projects is the distortions in the image due the material and manufacturing imperfections of lenses. As stated in Chapter 2, lenses are approximations to the perspective projection with a pin-hole camera and due to the geometric aberrations and material defects of the lens the image formation rules deviates from the ideal perspective projection. The problem of retrieving the image formation to perspective projection is reduced to construction of the distortion model, measuring distortions, determination of the distortion parameters, and reversing the distortions in the image to obtain the ideal image.

The problem of determination of image plane distance was solved in previous study [1] by introducing an iterative procedure to the 3D reconstruction method. The depth of the secondary target, d, is a function of image plane distance, d_0, while reconstruction of other two translations is dependent on depth. In the proposed method, the image plane distance is obtained by iterating the following equation,

$$d_0^{(i+1)} = \frac{d(d_0^{(i)})d_0^{(1)}}{d(d_0^{(i)}) - d_0^{(1)}}$$

(4.1)

until the relative error,

$$\left| d_0^{i+1} - d_0^i \right| < \varepsilon$$

(4.2)

decreases below a constant threshold, ε.

In the previous study [2], a method was proposed to determine the position of the principal point (center of perspective projection) with respect to camera casing, thus with respect to the world coordinate system. The proposed method utilizes the relative motion of the objects at different depths when the camera is moving. In the calibration procedure, the camera is mounted on an X-Z table which also mounted on a Y-rotary table. If the rotation axis does not coincide with the principal point, the location of the principal point with respect to world coordinate system and also with respect to the objects in the scene alternates (Figure 4.1). Thus, when the camera is rotated about the axis, a relative motion between the images of the objects at different depths occurs. The axis of rotation and the principal point are coincided by adjusting the position of the camera on the X-Z table to have no relative motion between the images of objects at different depths (Figure 4.1), the x and z position of the perspective projection center with respect to camera casing is obtained.
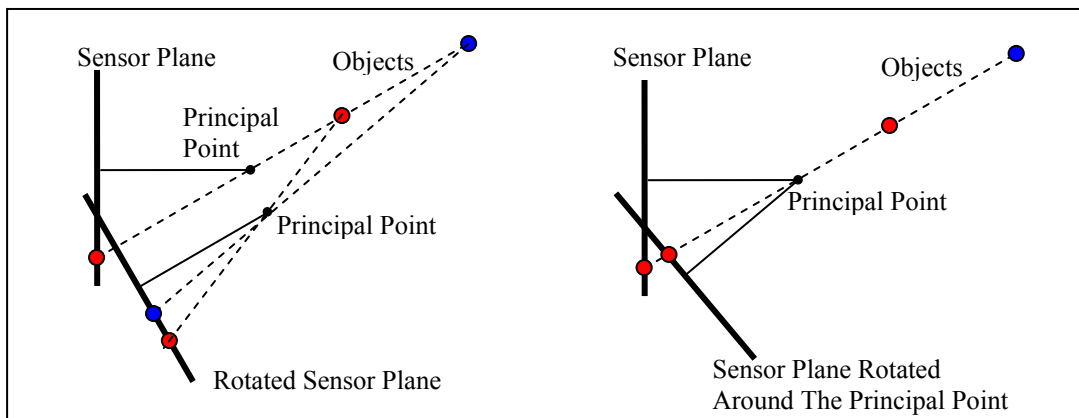

Figure 4.1. Determination of Principal Point of the Camera.

82

Although the experimental setup does not permit, it is theoretically possible to determine the y position of the perspective projection center with respect to the camera casing by applying the same procedure on Y-Z table and X-rotary tables. Moreover, it is possible to determine the optical axis and principal point by rotating the camera about an axis parallel to the z-axis and passing through the perspective projection center. The single non-rotating point on the image will be the principal point.

In previous study [4], a lens distortion model and a calibration procedure was proposed. On the other hand, the method was not implemented and the distortion correction was not considered in that study. In this study, a complete solution to the distortion model suggested in study [4] is proposed and implemented.

4.2 Distortion Model

In most computer vision research studies, only a radial distortion model is used, because of its dominancy to other types of distortions. In this type of distortion model, pixels appear closer or farther to the origin as a function of even powers of their undistorted distance to the origin. Hence, the horizontal and vertical components of radial distortion, $\delta u_i^{(r)}$ and $\delta v_i^{(r)}$, can be expressed as

$$\begin{bmatrix} \delta u_i^{(r)} \\ \delta v_i^{(r)} \end{bmatrix} = \begin{bmatrix} u_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4 + \ldots) \\ v_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4 + \ldots) \end{bmatrix} \tag{4.3}$$

where $u_i^{(ud)}$ and $v_i^{(ud)}$ are the undistorted horizontal and vertical image coordinates, $\{k_1, k_2, \ldots\}$ are the distortion coefficients. In most studies [27 -31], only k_1 and k_2 are considered and the other coefficients are assumed to be negligible. r_i is the radial distance of the undistorted image coordinates to the origin of the image plane and it can be expressed as

$$r_i^2 = (u_i^{ud})^2 + (v_i^{ud})^2 \tag{4.4}$$

The cause of radial distortion is bending of light rays more or less than the correct amount due to surface aberrations of the lens [30]. An image containing high amount of radial distortion is seen in Figure 4.2.

Figure 4.2. A Distorted Image

Tangential distortion is another type of distortion which is commonly encountered in computer vision problems. The main cause of a tangential distortion is the existence of an angle between the optical axis of the lens and image plane normal. The horizontal and vertical components of tangential distortion, $\delta u_i^{(t)}$ and $\delta v_i^{(t)}$, can be expressed as:

$$\begin{bmatrix} \delta u_i^{(t)} \\ \delta v_i^{(t)} \end{bmatrix} = \begin{bmatrix} 2 p_1 u_i^{(ud)} v_i^{(ud)} + p_2 [r_i^2 + 2(u_i^{(ud)})^2] \\ p_1 [r_i^2 + 2(v_i^{(ud)})^2] + 2 p_1 u_i^{(ud)} v_i^{(ud)} \end{bmatrix} \tag{4.5}$$

where p_1 and p_2 are the tangential distortion coefficients.

The proposed distortion model in study [4] is sum of radial and tangential distortion models. Than the cumulative horizontal and vertical components of the distortion, δu_i and δv_i can be expressed as:

$$\begin{bmatrix} \delta u_i \\ \delta v_i \end{bmatrix} = \begin{bmatrix} u_i - u_i^{(ud)} \\ v_i - v_i^{(ud)} \end{bmatrix} = \begin{bmatrix} \delta u_i^{(r)} + \delta u_i^{(t)} \\ \delta u_i^{(r)} + \delta u_i^{(t)} \end{bmatrix} \tag{4.6}$$

where u_i and v_i are the distorted image coordinates. By combining Equations (4.3), (4.5), and (4.6), the cumulative horizontal and vertical components of the distortion, δu_i and δv_i can be expressed as:

$$\begin{bmatrix} \delta u_i \\ \delta v_i \end{bmatrix} = \begin{bmatrix} u_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + 2p_1 u_i^{(ud)} v_i^{(ud)} + p_2[r_i^2 + 2(u_i^{(ud)})^2] \\ v_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + p_1[r_i^2 + 2(v_i^{(ud)})^2] + 2p_2 u_i^{(ud)} v_i^{(ud)} \end{bmatrix} \quad (4.7)$$

In Figures 4.3, 4.4, and 4.5, pure radial, pure tangential, and composite distortions are visualized, respectively. In these figures, the blue lines represent an original and undistorted image grid, while the green curves are obtained by the distorting the original lines by respective distortions. As seen in the figures, both distortion models do not affect the image center, although the distortion parameters used are extremely exaggerated. On the other hand, the affect of both distortion increases dramatically as the distance to origin increases.

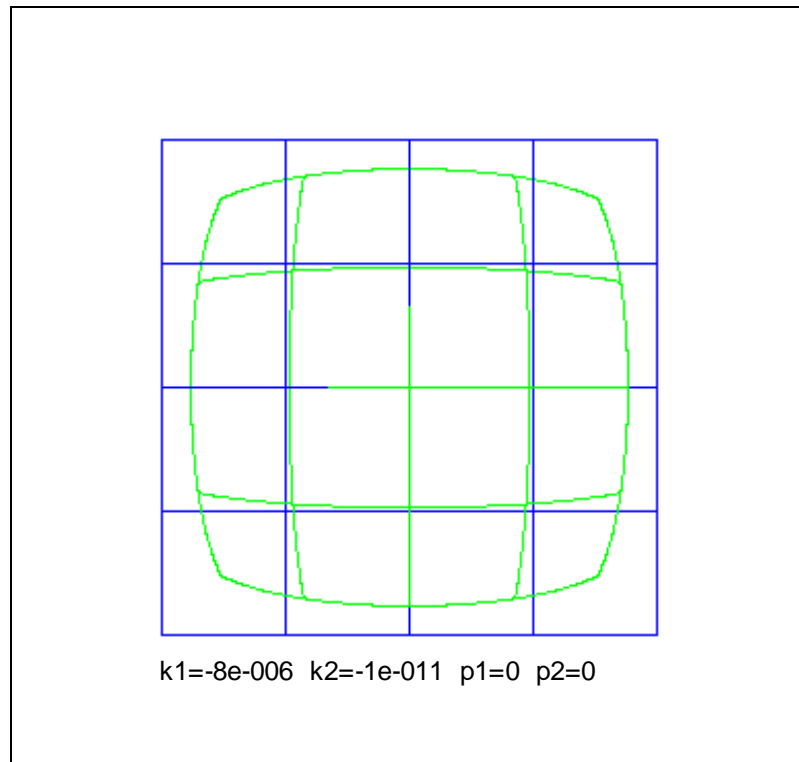k1=-8e-006 k2=-1e-011 p1=0 p2=0
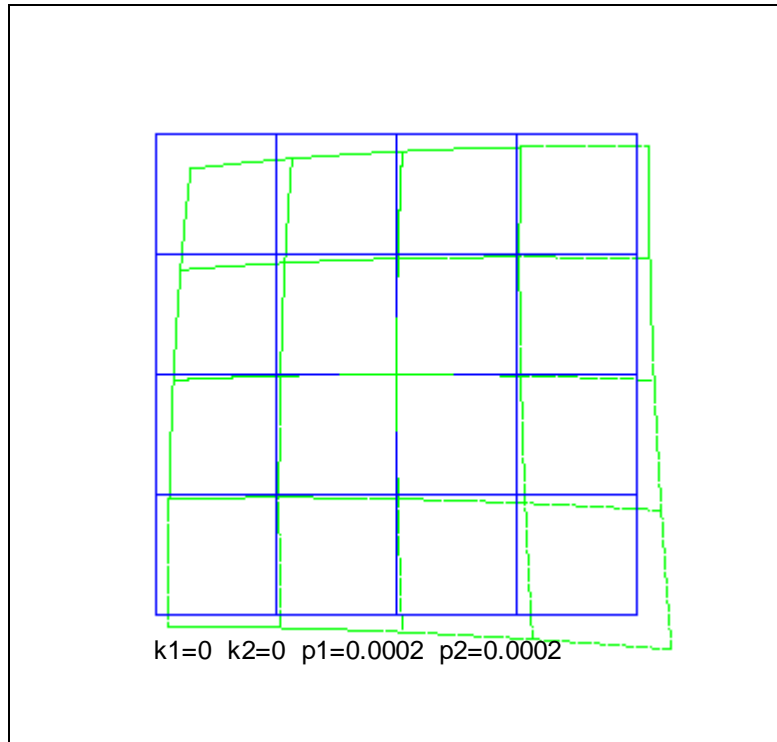
Figure 4.3. Pure Radial Distortion

85

k1=0 k2=0 p1=0.0002 p2=0.0002

Figure 4.4. Pure Tangential Distortion
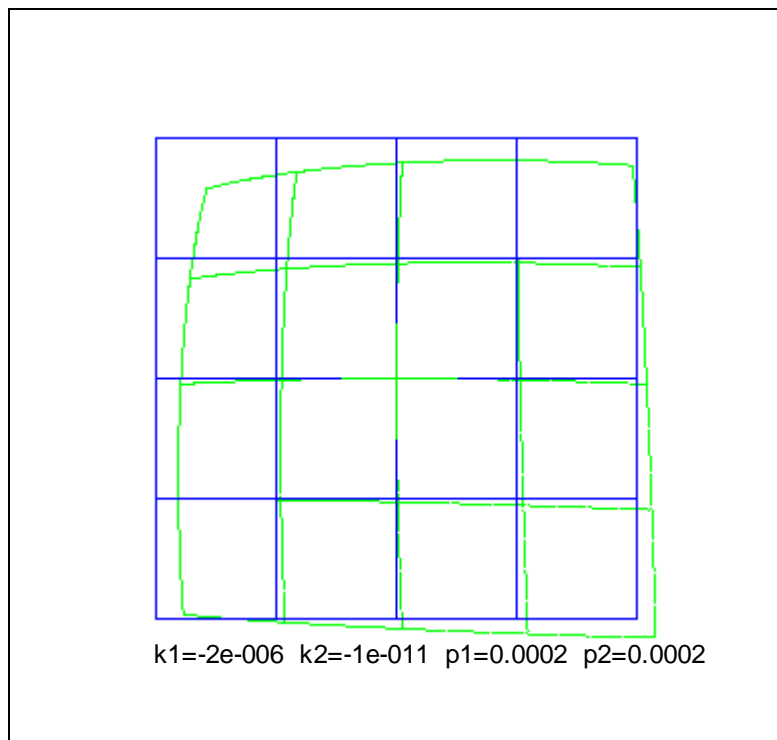

k1=-2e-006 k2=-1e-011 p1=0.0002 p2=0.0002

Figure 4.5. Radial and Tangential Distortion

The proposed calibration procedure in the study [4] utilizes the fact that the distortion at the pixels closer to the origin of the image plane is negligible even for the highly exaggerated distortion parameters. It is suggested to attach the secondary passive target on the calibration pattern and determine the calibration plane position by using the target and 3D reconstruction algorithm when the target is close to origin in the image.

4.3 Calibration Pattern and Distortion Measurement

It is possible to determine the distortion coefficients only if both the distorted and undistorted coordinates of a number of points on the image are known. Image plane coordinates already give the distorted coordinates of all points. However, in order to obtain the undistorted image coordinates, the geometry of the imaged scene and the position and orientation of the scene with respect to camera coordinate system should be known or calculated together with the distortion parameters. Then the exact locations of the features, e.g. lines and points, on the image plane are calculated by using ideal perspective projection and compared with their coordinates in the distorted image to obtain distortion parameters.

A straightforward solution of the problem would be using a planar calibration pattern accurately positioned in front of the camera with a known 3D configuration. The next step is the construction of an ideal image using the geometry and 3D configuration of the calibration pattern and internal camera parameters other than lens distortion parameters. Finally, it is possible to measure the distortion field on the distorted image by comparing it with the ideal image. Unfortunately, the positioning the calibration pattern in front of the camera accurately is not an easy job and generally impossible in on-site processes.

In the previous study [4], the use of a calibration pattern composed of equally spaced spots and containing the secondary passive target at its center is suggested. In this study, it is shown in Figures 4.3, 4.4, and 4.5 that the center of the image is

not affected remarkably even for very high values of distortion parameters and the suggested calibration procedure is implemented.

On the other hand, the calibration pattern is redesigned (Figure 4.5) to increase the reconstruction accuracy. The suggested calibration pattern in the study [4] contained a modified and a smaller secondary passive target. In this study, the secondary target is used as it is, but the size of calibration pattern is enlarged to cover the whole image when the secondary target is inside the distortion free region as shown in Figures 4.3, 4.4 and 4.5. The selection of spacing between the spots is another problem. Decreasing the spacing between the spots therefore using a denser spot pattern would increase the number of data points, thus the accuracy of distortion parameter determination process. However, matching the ideal positions of the spots and distorted spots becomes impossible, since ideal position of an outer spot may appear between the ideal position and distorted position of an inner spot depending on the amount of distortion and spacing between the spots. The spacing between the spots is selected as the minimum possible value which does not create matching problems.
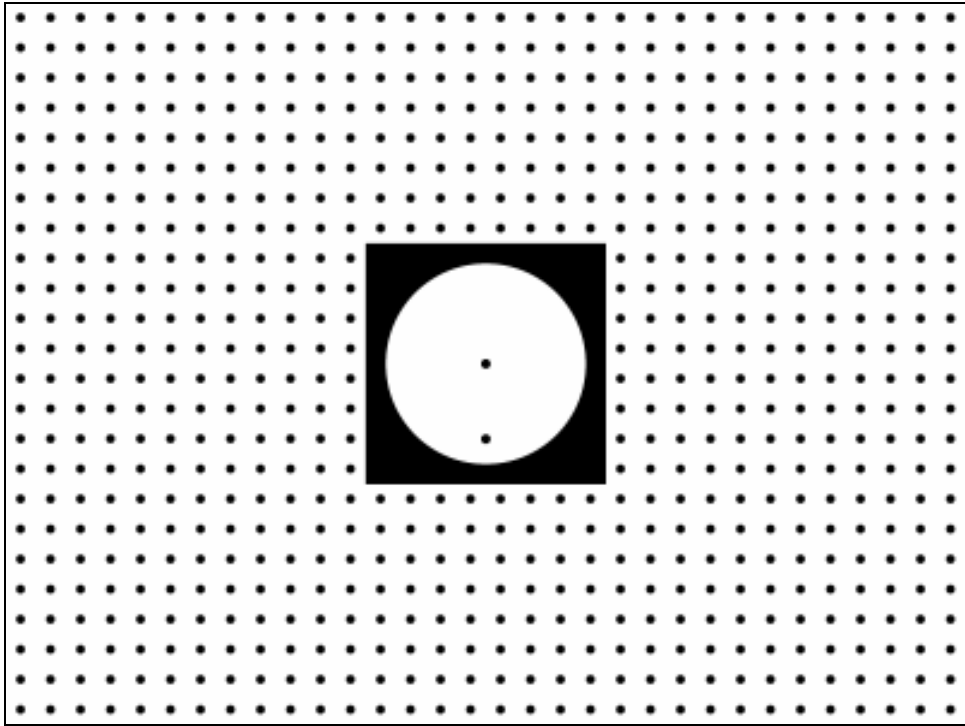
Figure 4.6. Calibration Pattern

The calibration pattern is positioned in front of the camera arbitrarily, but roughly 45 degree rotated positions are preferred due to fact that the accuracy of the 3D reconstruction method is higher for this rotation amount [1, 2]. Then, the 3D configuration of the secondary passive target is reconstructed using the image of the calibration pattern without distortion correction. Next, the ideal positions of the equally spaced black spots are determined using ideal perspective projection and the centers of the black spots in the distorted image are determined. The ideal positions and distorted positions are matched, starting from the closest point to origin.

