Reconstruction of Three Dimensional Coordinates of Multiple Targets Using Linear Sensors

Roumen H Georgiev

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of Human Biology

at the Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

April 2003

Abstract

Three dimensional (3D) image processing forms a major component in the methodology of medical imaging and other non-medical applications. It meets the important requirement for establishing the exact position of the human body in 3D space. The use of markers that are attached to different parts of the body, as well as to the instruments used in the diagnostic or treatment procedure, helps to identify the location of the instrument in relation to the body. In addition, medical imaging provides more detailed and helpful information for diagnosis, treatment and clinical follow-up. The main objective of this project, as emphasised in Chapter 1, is to prove the effectiveness of using multiple linear sensors in combination with cylindrical lenses in order to track multiple targets and to reconstruct 3D positions. A literature review of the theory and the main application examples during the years of research and technological development is presented in Chapter 2.

The first major subsection is covered in Chapter 3 and 4, which focuses on the use of one-dimensional sensors. In Chapter 3 comparisons are made between the linear versus the rectangular sensors and the cylindrical versus the spherical lenses. Linear sensors are specifically chosen because they contain substantially fewer elements, compared to the rectangular (2D) sensors. This factor can ensure better communication between the sensor and the hardware, higher sampling rates and simpler signal processing procedures. Linear sensors also provide less redundant data compared to 2D sensors and could be a cheaper technical solution for variety of applications currently based on rectangular sensors. It is suggested that a cylindrical lens be used in combination with the sensor, which converts the light from a point source into a straight line that relates orthogonally to the sensor. According to theory, optical aberrations could affect the exact position on the sensor when spherical lenses are used. The relevant theoretical and practical aspects are carefully studied when a linear sensor is used in combination with a cylindrical lens, thereby necessitating the application of models and tests. The combination of linear sensor with a cylindrical lens severely affects the intensity of the detected signal, especially for passive targets, which is considered a major disadvantage. In Chapter 4, additional signal processing including filtering, noise reduction and a centroid algorithm (peak centre of gravity), contribute to achieving a sub-pixel resolution. This chapter is also dedicated to the data acquisition and signal processing stage, which demonstrates that a resolution better than 0.1 mm is achievable by a linear sensor, thus confirming reports from the literature on this subject. The differences between active and passive targets are also explored in this chapter.

The second major subsection, covered in Chapter 5, deals with the reconstruction of 1D measurements into 3D coordinates. The theory is based on the perspective projection principle, where a point lying in the object space is projected onto an image point through the perspective centre of the lens, according to the "condition of co-linearity". Based on this

principle, two distinct algorithms are derived to use data from 2D cameras, namely Direct Linear Transformation and the Bundle Adjustment Method. These two algorithms are then implemented to reconstruct 3D coordinates from 1D sensor measurements. Further optimisation of the camera parameters in the Bundle Adjustment Method equations significantly improves the performance of the method by achieving sub-millimetre accuracy of the reconstructed coordinates. Experiments with simulated and measured data, are also presented in this chapter. Considering the algorithms for 3D position reconstruction based on 1D sensors it was shown that fewer parameters are needed in the process of calibration and optimisation compared to the case when 2D sensors are used.

The third major subsection in Chapter 6 focuses on tracking multiple active and passive targets. Tracking moving targets in space is usually done by continuously determining their 3D positions. The tracking of multiple targets can be done either individually or simultaneously. Active targets are more convenient because they can be flashed sequentially, unlike passive targets which all reflect the signal at the same time. Where targets are detected with a high level of noise, weighted methods prove to be an effective solution and, in order to solve the correspondence problem, a probabilistic model is suggested for target identification. Subsequently the tracking is accomplished by using a dynamic model for the reconstruction of the trajectory and the calculation of the motion parameters. Another major disadvantage of linear sensors, compared to rectangular sensors, is that they produce more ambiguity when multiple targets need to be identified.

Chapter 7 provides an overview of the fields where 3D tracking systems have been used and are currently being used, for both medical and non-medical applications. In some of these examples, systems based on linear sensors have already been developed but there is still a large potential for the use of 1D sensors as a solution to the trade-off between high level accuracy and performance speed at a reasonable estimated cost.

In conclusion, it has been shown that linear sensors in combination with cylindrical lenses are able to detect and track multiple moving targets with a high level of accuracy. They can work with both passive and active targets and are suitable for a large variety of applications. The 1D data flow compared to that of the 2D sensors is considerably reduced, thus increasing its performance speed. Although the lens distortion and other non-linearities produce large system errors, these can be eliminated following precise mapping procedures of the measured space combined with the optimisation of the camera parameters at the calibration stage. After this stage the accuracy of the reconstructed 3D coordinates of the measured targets will compete with that obtained from the 2D sensors.

Acknowledgments

First and foremost I would like to thank my supervisor Professor C L Vaughan for his valuable guidance and strong support throughout all the stages of this project, for his academic and scientific help during the discussions and his persistence and encouragement to complete every part of the research from the concept development, through various experiments, to the final writing.

Secondly, I would like to thank Mr Patrick Gross who implemented the simulation software "Optical Playground", used to perform some of the tests in the project.

Special thanks to Mr Harry Hall for his technical skills developing the components for the experimental setup.

I would like also to thank Ms Karin Lorenc for editing the text of the thesis.

Many thanks to my present and former colleagues and my friends for encouraging me to finish this research.

Finally, I would like to thank my parents for their distant support and care throughout this project.

Declaration

I Roumen Hristov Georgiev hereby declare that the thesis "Reconstruction of Three Dimensional Coordinates of Multiple Targets Using Linear Sensors" is my own unaided work, both in concept and execution, and apart from the guidance from my supervisor, I have received no assistance except for the construction of the experimental setup and implementation of the simulation software. Neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or any other university. I grant the University of Cape Town free licence to reproduce the above thesis in whole or in part, for the purpose of research.

Date: 14th April 2003 Signed: 

Table of Contents

List of Figures

List of Tables

Chapter One

Introduction

The measuring of objects, including their three dimensional (3D) position as well as their size and shape, has been a challenge to scientists and many years of research have been dedicated to solving this problem. The first attempts at applying these measurements are seen in land mapping, when terrestrial relief shapes are described and also in architecture, when buildings of historic value are restored. In the beginning the classic method of photogrammetry used a stereo pair of photographs from which the 3D points in the object space were calculated by means of mathematical transformations. Later, with the development of modern industry, this work expanded into fields such as robotics, machine vision, space research, medicine and even the entertainment industry.

Three dimensional image processing forms a large component of the methodology for medical imaging. It meets the important requirement for establishing the exact position of the human body in 3D space. Not only does it help the clinician to see the different organs and structures under scrutiny, but it also shows the location of external intervening objects such as instruments, contrast facilities