Figure 4.6 shows the matching between distorted and ideal positions of the calibration spots. In this figure, the ideal positions of the spots are marked as red points and yellow lines shows the matching.

Figure 4.6. Feature Matching.

The undistorted image coordinates, $u_i^{(ud)}$ and $v_i^{(ud)}$, and corresponding horizontal and vertical distortions, δu_i and δv_i, are calculated using the matches in Figure 4.6. To increase the number of data points and homogeneity of distribution of data points over the entire image plane, the feature matching method is implemented to permit the use of maximum four calibration images. It is preferred to use the four images of calibration pattern rotated approximately ±45° about x and y axes, due to the fact that the accuracy of the reconstr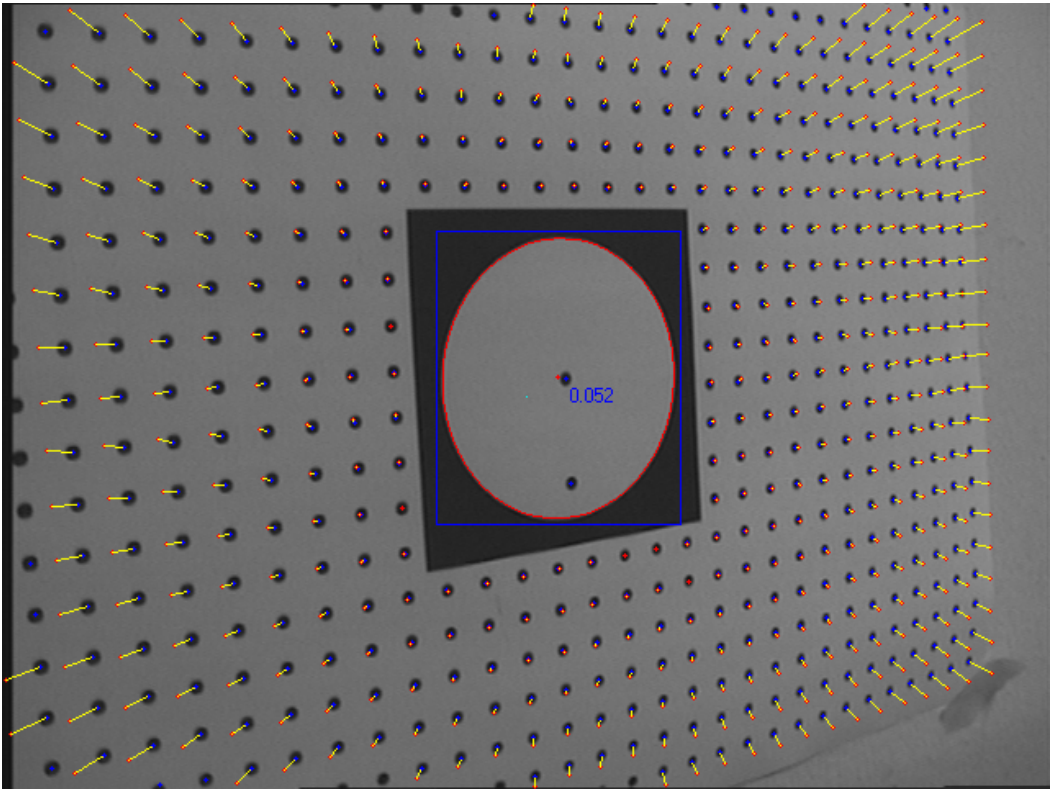uction method is high for these orientations. Also the secondary passive target on the calibration pattern should be centered at the calibration images.

4.4 Determination of Calibration Parameters

The determination of the distortion parameters is a well defined problem, once the distortion model is selected and the horizontal and vertical distortions, δu_i and δv_i, and the corresponding undistorted image coordinates, $u_i^{(ud)}$ and $v_i^{(ud)}$ are extracted. The distortion Equations (4.7) for each data obtained by feature matching with the unknown distortion parameters can be expressed as the following system of linear equations.

$$\begin{bmatrix} u_1^{ud} r_1^2 & u_1^{ud} r_1^4 & 2u_1^{ud} v_1^{ud} & (r_1^2 + 2(u_1^{ud})^2) \\ v_1^{ud} r_1^2 & v_1^{ud} r_1^4 & (r_1^2 + 2(v_1^{ud})^2) & 2u_1^{ud} v_1^{ud} \\ u_2^{ud} r_2^2 & u_2^{ud} r_2^4 & 2u_2^{ud} v_2^{ud} & (r_2^2 + 2(u_2^{ud})^2) \\ v_2^{ud} r_2^2 & v_2^{ud} r_2^4 & (r_2^2 + 2(v_2^{ud})^2) & 2u_2^{ud} v_2^{ud} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ u_N^{ud} r_N^2 & u_N^{ud} r_N^4 & 2u_N^{ud} v_N^{ud} & (r_N^2 + 2(u_N^{ud})^2) \\ v_N^{ud} r_N^2 & v_N^{ud} r_N^4 & (r_N^2 + 2(v_N^{ud})^2) & 2u_N^{ud} v_N^{ud} \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \delta u_1 \\ \delta v_1 \\ \delta u_2 \\ \delta v_2 \\ \vdots \\ \vdots \\ \delta u_N \\ \delta v_N \end{bmatrix} \qquad (4.8)$$

The unknown distortion parameter vector, $\mathbf{x} = \begin{bmatrix} k_1 & k_2 & p_1 & p_2 \end{bmatrix}$ is solved by pseudo inversion technique as follows

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \qquad (4.9)$$

90

where

$$\mathbf{A} = \begin{bmatrix} u_1^{ud} r_1^2 & u_1^{ud} r_1^4 & 2u_1^{ud} v_1^{ud} & (r_1^2 + 2(u_1^{ud})^2) \\ v_1^{ud} r_1^2 & v_1^{ud} r_1^4 & (r_1^2 + 2(v_1^{ud})^2) & 2u_1^{ud} v_1^{ud} \\ u_2^{ud} r_2^2 & u_2^{ud} r_2^4 & 2u_2^{ud} v_2^{ud} & (r_2^2 + 2(u_2^{ud})^2) \\ v_2^{ud} r_2^2 & v_2^{ud} r_2^4 & (r_2^2 + 2(v_2^{ud})^2) & 2u_2^{ud} v_2^{ud} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ u_N^{ud} r_N^2 & u_N^{ud} r_N^4 & 2u_N^{ud} v_N^{ud} & (r_N^2 + 2(u_N^{ud})^2) \\ v_N^{ud} r_N^2 & v_N^{ud} r_N^4 & (r_N^2 + 2(v_N^{ud})^2) & 2u_N^{ud} v_N^{ud} \end{bmatrix} \tag{4.10}$$

and

$$\mathbf{b} = \begin{bmatrix} \delta u_1 & \delta v_1 & \delta u_2 & \delta v_2 & \cdots & \cdots & \delta u_N & \delta v_N \end{bmatrix}^T \tag{4.11}$$

4.5. Distortion Correction

The 3D reconstruction method determines the 3D configuration parameters of the secondary passive target by inverse perspective projection using the image coordinates of the contour and spots of the secondary passive target in the image. Since the image formation deviates from the ideal perspective projection because of the lens distortion, the image coordinates of the pixels corresponding to target contour and spots should be corrected using the distortion parameters before the reconstruction process.

If the distortion Equation (4.7) composed of radial and tangential distortion is rearranged as

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} u_i^{ud} + u_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + 2p_1 u_i^{(ud)} v_i^{(ud)} + p_2[r_i^2 + 2(u_i^{(ud)})^2] \\ v_i^{ud} + v_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + p_1[r_i^2 + 2(v_i^{(ud)})^2] + 2p_2 u_i^{(ud)} v_i^{(ud)} \end{bmatrix} \tag{4.12}$$

it is seen that the distorted horizontal and vertical image coordinates, u_i and v_i, are non-linear functions

$$u_i = u_i^{(ud)} + f_u(u_i^{(ud)}, v_i^{(ud)}) \tag{4.13}$$

$$v_i = v_i^{(ud)} + f_v(u_i^{(ud)}, v_i^{(ud)}) \tag{4.14}$$

of undistorted horizontal and vertical image coordinates, $u_i^{(ud)}$ and $v_i^{(ud)}$, where

$$f_u(u_i^{(ud)}, v_i^{(ud)}) = u_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + 2p_1 u_i^{(ud)} v_i^{(ud)} + p_2[r_i^2 + 2(u_i^{(ud)})^2] \tag{4.15}$$

$$f_v(u_i^{(ud)}, v_i^{(ud)}) = v_i^{(ud)}(k_1 r_i^2 + k_2 r_i^4) + p_1[r_i^2 + 2(v_i^{(ud)})^2] + 2p_2 u_i^{(ud)} v_i^{(ud)} \tag{4.16}$$

and

$$r_i^2 = (u_i^{ud})^2 + (v_i^{ud})^2. \tag{4.17}$$

The problem of distortion correction can be defined as inversing the distortion using the distorted image coordinates and distortion parameters. As seen in Equations (4.13), (4.14), (4.15), (4.16) and (4.17) the distortions functions are coupled and include terms up to ninth order, and there is no explicit inverse function of these distortion equations.

One of the most common techniques in inverse distortion models is forming a lookup table using the distortion model parameters and undistorted image coordinates, and then calculating the undistorted image coordinates by 2D polynomial interpolation using the lookup table. As an alternative to this approach, in the recent study [31], the radial distortion model is inverted by Taylor series expansion.

This study also utilizes the Taylor series approximation to obtain the inverse functions of the distortion equations, Equations (4.13) and (4.14), corresponding to

compound radial and tangential distortion. The inverse equation can be written as follows if the higher order terms are omitted:

$$u^{(ud)} = u + \left(\frac{-f_u(u,v) + f_v(u,v)\dfrac{\delta f_u(u,v)}{\delta v} - f_u(u,v)\dfrac{\delta f_v(u,v)}{\delta v}}{1 + \dfrac{\delta f_u(u,v)}{\delta u} + \dfrac{\delta f_v(u,v)}{\delta v}} \right) \qquad (4.18)$$

$$v^{(ud)} = v + \left(\frac{-f_v(u,v) + f_u(u,v)\dfrac{\delta f_v(u,v)}{\delta u} - f_v(u,v)\dfrac{\delta f_u(u,v)}{\delta u}}{1 + \dfrac{\delta f_u(u,v)}{\delta u} + \dfrac{\delta f_v(u,v)}{\delta v}} \right) \qquad (4.19)$$

where $u^{(ud)}$ and $v^{(ud)}$ are undistorted image coordinates corresponding to distorted image coordinates u and v. The partial derivatives of Equations (4.15) and (4.16) are derived analytically and substituted into inverse distortion Equations (4.18) and (4.19) during implementation.

The performance of the inverse distortion formulas is evaluated with synthetically distorted images and the results are presented in Figures 4.7, 4.8, 4.9, and 4.10. In these figures, undistorted pixels are plotted with blue color, green colored pixels are the distorted ones and finally the red colored pixels indicate the corrected coordinates by the proposed inverse distortion formulas.

k1=-4e-006 k2=-1e-011 p1=0 p2=0
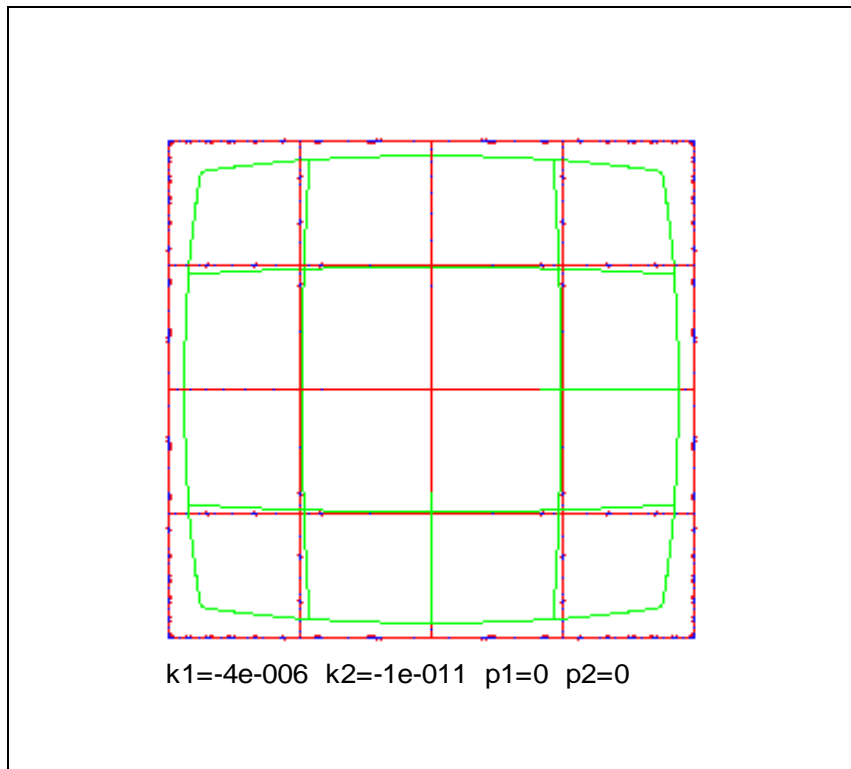
Figure 4.7. Distortion Correction (Pure Radial Distortion).

k1=0 k2=0 p1=0.0002 p2=0.0002
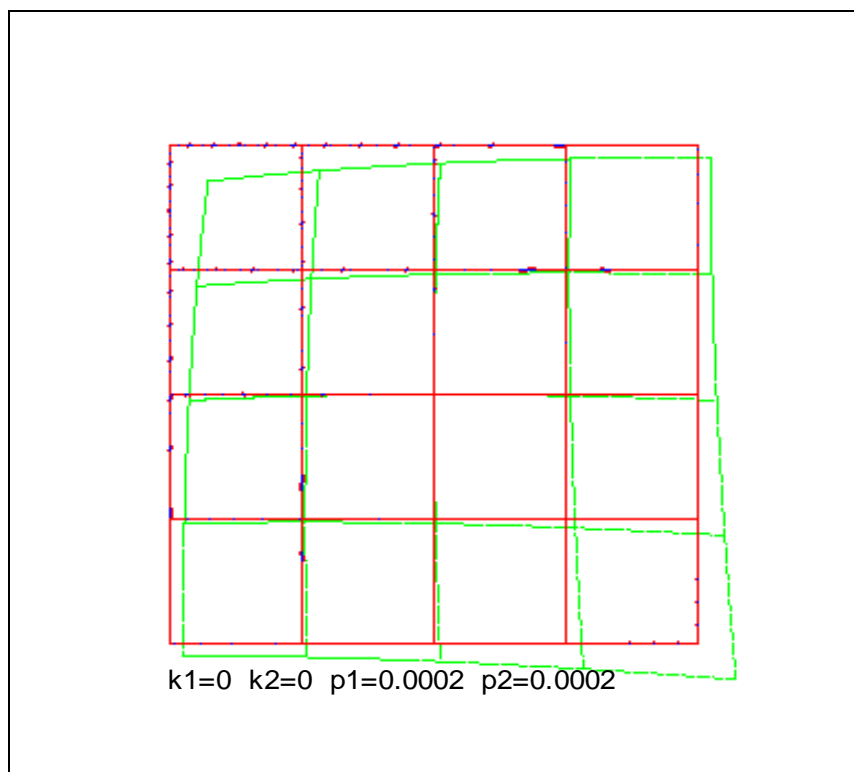
Figure 4.8. Distortion Correction (Pure Tangential Distortion).

94

k1=-4e-006 k2=-1e-011 p1=0.0002 p2=0.0002
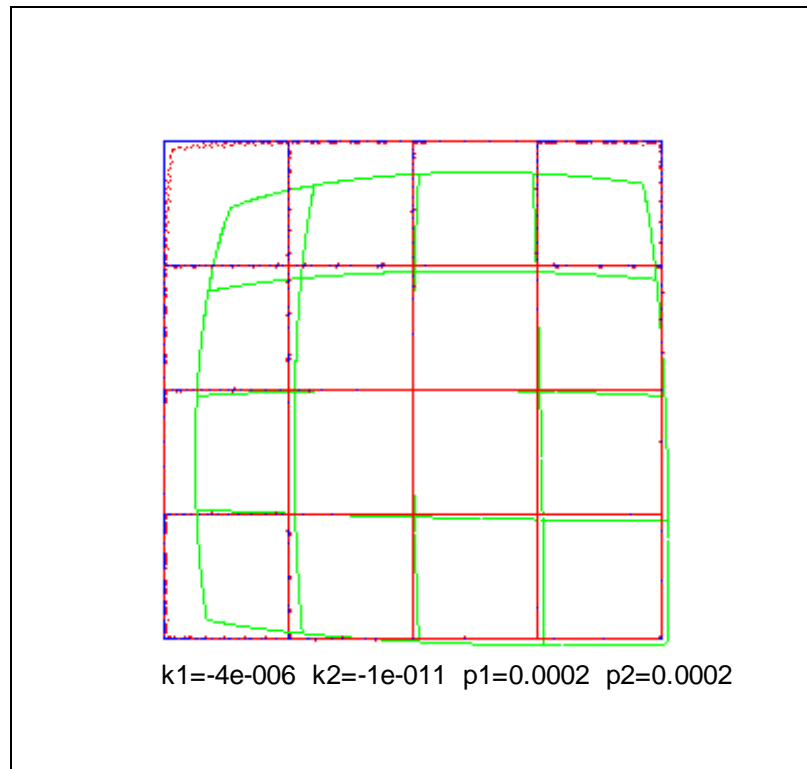
Figure 4.9. Distortion Correction (Composite Distortion).

k1=-2.0576e-007 k2=-1.2091e-013 p1=3.3869e-006 p2=2.5558e-006

Figure 4.10. Distortion Correction (Determined Camera Distortion).

As seen in Figures 4.7 and 4.8, the inverse distortion method obtained by Taylor series expansion can correct the pure radial distortions with a maximum one pixel error. Also, Figure 4.9 shows that the inverse distortion method can correct the high amount of composite distortion with a maximum 6 pixels error. Finally, in Figure 4.10, distortion parameters determined for the camera is used to simulate the performance of the inverse distortion method. The maximum correction error for camera distortion parameters does not exceed one pixel.

The results of the simulations presented in Figures 4.7, 4.8, 4.9, and 4.10 show that the proposed and implemented inverse correction method is accurate enough to correct the radial and tangential distortions of the camera system used in this study. The inverse distortion formulas are combine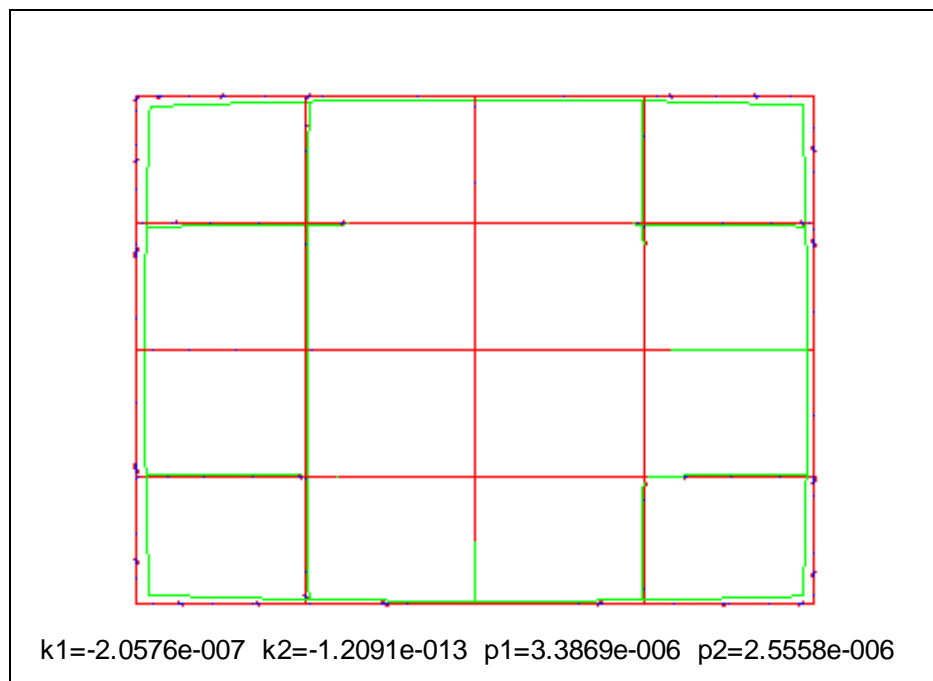d with the 3D reconstruction method, the image coordinates of the target contour and spot centers are corrected before the inverse perspective projection.

4.6 Calibration Procedure and Results

In this Chapter, a calibration pattern and the distortion measurement method are presented. A method is proposed to determine the distortion parameters of a composite radial and tangential distortion model using the measured distortions. Finally, distortion correction method for the distortion model is presented.

The proposed lens distortion model, distortion measurement, distortion parameters determination and distortion correction methods implies the following calibration procedure.

1. Position the calibration pattern in front of the camera, capture the four images of the pattern rotated ±45° about x and y axes with respect to camera and centering the secondary passive target on the calibration pattern.
2. Set the distortion parameters to zero, initially.

3. Reconstruct the calibration pattern plane configuration parameters using the image of the secondary passive target by distortion correction embedded 3D reconstruction method.

4. Calculate the ideal positions of the spots of the calibration pattern on the image plane using the reconstructed 3D configuration of the calibration pattern plane.

5. Determine the centers of the spots of the calibration pattern viewed in the distorted images.

6. Match the ideal and actual positions of the calibration spots and calculate the horizontal and vertical distortions and the corresponding undistorted coordinates.

7. Determine the distortion parameters.

8. Iterate the steps 3 to 7, until the relative absolute distortion parameter errors become smaller than a required threshold value.

The proposed calibration procedure is implemented such that it is fully autonomous and does not require assistance except roughly positioning the calibration pattern in front of the camera. The distortion parameters obtained as result of the proposed calibration procedure are as follows.

$$\begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} -2.05 \cdot 10^{-7} \\ -1.21 \cdot 10^{-13} \\ 3.39 \cdot 10^{-6} \\ 2.56 \cdot 10^{-6} \end{bmatrix}$$

The absolute relative error limit for distortion parameters are taken as 0.01 in the calibration procedure, and it converges in the third iteration.

The determination of image plane distance, principal point and center of perspective projection was solved in previous studies [1, 2]. In this study, the remaining part of internal camera calibration, the lens distortion, is solved using a composite distortion model composed of radial and tangential distortions as suggested in the previous study [4].

CHAPTER 5

IMPLEMENTATION AND CASE STUDIES

5.1 Overview

In this Chapter, the implementation of the proposed methods in Chapters 2, 3, and 4 is discussed. The hardware used and overall algorithm of the software developed are presented. The results of a set of case studies which are designed to test the software are discussed.

5.2. Implementation

The software developed includes the all proposed methods in this study and almost all methods developed in previous studies [1-4]. A manual for the software is presented in Appendix A. The hardware, camera, frame grabber, and computer used in this study are different than the ones used in previous studies.

5.2.1. Hardware

In the previous studies [1-4], a CID camera, CIDTEC 2250D was preferred to test the developed methods because of its contiguous pixel structure, non-destructive read-out property, and very low noise signal ratio. Unfortunately, this camera broke down during the last phase of the earlier study hence currently is not available. Therefore, and it has to be replaced by a CCD camera, SONY EVI D-31, which is borrowed from the Mechatronic Design Laboratory of the Mechanical Engineering Department of METU. The specifications of the camera are presented in Table 5.1, [32].

Table 5.1. Camera Specifications

	SONY EVI D-31
Video Signal	NTSC
Image Sensor	1/3"Hyper HAD Color CCD
Effective Pixels	768(H)X492(V)
H. Resolution	460TV lines
V. Resolution	350TV lines
Lens	X12 Power Zoom, to f=5.4 to 64.8mm, F1.8 to F2.7
S/N Ratio	more than 48dB
Frame Rate	25 FPS
Video Out	Composite, S-Video

An Imagenation PXC200A frame grabber is used digitize the video signal of the camera. Table 5.2 shows the specifications of the frame grabber [33].

Table 5.2. Frame Grabber Specifications

	Imagenation PXC200A
Input Video Formats	NTSC, PAL, SECAM, S-Video
Input Video Signal	1 V peak-to-peak, 75 Ω
Resolution	NTSC: 640x480 PAL/SECAM 768x576
Frame Rate	25 FPS
Output Formats	Color: YCrCb 4:2:2; RGB 32 24 16 15 Monochrome: Y8
Video Noise	≤1 LSB RMS

A Pentium IV 2.8 GHz 512 MB DDR Ram personal computer is used to mount the frame grabber and test the developed and implemented methods in this study. All processing speeds or times presented in this manuscript are measured for this setup. S-Video inputs and outputs of the frame grabber and the camera are utilized.

5.2.2 Overall Algorithm

All methods discussed in this study, including the methods developed in the previous studies [1-4] are combined in the software developed using C++. In order to provide compatibility between the processes, all of the alternative methods of the same process are designed to have same input and output structure.

The software developed is composed of mainly two parts: the interface and the background service. The interface provides selection options for the method and parameters for all processes. On the other hand, the background service of the program processes the selected input with selected methods and parameters. This type of structure makes it possible to switch methods or parameters without stopping the process.

The flow diagram of the overall algorithm is presented in Figure 5.1. The process of 3D reconstruction starts with the selection of input source. It is possible to select a digital image file, camera or virtual perspective projection as the input source. Except the digital image file input case, all processes are repeated continuously. A 640x480 8-bit grayscale image is obtained if the digital image file or camera is selected as the input source. On the other hand, a 640x480 1-bit binary image is obtained if the virtual perspective projection is selected as the input source. In the image cropping process, the sub-region of the image estimated by target tracking method is cropped. The histogram of the estimated target region is also formed during image cropping. The remaining processes can be grouped into three main parts: target detection, reconstruction, and target tracking.
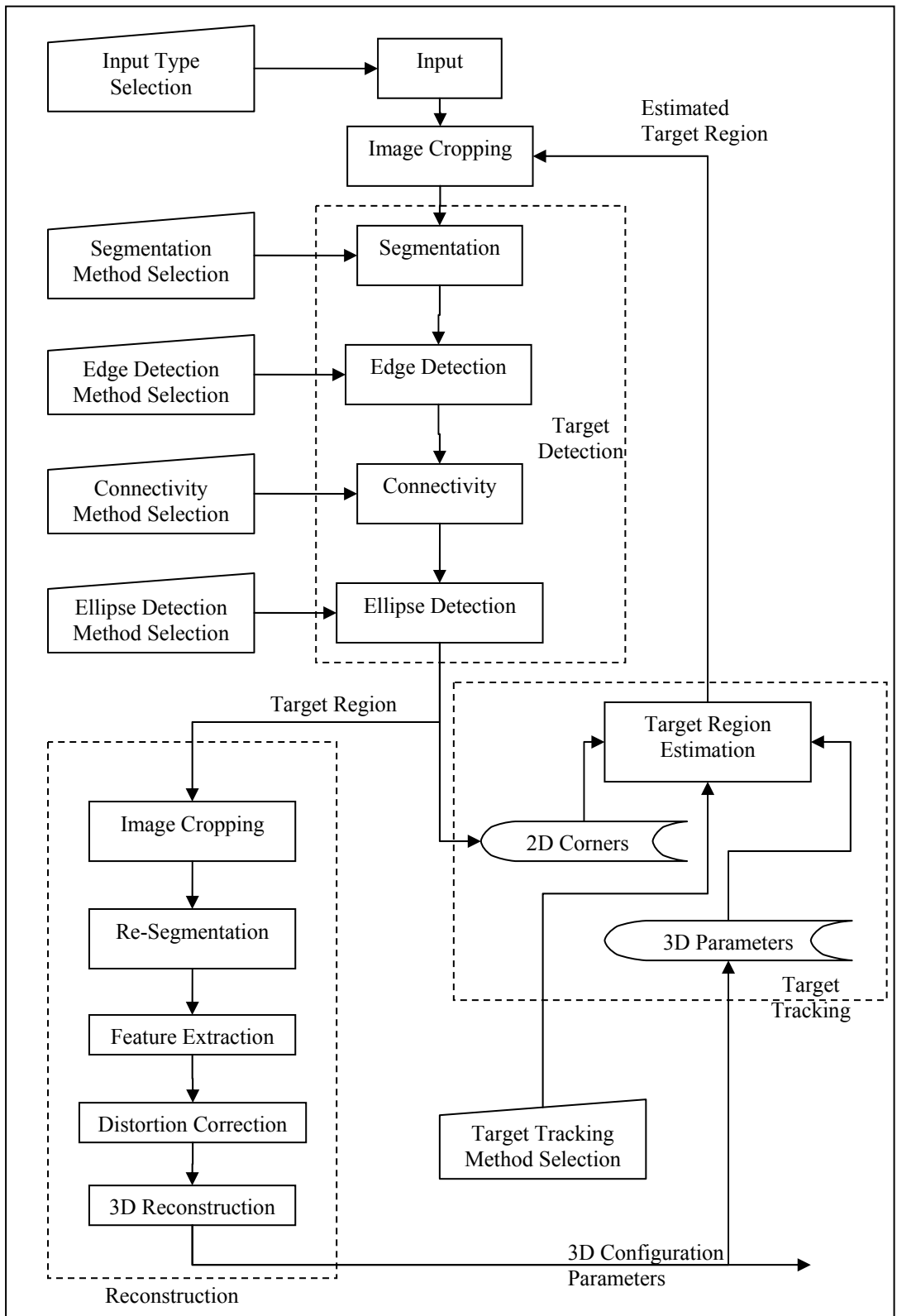
Figure 5.1. Flowchart of the Overall Algorithm.

The target detection part is composed of segmentation, edge detection, connectivity analysis, and ellipse detection steps. If the ellipse corresponding to the target's contour in the image is detected in the target detection part, the coordinates of the target region are fed to 2D corners data block of the target tracking part. The reconstruction part follows the target detection part. If no ellipse is detected, then the process returns to input block.

The reconstruction part is composed of image cropping, re-segmentation, feature extraction, distortion correction, and finally 3D reconstruction methods. The smallest rectangular image area containing the target is cropped and re-segmented first. The pixels corresponding to contour and the spots of the target are detected by feature extraction method. The distortion correction method, calculates the undistorted image coordinates of these feature pixels. Finally the 3D parameters are calculated by 3D reconstruction method. The 3D configuration parameters are fed to the 3D parameters data block of the target tracking part.

The target tracking part of the algorithm is composed of 2D corners and 3D parameters data block and target region estimation method. According to the selected target tracking strategy, the image coordinates of the target are estimated using either 2D corners or 3D parameters obtained from previous images.

5.3. Case Studies

The 3D reconstruction method was tested using a positioning unit and associated absolute reconstruction errors were presented in the previous study [2]. However, the current study integrates the autonomous thresholding and distortion correction methods to the 3D reconstruction and both of these methods may affect the resulting ellipse contour in the image; thus, the 3D configuration parameters may differ. However, during on-site applications, generally an alternative and more accurate measurement technique does not exist, and it is not possible to measure the errors of the 3D reconstruction method absolutely.

In order to test the performance of the overall 3D reconstruction method with the integrated autonomous thresholding and distortion correction methods, two test procedures are designed and used. These tests can also be used to measure the accuracy of the 3D reconstruction method during onsite applications, since these tests do not utilize an alternative measurement technique. In the following subsections these two test procedures are explained and the results of the experiments are presented.

5.3.1. Distance and Angle Measurement

This case is designed to measure the translational and rotational reconstruction accuracy of the overall 3D reconstruction method with changing illumination conditions. A calibration object composed of two secondary passive targets (Figure 5.2) is utilized. The angle between the faces of the object and spacing between the secondary targets are known.
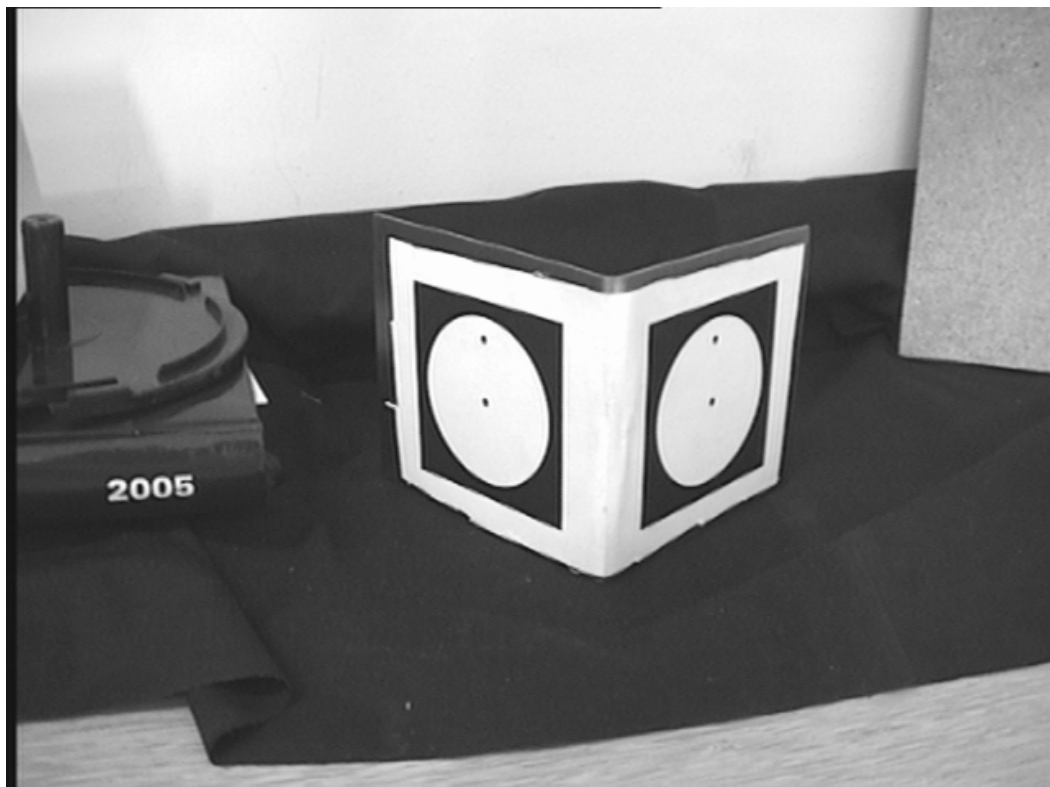

Figure 5.2. Distance Measurement Pattern.

In this test, the object is positioned in front of the camera arbitrarily as shown in Figure 5.2. 3D configurations of each secondary passive target with respect to camera are reconstructed separately. Using the reconstructed 3D configurations, the distance and angle between the secondary passive targets are calculated and compared with actual distance and angle.

In this study, this test is repeated for two different illumination conditions, three different camera locations and with and without distortion correction. The results of the experiments are presented in Figures 5.3, 5.4, 5.5, 5.6, 5.7, and 5.8. In these figures, approximately first half of the measured data is obtained at natural illumination conditions (no direct sun light into the room, curtains closed) while the second half is obtained with some extra lighting (four ceiling florescent lambs are switched on).
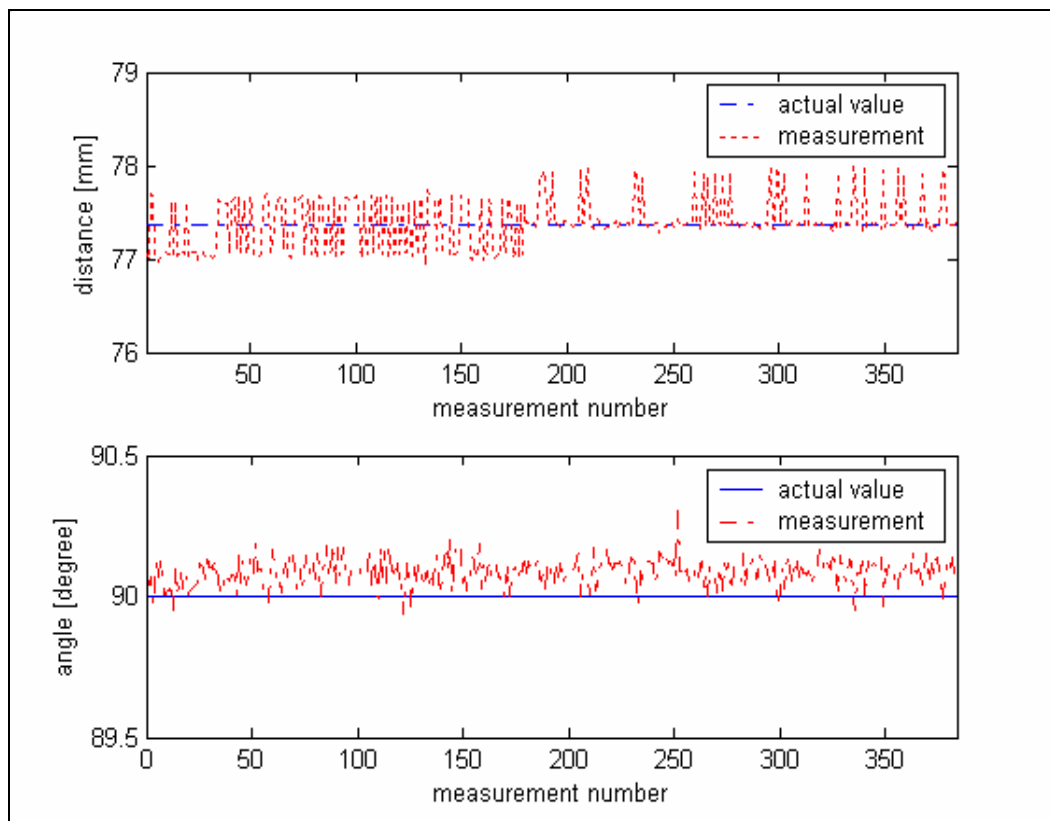

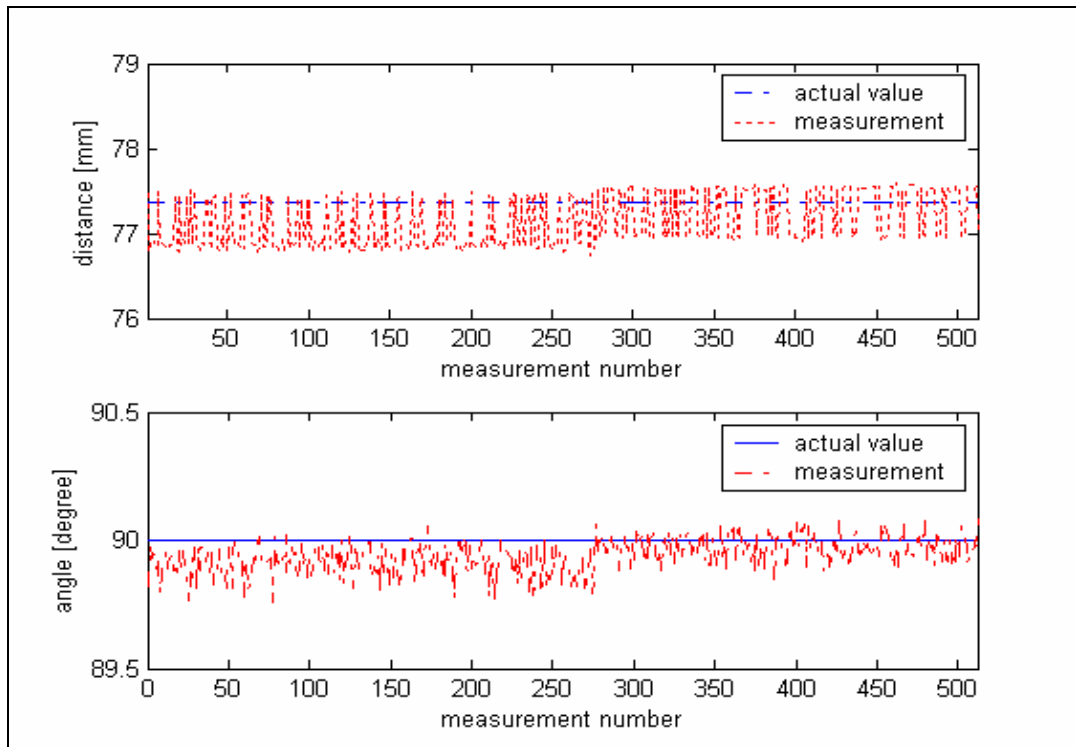**Figure 5.3.** Distance and Angle Measurement, Position 1.

Figure 5.4. Distance and Angle Measurement, Position 1, Distortion Corrected.
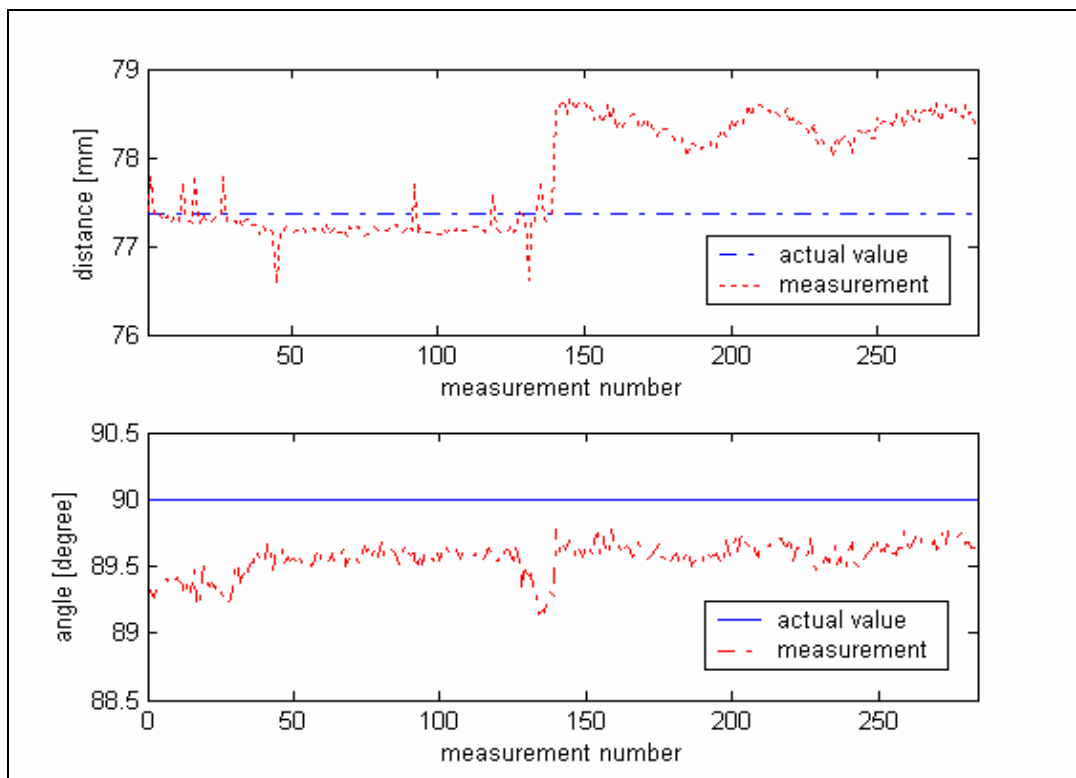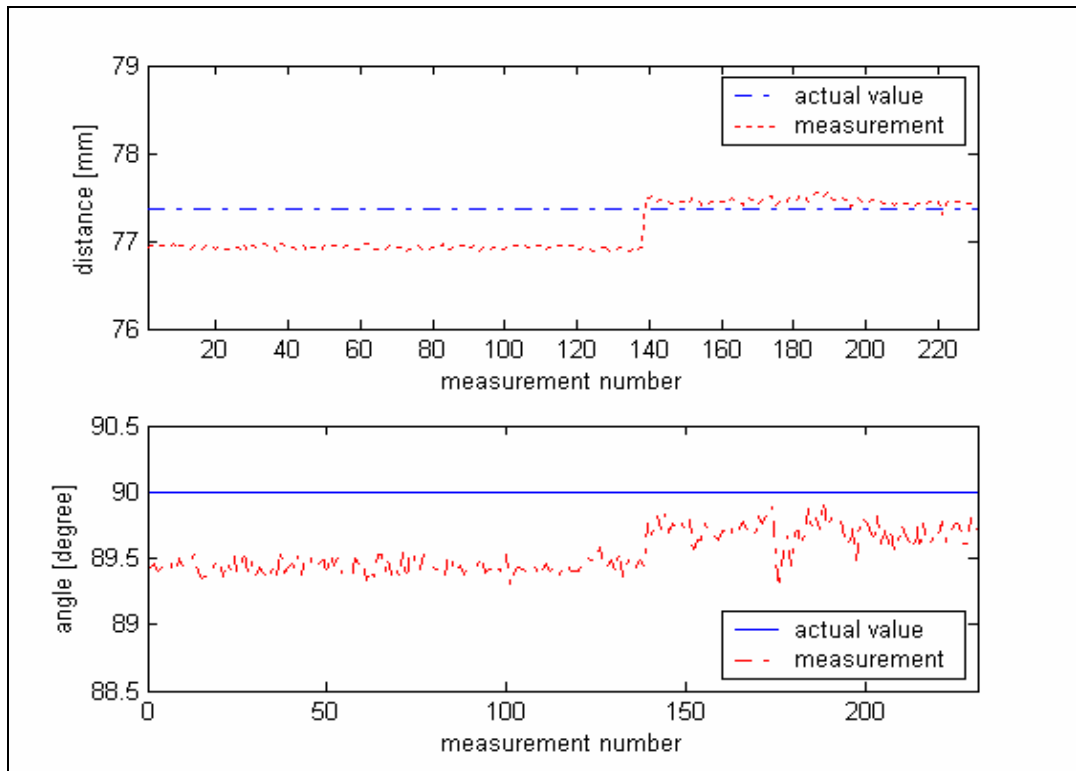

Figure 5.5. Distance and Angle Measurement, Position 2.

105

Figure 5.6. Distance and Angle Measurement, Position 2, Distortion Corrected
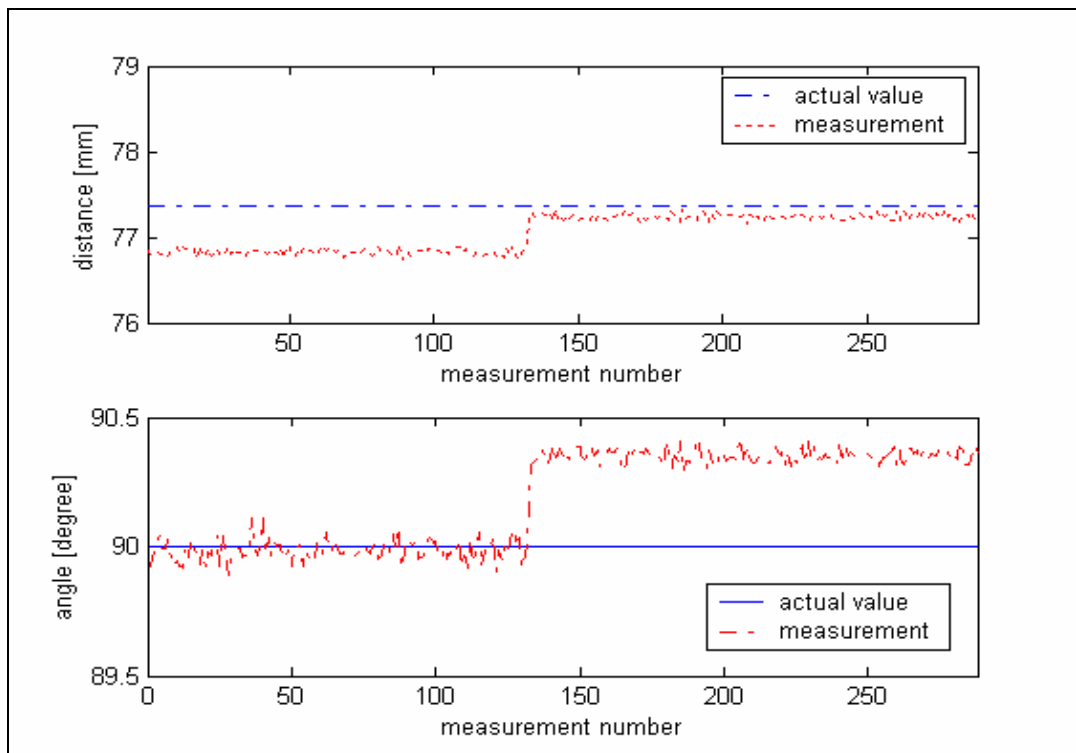
Figure 5.7. Distance and Angle Measurement, Position 3.

106
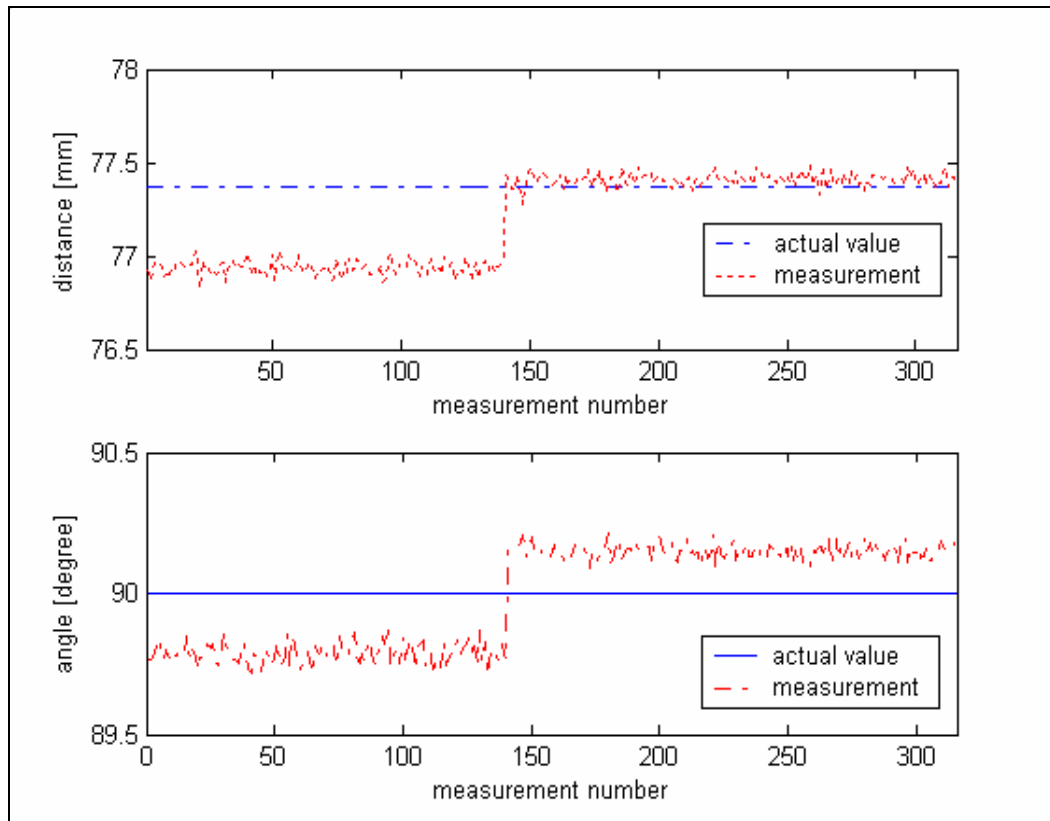
Figure 5.8. Distance and Angle Measurement, Position 3, Distortion Corrected

As seen in these figures, the reconstructed distance and angle between the secondary passive targets are found to be smaller for the natural illumination condition case. If a constant threshold is used, it is expected to have a smaller target in the binary image, which will result in overestimated distances for natural illumination conditions. However the minimum error thresholding method reverses the situation by returning a smaller threshold value for the natural illumination case. The amount of deviations in ellipse center and center spot are also affected by the illumination conditions and threshold value. Since this deviation is also employed in 3D reconstruction method to compute the rotational 3D parameters, the measured angles are also affected by the illumination conditions.

The effects of distortion correction are remarkable in all figures presented above. The distortion correction method reduces the reconstruction errors for extra lighting condition since the distortion parameters are also calculated using extra lighting.

107

5.3.2. g Calculation.

This test is designed to measure the angular and translational reconstruction performance of the 3D reconstruction method and also target tracking performance of the proposed method. In this test, it is aimed to calculate the gravitational acceleration using a secondary passive target attached pendulum. The swinging motion of the secondary passive target on the pendulum is tracked; the reconstructed rotational and translational parameters of the pendulum and the time vector are exported. The exported data is processed to calculate the period and length of the pendulum. The gravitational acceleration is estimated using the following equation.

$$T = 2\pi \sqrt{\frac{l}{g}} \tag{5.1}$$

where, T is the period, l is the length of the pendulum and g is the gravitational acceleration. Note that in Equation (5.1), air resistance is neglected and the swinging object is assumed to be a point mass with zero mass moment of inertia. The air resistance and non-zero mass moment of inertia will decrease the frequency of oscillations, increases the period of the pendulum, thus gravitational acceleration g is underestimated.

The pendulum used in this study is composed of a planar mass and two thin ropes as depicted in Figure 5.9. The mass of the ropes and moment of inertia of the mass are neglected and the secondary passive target is attached at the center of gravity of the planar mass.
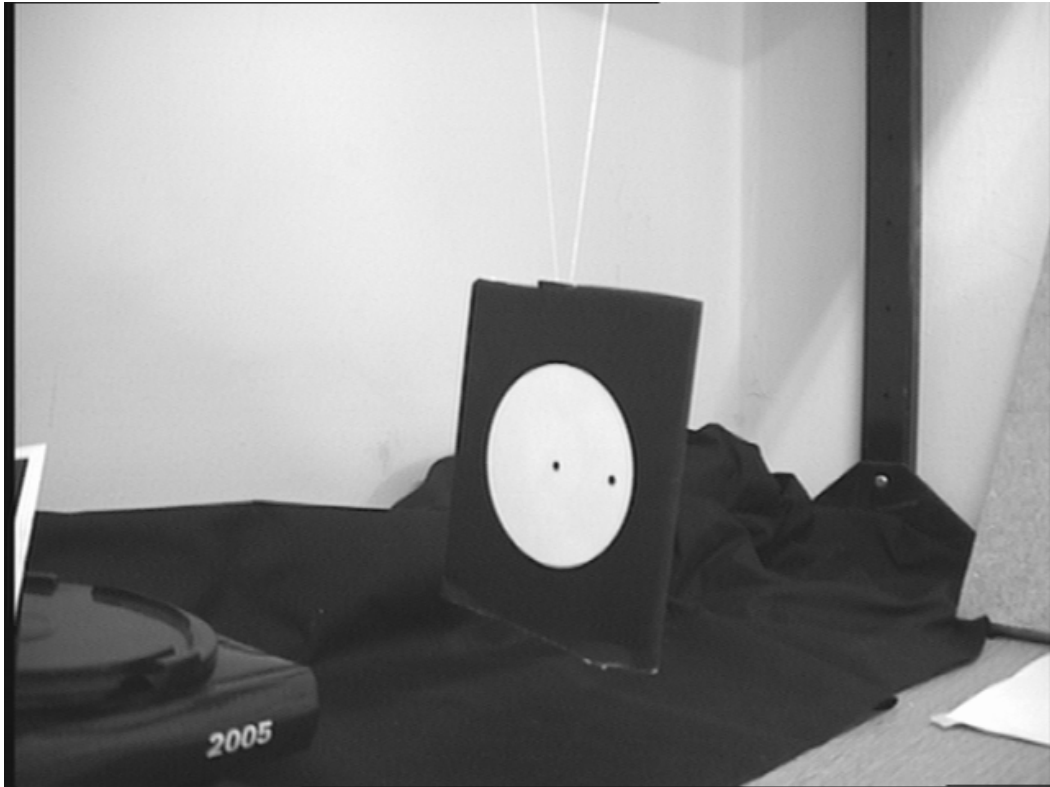
Figure 5.9. Pendulum

The swinging motion of the pendulum is tracked and the 3D configuration parameters and the time vector are extracted. In Figure 5.10, the reconstructed translational parameters are presented. The period of the motion of the pendulum is calculated as the average period of the three translational parameters. The period of a translational parameter is approximated by dividing the duration between the first and last local maximum points by number of cycles. For the data shown in Figure 5.10, the period of the pendulum is measured as 1.338 s.
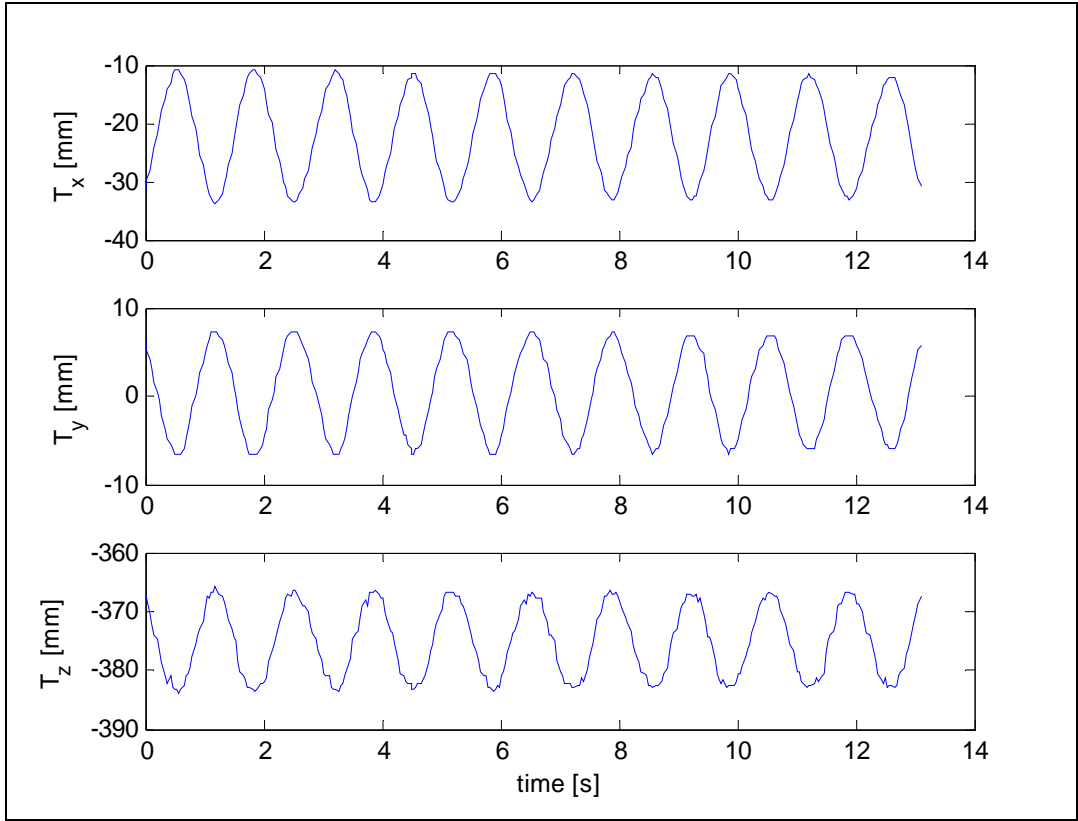
Figure 5.10. Reconstructed Translational Parameters, Swing Motion of Pendulum

The center of rotation and length of the pendulum are calculated using the three translational parameters and surface normal vector of the secondary target. A sphere which can be characterized as

$$r^2 - (x-a)^2 - (y-b)^2 - (z-c)^2 = 0 \tag{5.2}$$

is fitted to the translational parameters with the constraint that the radius vector must be perpendicular to the surface normal vector of the pendulum (target) and can be expressed as,

$$\begin{bmatrix} a-x & b-y & c-z \end{bmatrix} \cdot u_3 = 0 \tag{5.3}$$

110

where r is the radius and (a,b,c) is the center of the sphere, u_3 is the surface normal vector, x, y and z is the translations of the secondary passive target with respect to camera coordinate system.

Arranging the Equations (5.2) and (5.3) yields the following equations,

$$\begin{bmatrix} 2x & 2y & 2z & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ r^2 - a^2 - b^2 - c^2 \end{bmatrix} = x^2 + y^2 + z^2 \qquad (5.4)$$

$$\begin{bmatrix} u_{3x} & u_{3y} & u_{3z} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = u_{3x}x + u_{3y}y + u_{3z}z \qquad (5.5)$$

where, u_{3x}, u_{3y} and u_{3z} are the components of the surface normal vector, u_3, in x, y, and z directions with respect to the camera coordinate system. The following system of equations is obtained for N data points.

$$\begin{bmatrix} 2x_1 & 2y_1 & 2y_1 & 1 \\ u_{3x,1} & u_{3y,1} & u_{3z,1} & 0 \\ 2x_2 & 2y_2 & 2z_2 & 1 \\ u_{3x,2} & u_{3y,2} & u_{3z,2} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 2x_N & 2y_N & 2z_N & 1 \\ u_{3x,N} & u_{3y,N} & u_{3z,N} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ r^2 - a^2 - b^2 - c^2 \end{bmatrix} = \begin{bmatrix} x_1^2 + y_1^2 + z_1^2 \\ u_{3x,1}x_1 + u_{3y,1}y_1 + u_{3z,1}z_1 \\ x_2^2 + y_2^2 + z_2^2 \\ u_{3x,2}x_2 + u_{3y,2}y_2 + u_{3z,2}z_2 \\ \vdots \\ \vdots \\ x_N^2 + y_N^2 + z_N^2 \\ u_{3x,N}x_N + u_{3y,N}y_N + u_{3z,N}z_N \end{bmatrix} \qquad (5.6)$$

where x_i, y_i and z_i are the translations and $u_{3x,i}$, $u_{3y,i}$, and $u_{3z,i}$ are the surface normal vector components of the ith data set. Equation (5.6) is solved by pseudo-inversion and the unknowns, a, b, c and r are obtained.

Unfortunately, the sphere fitting method discussed above is not able to estimate the length of the pendulum for the data presented in Figure 5.10, since the amplitude of the motion is not large enough to get a good spherical fit. On the other hand, increasing the amplitude of the pendulum motion results in frames with interlacing and motion blur effect (Figure 5.11) because the scanning speed of the camera is 8.5 Hz and it is interlaced, although the camera can send 25 frames to the frame grabber per second.
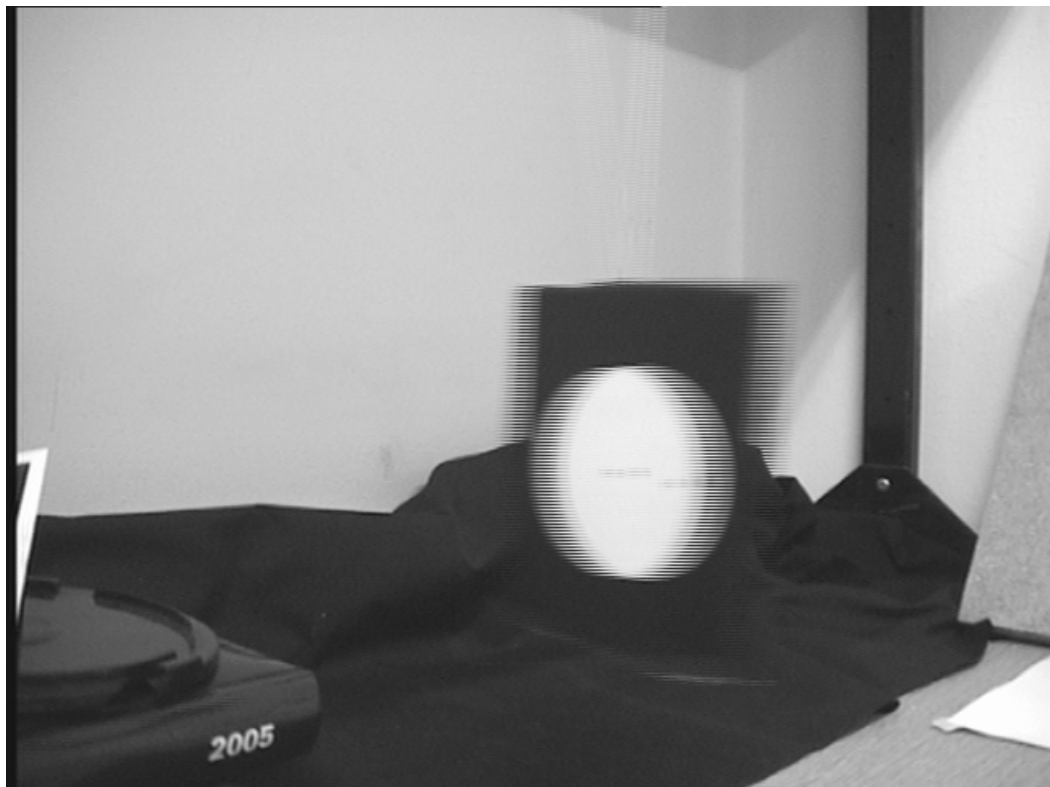
Figure 5.11. Interlacing and Motion Blur.

Although the solution of the interlacing and motion blur effect problem is using a high speed scanning camera, in order the complete the analysis in this study, the problem is solved by using frames in which the pendulum is stopped at five different positions. The reconstructed translational parameters of the target for these frames are presented in Figure 5.12. The sphere fitting method estimated the radius, the length of the pendulum, as 443.71 mm for the data presented in Figure 5.12, where the actual length is 444±0.5 mm.
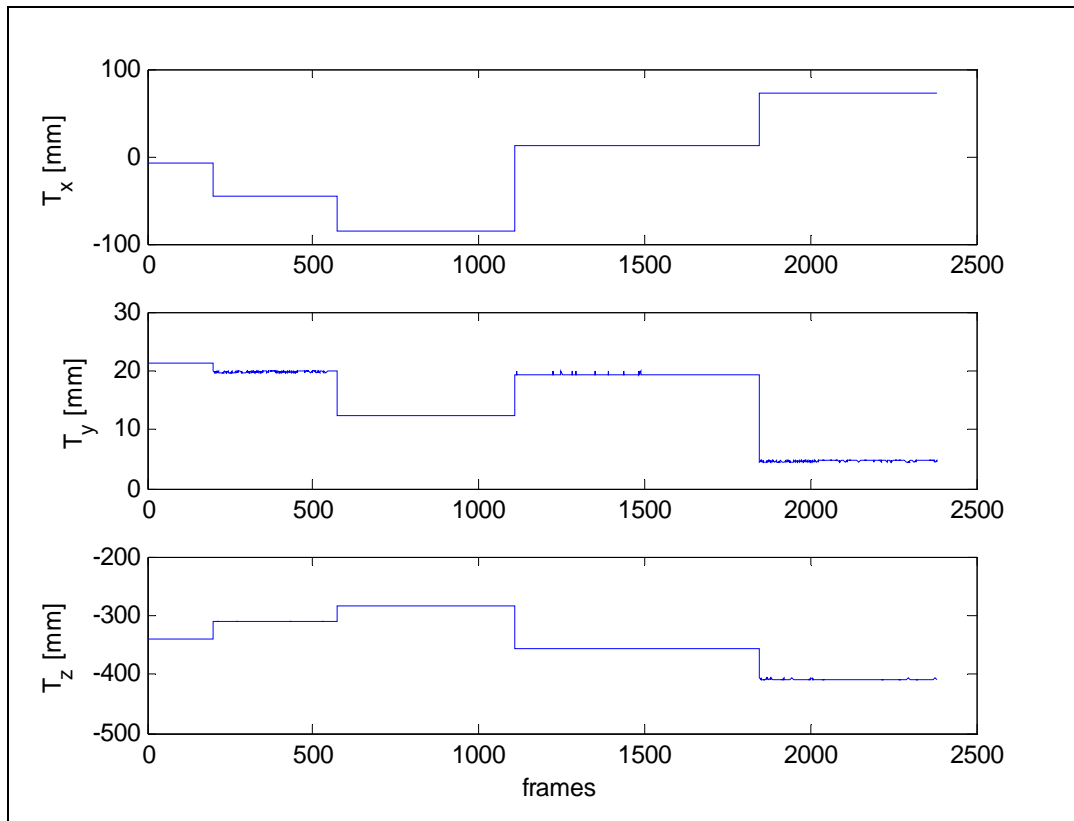
Figure 5.12. Reconstructed Translational Parameters, Positioned Pendulum

The gravitational acceleration is estimated as 9.785m/s^2 using the calculated period and length of the pendulum and Equation (5.1). The experiment is repeated for another four set of camera positions and the results are presented in Table 5.1. The gravitational acceleration is estimated as 9.761 m/s^2 on the average.

Table 5.3. Results of Gravity Calculation.

Data Set	Length [mm]	Period [s]	g [m/s^2]
1	443.71	1.338	9.785
2	443.07	1.339	9.757
3	444.53	1.339	9.788
4	441.52	1.338	9.736
5	443.30	1.340	9.741
Average	**443.23**	**1.338**	**9.761**

The fact that gravitational acceleration calculation results presented in Table 5.1, are slightly smaller than the actual value 9.806 m/s^2 in Ankara [9], mean that the 3D reconstruction method is accurate, since it is expected to obtain underestimated gravitational acceleration values due to assumptions in the period Equation (5.1). Also obtaining similar results for the different camera positions validates the precision of the 3D reconstruction method.

Also this test can be simplified and the precision of the 3D reconstruction algorithm for moving targets can be estimated by investigating the reconstructed rotational parameters R_x and R_y. Note that these parameters determine the orientation of the target plane with respect to camera coordinate system. Since the target plane and the plane of swinging motion coincide for this case, the reconstructed rotational parameters R_x and R_y of the swinging target should be constant over the time. Figure 5.13 represents the reconstructed rotational parameters that determine the target plane orientation with respect to camera coordinate system. As seen in the figure, the orientation of target plane is almost constant for the secondary passive target attached on the swinging pendulum.
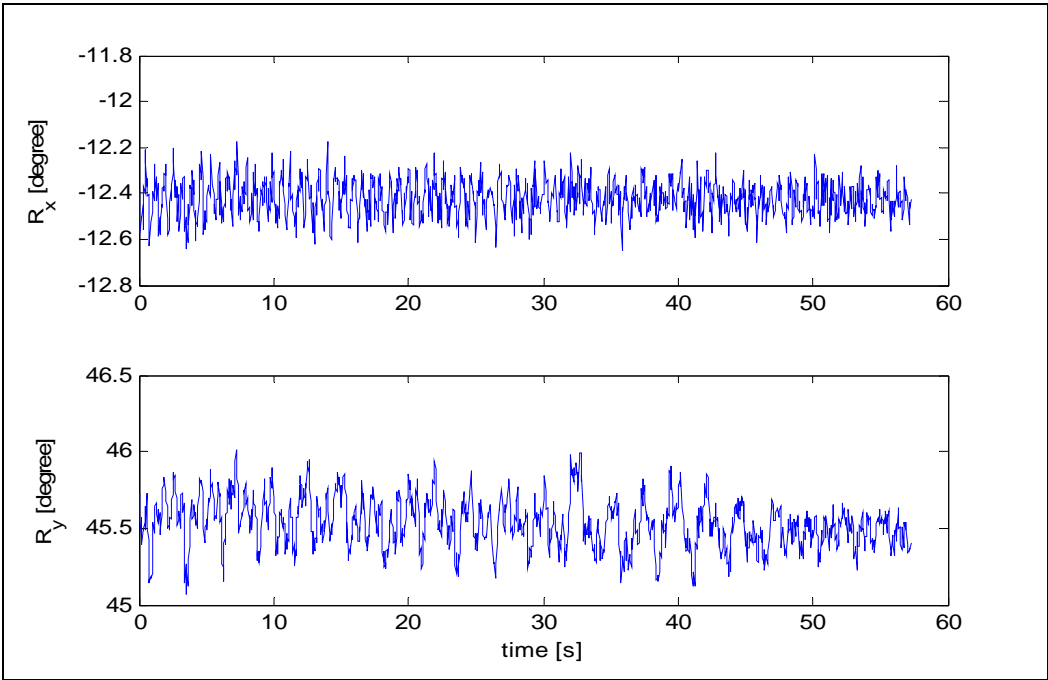

Figure 5.13. Reconstructed Rotational Parameters, Swinging Pendulum

114

5.4. Results

In this chapter, the structure of the overall 3D reconstruction algorithm and the hardware used are presented. Two tests designed to measure the on-site application performance of the setup are explained and the results of these tests are presented and discussed.

The overall 3D reconstruction method is succeeded to detect the target and reconstruct the 3D configuration parameters in all of the tests. However the low scanning speed of the camera results in interlacing and motion blur effects in the frames corresponding to scenes in which the target moves rapidly. For these frames the target can not be detected since the interlacing effect deforms the contour of the target.

The first test designed and applied to measure the accuracy of the 3D reconstruction method is utilized an object with two secondary passive targets attached on its two perpendicular surfaces. In this test, the distance between the centers and angle between the surface normal vectors of the secondary passive targets are calculated using the reconstructed 3D configurations. Experiments are conducted for different camera positions and orientations and the effects of lighting conditions and the distortion correction on the reconstructed parameters are presented and discussed.

The second test utilizes the swinging motion of a pendulum with a secondary passive target attached on its mass center to measure the reconstruction accuracy of the 3D reconstruction method and target tracking performance of the proposed method for moving targets. Unfortunately, the interlacing effect due to scanning process of the camera do not permit tracking and reconstructing the 3D configuration of the swinging pendulum with large amplitudes which is necessary to calculate the length of the pendulum. On the other hand, for small amplitude swinging motions, the target detection and tracking methods are succeeded and the period of the pendulum is calculated using the reconstructed translational parameters over the time. To complete the test, the length of the pendulum is

calculated using the reconstructed 3D configurations corresponding frames in which the pendulum is stopped at larger angular displacements with respect to equilibrium position. The gravitational acceleration is calculated using the length and period of the pendulum. The resulting gravitational acceleration calculated for five different camera positions are slightly smaller than the actual value. Considering the fact that air resistance and moment of inertia are neglected, the reconstructed gravitational acceleration values show that the method is accurate. Also obtaining similar results for different camera positions shows the precision of the overall 3D reconstruction method. The reconstructed rotational parameters that define the orientation of the target plane with respect to camera coordinate system are almost constant as they are expected for a planar motion; this result also validates the precision of the overall method.

CHAPTER 6

SUMMARY AND CONCLUSION

In this chapter, the work done in this study is summarized focusing on the methods proposed, developed, and implemented. Contributions of these methods to the performance of the 3D reconstruction system are evaluated. The major findings of the study are concluded and some recommendations for future work are given.

6.1. Summary

The aim of this study is to develop a robust, fast, autonomous and accurate, 3D reconstruction method utilizing the achievements of the previous studies [1-4]. Naturally, the design objectives of the previous studies, advantages and disadvantages of proposed methods determine this study's approach to the problem. Although the sum of the methods proposed by the previous studies constitutes a fully autonomous 3D reconstruction method, there occurred compatibility, reliability and speed problems when they are combined. The goal of this study is to investigate the previously proposed methods in terms of robustness, speed and accuracy and modify or replace these methods to achieve a 3D reconstruction method satisfying the design requirements.

The processing sequence formed in the previous studies is retained and a step-wise approach is preferred for the evaluation of the previously developed methods. Starting from the image formation and grabbing, all methods proposed are checked, improved, and combined sequentially to avoid compatibility problems. The improved versions of these methods through modification or replacement are used as the primary set while their original versions are kept as alternatives for comparison purposes.

The existing hardware is updated first. A Sony EVI-D31 color CCD camera, Imagenation PXC200A color frame grabber, and a Pentium IV 2.8GHz 512MB DDR ram PC are used because of their availability. Although this hardware supports using color images, grayscale images are grabbed and used considering possible future applications.

Although readily available software platforms like Matlab 6.5® provide some user friendly image acquisition and image processing toolboxes, and therefore are commonly used in machine vision studies, Borland® C++ 6.0 compiler is used for implementation because of speed considerations. A modular approach is utilized in the implementation of methods to provide the necessary flexibility and reliability. All modules and sub-modules are designed to return a success or error code. The same input and output structures are utilized by the alternative methods.

The image capturing techniques implemented in the previous studies [1-4] are updated and utilized. A virtual positioning and perspective projection model is developed for simulation purposes. This model is capable of creating 100 ideal frames of the moving secondary passive target in virtual 3D environment per second. A direct access to the frame grabber's buffer is provided in order to get and process the images of the scene containing the moving secondary target. For comparison of final states of the methods developed with their older versions, a set of digital image file routines are also included.

The segmentation performance of a global threshold method employing the minimum error thresholding technique [21] was experimented and approved in the previous studies [3, 4]. This method is discussed, implemented, and tested considering the objectives of this study. The speed of the method is enhanced 45 times by avoiding the repetitive calculations in the corresponding equations.

The use of a Sobel filter and an edge thinning operation to follow are replaced by a Laplace operator in the edge detection process. So, only one pixel thick edges are

obtained without any deformation caused by the edge thinning operation. Furthermore, the speed of the edge detection process is enhanced 13 times.

The problems in the connectivity analysis proposed by the previous study [4] are solved by removing some faulty implementations. However, this version of the algorithm does not provide any noticeable speed enhancement. Therefore, a new one pass connectivity analysis algorithm is designed and implemented, which is compatible with edge detection method. The speed of the new algorithm is 20 times faster than the original one.

The governing process of the previous study [4] was the ellipse detection. In the previous study [4], the Bennett's algorithm [25] was modified to enhance the detection performance and speed. However the results of the experiments conducted in this study show that this implementation was not optimum in terms of detection success and speed. The use of distinct accumulation spaces is introduced to evaluate each connected path separately. The quality factor and symmetry measure definitions are updated. Decision rules are re-defined using the results of experiments conducted with synthetic images. Although the correct detection rate is increased to 90% and false detection rate is reduced to 10% approximately, the order of the algorithm does not permit any appreciable speed enhancement.

The direct ellipse fitting method proposed in a previous study [26] is discussed and implemented. A relative error criterion is defined and a decision rule is determined utilizing the experiments with synthetic images. Absolutely 100% correct detection and 0% false detection rates are achieved. The results of experiments with virtual frames, real frames and real images show that ellipse detection with the direct ellipse fitting method [26] is approximately 60 times faster than the Bennett's algorithm [25].

The implementation of the 3D reconstruction algorithm is reviewed considering the robustness requirement of this study. The preliminary operations, contour following, and spot detection methods are replaced with newly designed ones. The

3D reconstruction algorithm is rearranged to match the reliability standards applied to the other implemented methods.

The speed of the target detection method composed of segmentation, edge detection, connectivity analysis, and ellipse detection methods is reached up to the frame grabbing speed of 25 Hz, but still highly depends on the complexity of the scene, number of edge pixels, and connected paths. A target tracking method is introduced in order to reduce the size of processed region and thus the processing load, to fix the ellipse detection speed at 25 Hz, and finally to reduce the CPU utilization. Target's position in the processing frame is estimated using either 2D image plane coordinates or 3D configuration parameters of the target extracted for the previous frames. Since the accuracy of the target tracking is not so critical as long as the target remains inside the estimated region, a polynomial fitting technique is preferred because of its low complexity and computational load. Some experiments are conducted for both synthetic frames and real frames including a moving target, with various numbers of last data points and various polynomial orders. The results show that tracking the target using a second degree polynomial fitted to 2D image coordinates of the smallest rectangular region in the last five frames exhibits the best performance if the maximum and average estimation errors are compared. However, the other tracking strategies may also be used to track the target, as long as the estimated region is enlarged enough to handle estimation errors. In the experiments, it is observed that the processing speed is increased to the frame grabbing speed of 25 Hz and the CPU utilization is reduced down to 50-60 % range with the addition of target tracking, independent of complexity of the scene.

The use of a composite radial and tangential distortion model and a method for the determination of model parameters were proposed in the study [4]. However, no suggestions were given on the distortion measurement and distortion correction methods. A calibration pattern including the secondary passive target is redesigned. The image region unaffected or negligibly affected by the lens distortion is determined and used to reconstruct the 3D configuration of the calibration pattern

plane utilizing the secondary passive target on the calibration pattern. An autonomous method is proposed to measure the distortions on the entire image plane. In this method, ideal locations of the calibration spots are calculated and matched with the spots in the image in order to obtain the horizontal and vertical distortion components. Parameters of the distortion model are determined by fitting the model to the measured distortion components. The inverse distortion model is constructed by using Taylor series expansion. The accuracy level of the distortion inversion is obtained using synthetically distorted images. This distortion correction method is combined with the 3D reconstruction method to correct the image coordinates of the pixels corresponding to the target's contour and spots.

Two test procedures are designed to measure the on-site performance of the overall method composed of target detection, 3D reconstruction, and target tracking. In the first test, an object with two secondary passive targets mounted on its perpendicular faces is utilized. The accuracy level of the method is measured by comparing the reconstructed the angle between the faces and distance between the target centers with the actual values. In the second test, the gravitational acceleration is calculated using the motion of a secondary passive target mounted pendulum. The length and period of the pendulum is reconstructed in order to evaluate the gravitational acceleration approximately.

Experiments are conducted with natural and extra lighting conditions, with and without distortion correction, and for different camera positions for the first test procedure. The results show that the accuracy of the overall reconstruction method is higher for extra lighting conditions and distortion corrected cases. However, the autonomous thresholding and distortion correction methods reduce the reconstruction errors below 0.5 mm for distances and 0.5° for angles even for the natural lighting conditions.

In the experiments for second test procedure conducted for different camera positions, the gravitational acceleration is obtained as 9.761m/s^2 on the average. This result is close to its actual value in Ankara and validates the accuracy of the 3D

reconstruction method, if the assumptions and simplifications in the pendulum based gravity analysis are considered (air resistance and moment of inertia are neglected). The results obtained between 9.736 m/s^2 and 9.788 m/s^2 for different camera positions validate of the precision of the 3D reconstruction method.

6.2. Conclusions

The method developed and implemented in this study is capable of detecting the secondary passive target and reconstructing its 3D configuration with a constant processing of 25 Hz. Since this speed is equal to frame grabbing speed, the system can be regarded as it is working in real time. Moreover, the CPU utilization does not exceed % 60, showing that the maximum solution speed of method is even higher.

No runtime errors occurred in any of the experiments including the ones that tests the robustness of the method in extreme conditions such as no lighting, no or more than one secondary passive targets, target with no internal spots, and agile motions of the target or the camera. In the experiments, the method continues to process the subsequent frames although the target is not detected or the method is unable to reconstruct the 3D configuration parameters in the previous frame.

Results of the test cases show that the accuracy of the 3D reconstruction method is enhanced by introducing the distortion correction and autonomous thresholding methods. In the experiments conducted with various camera positions and lighting conditions, it is observed that the method is able to reconstruct the distances and angles with the errors less than 0.5 mm and 0.5° respectively.

This study attains its objectives since the facts given above imply that the developed method in this study is able to reconstruct the 3D configuration parameters of the secondary passive target using single camera images autonomously, in real time with a desired accuracy and robustness.

6.3. Recommendations for Future Work

Almost all machine vision systems developed up to now are application dependent, and this study does not claim the opposite. Thus the essential future improvements may emerge during the adaptation of the proposed method to real life applications. However it is possible to list some general recommendations to increase the performance of the method in terms of its accuracy and speed.

- The accuracy of the 3D reconstruction method can be enhanced by increasing the resolution of the camera and/or employing sub-pixel measurement and/or computation techniques.

- Temporal resolution of the reconstructed 3D configuration parameters can be increased by utilizing a high speed camera.

- Color images and colored target may be used for further reduction of the computational load due to target detection.

- The overall 3D reconstruction method can be embedded into a digital image processing board.

- Pan, tilt and zoom motions of the camera can be controlled by means of servomotors to increase the reconstruction accuracy especially when following moving targets.

REFERENCES

[1]   **Kılınç, İ. A., (1994)**, "Determination of the Position and Orientation of Rigid Bodies by Using Single Camera Images", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[2]   **Acar, E. U., (1995)**, "Experimental Investigation and Implementation of a 3D Configuration Reconstruction Algorithm for an Object Using a Single Camera Image", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[3]   **Özkılıç, S. and Platin B. E., (2002),** "Elliptical Target Recognition for Automation of a 3-D Configuration Reconstruction Algorithm of an Object Using a Single Camera Image", Proceedings of the 6th International Conference on Mechatronic Design and Modeling, September 2002, Cappadocia, Turkey.

[4]   **Özkılıç, S., (2003),** "Performance Improvement of a 3D Configuration Reconstruction Algorithm for an Object Using a Single Camera Image", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[5]   **Arslan, E. F., (1993),** "Design and Development of a Software to Generate Data Files of 2D Objects for Drafting Software Packages", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[6]   **Çokel, E., (1996),** "Development of an Image Analysis and Interpretation Software for Navigation of Mobile Robots", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[7]   **Özdemir, K., (1996),** "Estimation of Mobile Robot Location by Using a Single CCD Camera", M.SC Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[8]   **Çetin, O. L., (2000),** "Improvement of an Image Processing System for Autonomous Mobile Robot Navigation", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[9]   **Shafiq, M. S., (1998),** "Motion Tracking in Gait Analysis", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[10]   **Güler, H. C., (1998),** "Biomedical Modeling of Lower Extremity and Simulation of Foot During Gait", Ph. D. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[11]   **Karpat, Y., (1998),** "Development and Testing of Data Acquisition Tools for a Gait Analysis System", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[12]   **Konukseven, E. İ,, (1996),** "Moving Part Recognition and Automatic Pick and Place Using an Industrial Robot", Ph. D. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[13]   **Şeran, O., (2003),** "Visually Guided Robotic Assembly", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[14]   **Bayraktar, H., (2004),** "Development of a Stereo Vision System for an Industrial Robot", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[15]   **Fidan, U., (2002),** "Design and Construction of a 3D Laser Rangefinder for Indoor Applications", M.Sc. Thesis, Mechanical Engineering Department, Middle East Technical University, Ankara, Turkey.

[16]   **Olgaç, N., Gan, Z. and Platin, B. E., (1989)**, "3D Reconstruction of Object Configurations by Hybrid Projection Analysis Using a Single Camera Image", Proceedings of the 1st National Conference on Applied Mechanisms and Robotics, Vol. 1, Paper No. 89AMR-4C-6, November, Cincinnati, USA.

[17]   **Platin, B. E., Gan, Z. and Olgaç, N., (1990)**, "3D Object Configuration Sensor Utilizing Single Camera Images", ASME Paper No. 90-WA/DSC22, presented at the Winter Annual Meeting, November 25-30, Dallas, USA.

[18]   **Olgaç, N., Craig, P. D. and Platin, B. E., (1991)**, "Advancements in Utilizing Monocular Object Configuration Sensor, Part II", Advances in Instrumentation, DSC-Vol. 30, ASME Book No. H00691, pp. 21-24, also presented at the Winter Annual Meeting, December 1-6, Atlanta, USA.

[19]   **Sankur, B. and Sezgin, M.Y., (2001),** "Image Thresholding Techniques: A Survey over Categories", Journal of Electronic Imaging, Vol. 13, No. 1, pp. 146-165.

[20]   **Alatan, A., (2003),** "EE701 Robot Vision Lecture Notes", Electrical and Electronics Engineering Department, Middle East Technical University, Ankara, Turkey.

[21]   **Kittler, J. and Illingworth, J., (1986)**, "Minimum Error Thresholding", Pattern Recognition, Vol. 19, pp. 41- 47.

[22]   **Sonka, M., Hlavac V. and Boyle, R., (1999),** "Image Processing, Analysis, and Machine Vision", PWS Publishing, California, USA.

[23]   **Jain, R., Kasturi, R. and Schunck, B. G., (1995),** "Machine Vision", McGraw-Hill, New York, USA.

[24]   **Shu, J. S. P., (1989)**, "One-Pixel-Wide Edge Detection", Pattern Recognition, Vol. 22, pp. 41- 47.

[25]   **Bennett, N., Burridge, R. and Saito, N., (1999)**, "A Method to Detect and Characterize Ellipses Using the Hough Transform", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, pp. 652- 657.

[26]   **Fitzgibbon, A. W., Pilu, M. and Fisher, B. F., (1999)**, "Direct Least Square Fitting of Ellipses", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, pp. 476- 480.

[27]   **Weng, Y., Cohen, P. and Herniou, M., (1992),** "Camera Calibration with Distortion Models and Accuracy Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, pp. 965- 980.

[28]   **Altunbaşak, Y., Mersereau, R. M. and Patti, A.J., (2003),** "A Fast Parametric Motion Estimation Algorithm with Illumination and Lens Distortion Correction", IEEE Transactions on Image Processing, Vol. 12, pp. 395- 408.

[29]   **Karras, G. E., Mountrakis, G., Patias, P. and Petsa, E. (1998),** "Modeling Distortion of Super-Wide-Angle Lenses for Architectural and Archaeological Applications", International Archives of Photogrammetry & Remote Sensing, Vol. 32(5), pp. 570-573.

[30]   **Lenz, R. and Tsai, R., (1988)**, "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3D Machine Vision Methodology", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 10, No. 5, September 1988.

[31]   **Mallon, J. and Whelan, P. F., (2004)**, "Precise Radial Un-distortion Images", 17th International Conference on Pattern Recognition, Vol. 1, pp 18-21.

[32]   EVI-D31 Color Video Camera User Manual, Sony Corporation.

[33]   PXC200 Precision Color Frame Grabber Manual, Imagenation Corporation.

APPENDIX A

IMAGEPRO 2.0

Methods proposed in this study and former methods proposed in previous studies are integrated into a software called ImagePro 2.0 to conduct experiments. In Appendix A, general features of the software are explained.

The software composed of two main parts: the interface (Figure A.1) and the background service. The interface makes it possible to change the processing options and parameters as well as to display options without corrupting the operation. In the mean time, the background service part processes the selected input source according to the selected options. In the following sections, the functions of the interface and a user manual are presented.
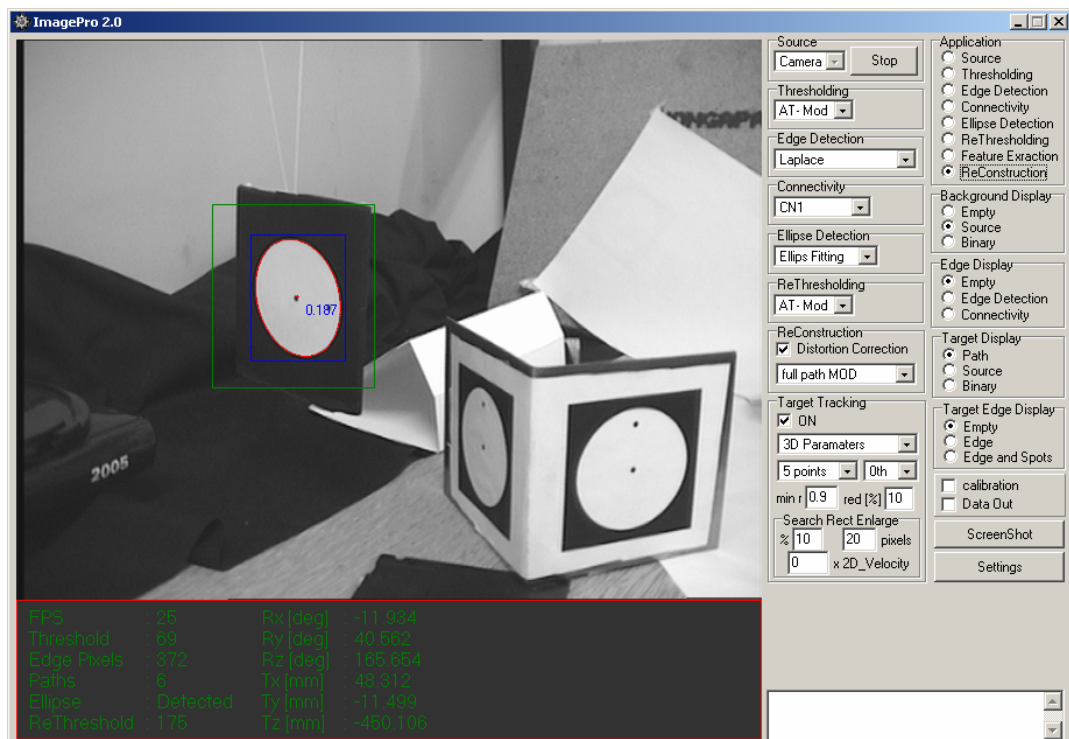
Figure A.1. ImagePro 2.0 Interface

A.1 Interface

The interface of the software is composed of two main areas: the first area is the control panel and the second one is the results area as shown in Figure A.2.
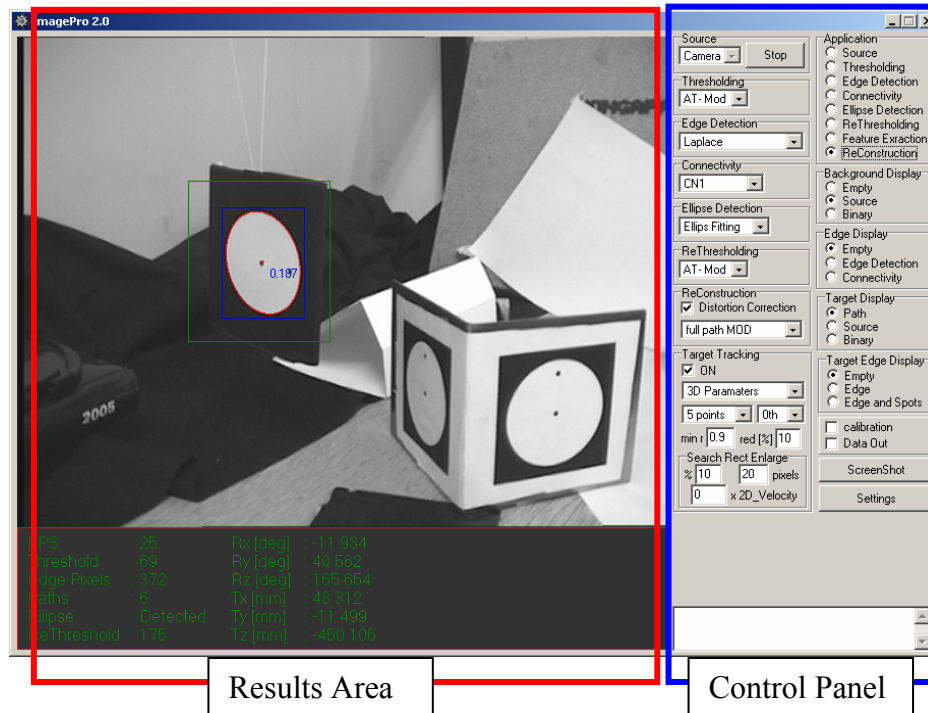

Figure A.2. Parts of the Interface

A.1.1 Control Panel

The control panel is composed of sub-panels (Figure A.3) that enable the user to switch between alternatives methods, to adjust the parameters, and to change display options during an operation.
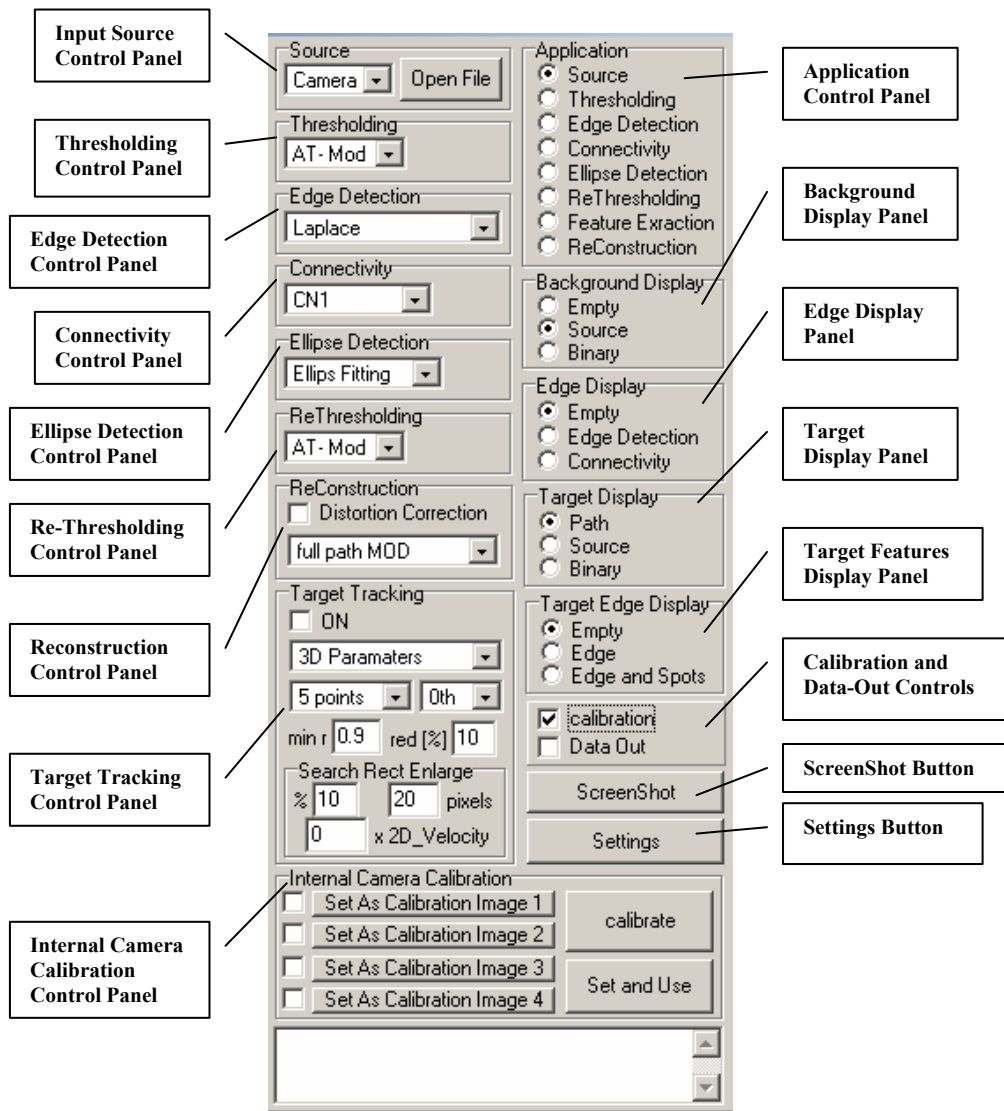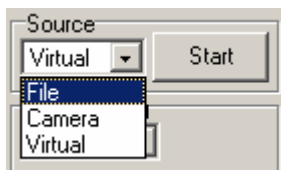
Figure A.3. Control Panel

Input Source Control Panel:

This panel is composed of a pull down menu that enables switching between the input source options and a button that provides starting and stopping the processing for "Camera" and "Virtual" input case and selecting digital image file for the "File" input case.

Thresholding Control Panel:

Thresholding control panel is composed of a list box and a text box. The threshold determination method is selected using the list box. "Manual" option enables using a constant user-defined threshold value entered to the text box "Val". "AT-Org", "AT-Mod" is the Kittler's and Illingwort's minimum error thresholding method with the original and modified equations, respectively, discussed in the segmentation section of Chapter 2.

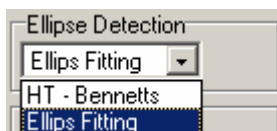Edge Detection Control Panel:

This panel enables switching between the edge detection methods discussed in Chapter 2.

Connectivity Control Panel:
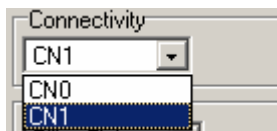
This panel enables switching between the connectivity analysis methods discussed in Chapter 2. CN0 is the corrected version of the proposed method in the previous study [4], CN1 is the connectivity analysis method designed in this study as stated before.

Ellipse Detection Control Panel:

The single list box in this panel provides choosing the ellipse detection method. The first alternative is Bennett's algorithm and second one is direct ellipse fitting.

132

Re-Thresholding Control Panel:
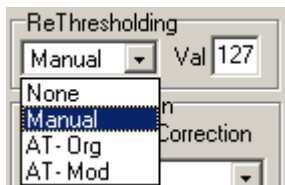
This panel presents the same options with Thresholding Control Panel but the "None" option is added to by-pass the re-thresholding method.

Reconstruction Control Panel:
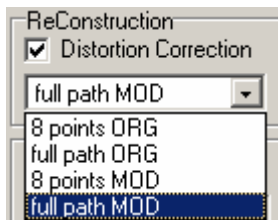
The panel is composed of a check box and a list box. The check box provides enabling or disabling the distortion correction method discussed in Chapter 4, for correcting the image coordinates the pixels corresponding to the target's contour and spots. The list box offers four choices for the pixels used to determine the parameters of the ellipse corresponding to the target's contour. "8 points ORG" and "full path ORG" are the original methods proposed in the previous study [2], while the "8 points MOD" and "full path MOD" are the improved versions respectively in terms of reliability.

Target Tracking Control Panel:
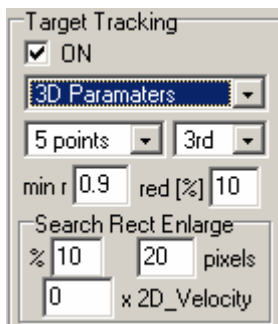
This panel enables the selecting the target tracking strategy and options discussed in Chapter 3. The check box "ON" enables the tracking while processing camera and virtual input source. It is possible to select "3D Parameters" or "2D Corners"

133

tracking strategy, number data points and order of polynomial using the list boxes. The text box "min r" is used to set the minimum value of the correlation coefficient of the polynomial fit below which the degree of polynomial and number of data points are reduced during tracking. "red [%]" text box designed to set the reduction ratio. This number is used to slow down the shrinkage of the estimated target region and guarantee the target is fully inside the estimated region. The other text boxes are for setting the enlargement coefficients discussed in Chapter 3.

Application Control Panel:
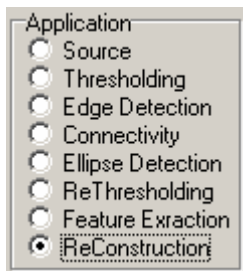
This panel shows the processing sequence from up to bottom. The input source is processed up to the selected level.

Background Display Control Panel
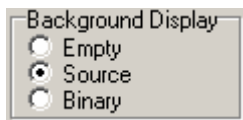
This panel enables setting the image displayed on the background of the results area. If "Empty" option is selected then a black background is displayed. The input source is displayed if the "Source" is selected and the binarized image is displayed if the "Binary" option is selected.
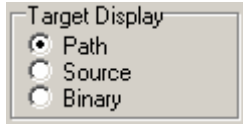
Edge Display Control Panel:

The panel is designed for setting the edge display option. The selection "Edge Detection" will draw the edge pixels determined by the selected edge detection

method on the background image, while the "Connectivity" displays the different colored connected paths.

Target Display Control Panel:

This panel determines how the detected ellipse is displayed in the results area. The "path" option displays the detected ellipse. "Source" option shows grayscale image of the target and "Binary" option displays the binary image of the target region.

Target Features Display Control Panel:
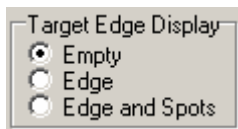
The panel is designed for selecting display option for the extracted features of the target. Edge and Spots of the target is displayed according to selection.

Calibration and Data-Out Controls:
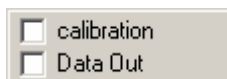
If "Data Out" box is checked "dataout.txt" data file is created and the reconstructed 3D parameters are written into the file. "calibration" check box enables the "Internal Camera Calibration" panel and forces the background service to compute the ideal positions of the calibration spots discussed in Chapter 4.

Internal Camera Calibration Control Panel:
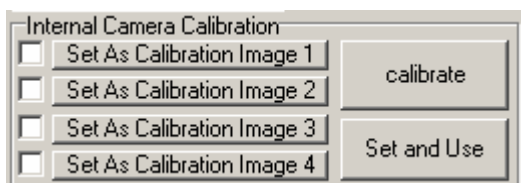
This panel is used to set up to four calibration images using the four buttons on the left part of the panel. The check boxes enable removing the previously set

calibration images. "calibrate" button calculates the distortion parameters and "Set and Use" button associate the distortion parameters to the input source in use.
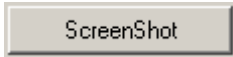
ScreenShot Button:

This button freezes the process and saves the image of the result area to a file with the name entered in the appearing save dialog box.

Settings Button:

This button displays the "Setting" menu for changing parameters corresponding to the input sources.

Settings Menu:

This menu (figure A.3) is designed for setting the dimensions of the secondary passive target, image plane distance, and principal point of the image plane, pixel dimensions corresponding to the input sources. Also, distortion parameters for the file and camera inputs and translational and rotational velocities of the target for virtual input case are set using this menu.
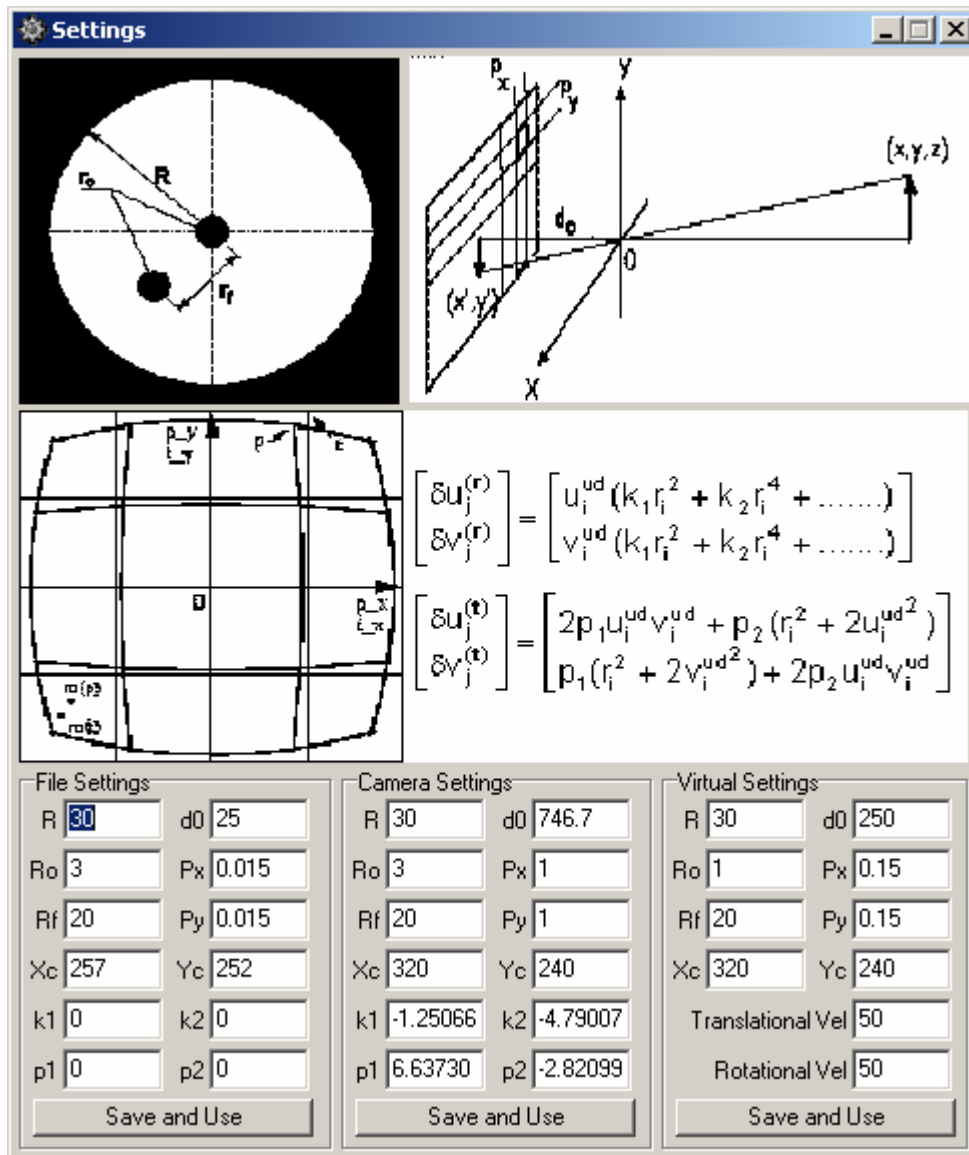
Figure A.4. Settings Menu

A.1.2 Results Area

The results area (figure A.5) is composed two parts. In the first part the image of the scene is displayed according to the selected display options. The green rectangle is the estimated target region by target tracking method and the blue one is the region in which the target is detected. In the second part the results of the methods are presented.

Figure A.5. Results Area

A.2. User Manual

The ImagePro 2.0 software has two operating modes. The first one is 3D reconstruction and the second mode is internal camera calibration. In this section usage of these operating modes are explained.

A.2.1. 3D Reconstruction

It is possible to process the images or frames gathered from the available three input sources: File, Camera and Virtual Perspective Projection.

File Processing:

- Select the "File" option from the "Input Source Control Panel".

- The button inside the "Input Source Control Panel" is automatically renamed as "Open File". Press the button and select the digital image file to be processed.

- Once the file is selected the ImagePro 2.0 software automatically processes the file according the method and display options set through the sub-panels of the "Control Panel" and the presents the results in the "Results Area".

- The process is repeated when a method option or the selected file is changed using the sub-panels of the "Control Panel"

Camera Processing:

- Turn on the camera

- Select the "Camera" option from the "Input Source Control Panel".

- The button inside the "Input Source Control Panel" is automatically renamed as "Start". Press the button to start processing. Once the process is started, the button is renamed as "Stop" and it is used to stop processing.

- Once the process is initiated by pressing the "Start" button, the software processes the frames obtained through frame grabber continuously according to the selected options using the sub-panels of the "Control Panel".

- It is possible to change the processing or display options without stopping the operation.

Virtual Processing:

- Select the "Virtual" option from the "Input Source Control Panel".

- The button inside the "Input Source Control Panel" is automatically renamed as "Start". Press the button to start processing. One the process is started, the button is renamed as "Stop". To stop processing press the button.

- Once the process is initiated by pressing the "Start" button the software processes the frames obtained by virtual perspective projection method

continuously according to the selected options using the sub-panels of the "Control Panel".

- To move the target use the following keys on the keyboard:

4-6	T_x Translation
2-8	T_y Translation
Insert-Delete	T_z Translation
Up-Down	R_x Rotation
Left-Right	R_y Rotation
Home-End	R_z Rotation

- It is possible to change the processing options without stopping the operation.

A.2.2. Internal Camera Calibration

It is possible to determine internal camera parameters using "File" and "Camera" input sources. The process is identical for both cases.

- Use "Start/Stop" button to obtain still images of the calibration pattern for "Camera" input source or "Open File" button for "File" input source.
- Set "Reconstruction" option from the "Application Control Panel" and check calibration check box in the "Calibration and Data-Out Control Panel".
- Press "Set as Calibration Image 1" button in the "Internal Camera Calibration Control Panel".
- Repeat the above steps for other three calibration images (optional).
- Press the "calibrate" button in the "Internal Camera Calibration Control Panel".
- Press the "Set and Use" button in the "Internal Camera Calibration Control Panel" to associate the distortion parameters with the input source in use.

APPENDIX B

SYNTHETIC IMAGE LIBRARY

Appendix B is devoted to the characteristics of images in the synthetic image library. The images in the library are binary and each contains a white elliptical region or a white rectangular region located at the image center. These images are used to measure the performances of the ellipse detection methods discussed in Chapter 2. The width, rotation and aspect ratio of the rectangular images and major length, rotation and aspect ratio of the elliptical images are presented in the following figures and also in Table B.1 and B.2.
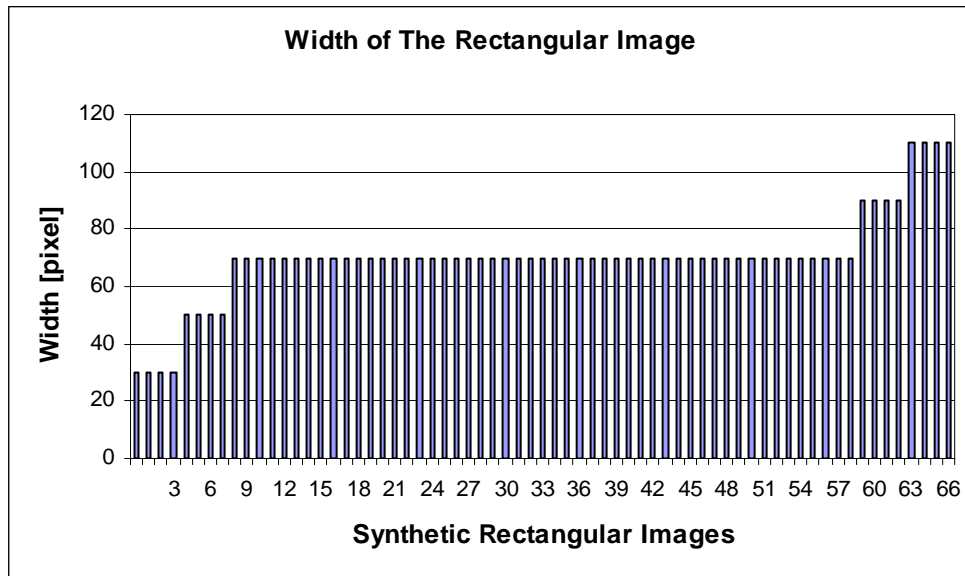
Figure B.1. Width of Synthetic Rectangular Images
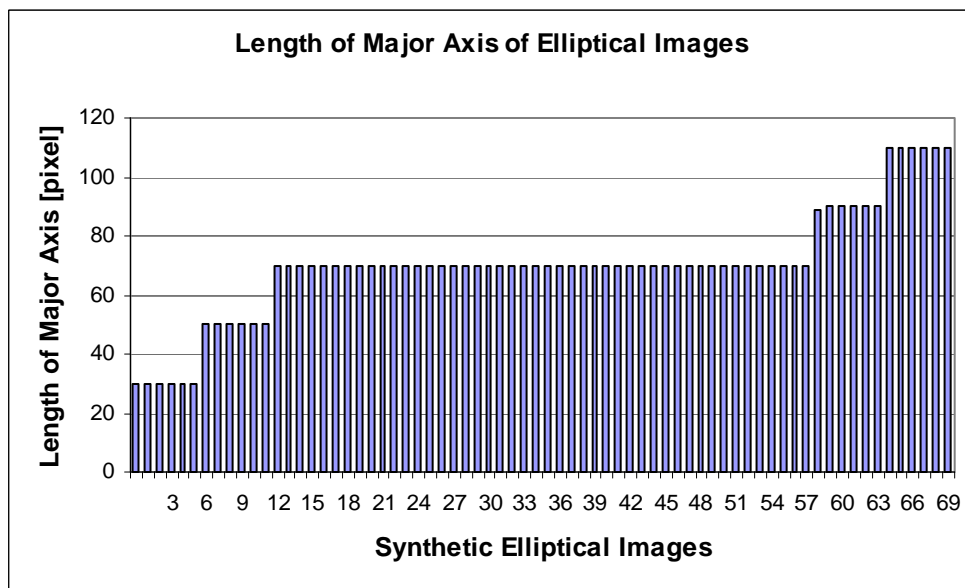

Figure B.2. Major Length of Synthetic Elliptical Images
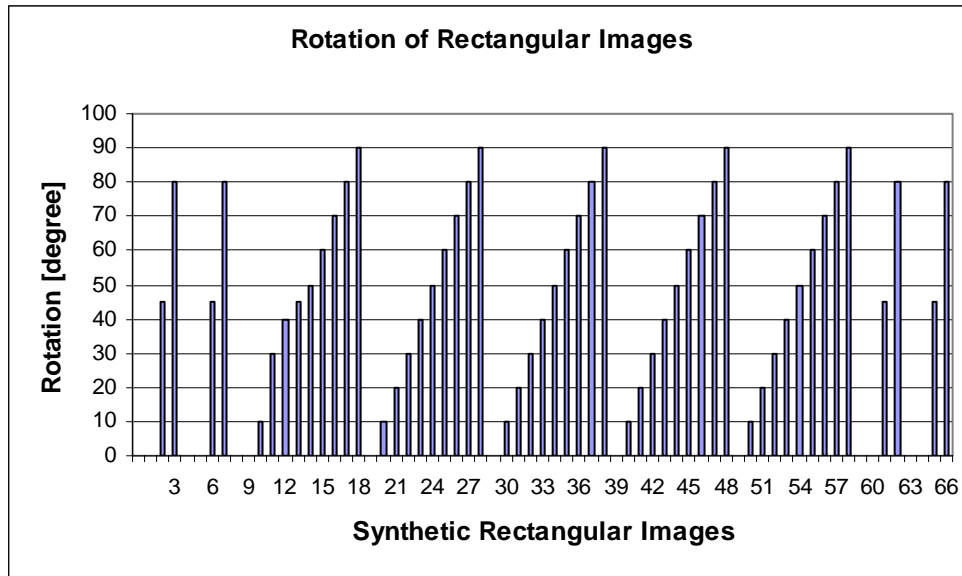
142

Figure B.3. Rotation of Synthetic Rectangular Images
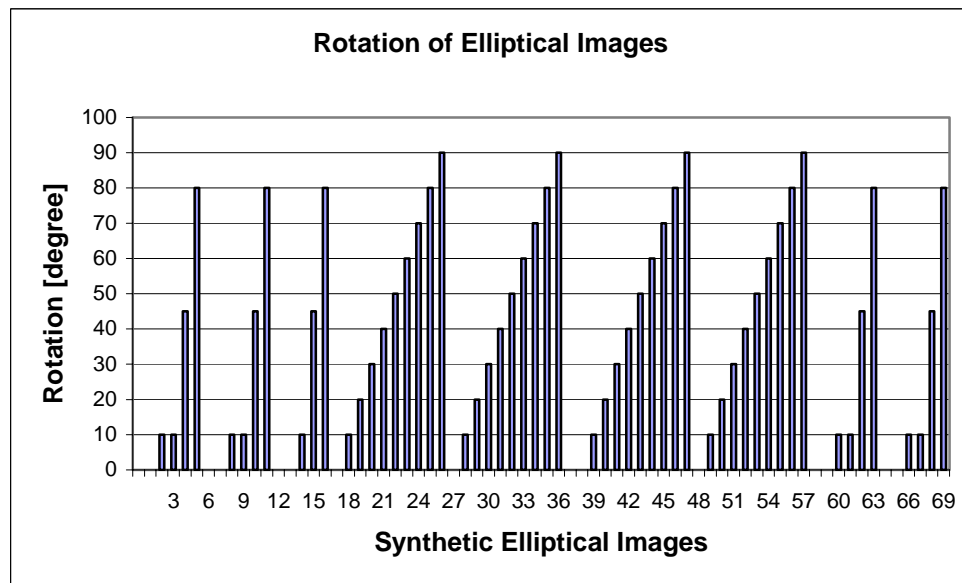

Figure B.4. Rotation of Synthetic Elliptical Images
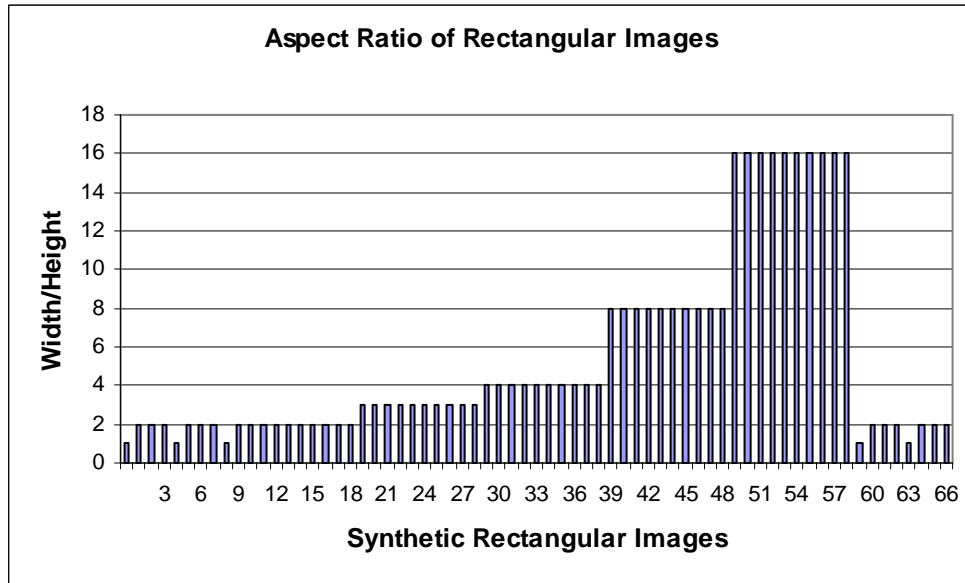
143

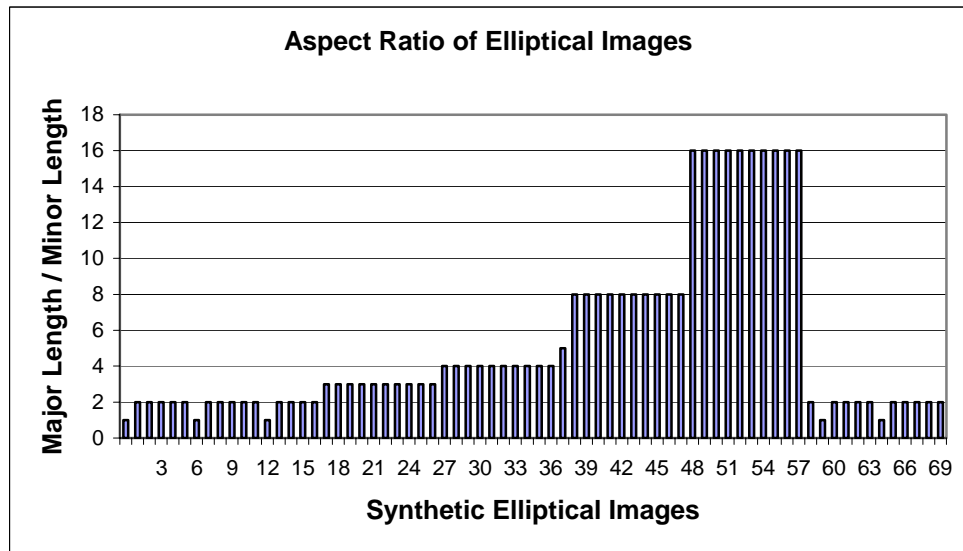**Figure B.5.** Aspect Ratio of Synthetic Rectangular Images

Figure B.6. Aspect Ratio of Synthetic Elliptical Images

144

Table B.1. Characteristics of Synthetic Rectangular Images

Image No	Width [pixel]	Rotation [degree]	Width / Height	Image No	Width [pixel]	Rotation [degree]	Width / Height
1	30	0	1	35	70	50	4
2	30	0	2	36	70	60	4
3	30	45	2	37	70	70	4
4	30	80	2	38	70	80	4
5	50	0	1	39	70	90	4
6	50	0	2	40	70	0	8
7	50	45	2	41	70	10	8
8	50	80	2	42	70	20	8
9	70	0	1	43	70	30	8
10	70	0	2	44	70	40	8
11	70	10	2	45	70	50	8
12	70	30	2	46	70	60	8
13	70	40	2	47	70	70	8
14	70	45	2	48	70	80	8
15	70	50	2	49	70	90	8
16	70	60	2	50	70	0	16
17	70	70	2	51	70	10	16
18	70	80	2	52	70	20	16
19	70	90	2	53	70	30	16
20	70	0	3	54	70	40	16
21	70	10	3	55	70	50	16
22	70	20	3	56	70	60	16
23	70	30	3	57	70	70	16
24	70	40	3	58	70	80	16
25	70	50	3	59	70	90	16
26	70	60	3	60	90	0	1
27	70	70	3	61	90	0	2
28	70	80	3	62	90	45	2
29	70	90	3	63	90	80	2
30	70	0	4	64	110	0	1
31	70	10	4	65	110	0	2
32	70	20	4	66	110	45	2
33	70	30	4	67	110	80	2
34	70	40	4				

Table B.2. Characteristics of Synthetic Elliptical Images

Image No	Major Length [pixel]	Rotation [degree]	Major Length / Minor Length	Image No	Major Length [pixel]	Rotation [degree]	Major Length / Minor Length
1	30	0	1	36	70	80	4
2	30	0	2	37	70	90	4
3	30	10	2	38	70	0	5
4	30	10	2	39	70	0	8
5	30	45	2	40	70	10	8
6	30	80	2	41	70	20	8
7	50	0	1	42	70	30	8
8	50	0	2	43	70	40	8
9	50	10	2	44	70	50	8
10	50	10	2	45	70	60	8
11	50	45	2	46	70	70	8
12	50	80	2	47	70	80	8
13	70	0	1	48	70	90	8
14	70	0	2	49	70	0	16
15	70	10	2	50	70	10	16
16	70	45	2	51	70	20	16
17	70	80	2	52	70	30	16
18	70	0	3	53	70	40	16
19	70	10	3	54	70	50	16
20	70	20	3	55	70	60	16
21	70	30	3	56	70	70	16
22	70	40	3	57	70	80	16
23	70	50	3	58	70	90	16
24	70	60	3	59	89	0	2
25	70	70	3	60	90	0	1
26	70	80	3	61	90	10	2
27	70	90	3	62	90	10	2
28	70	0	4	63	90	45	2
29	70	10	4	64	90	80	2
30	70	20	4	65	110	0	1
31	70	30	4	66	110	0	2
32	70	40	4	67	110	10	2
33	70	50	4	68	110	10	2
34	70	60	4	69	110	45	2
35	70	70	4	70	110	80	2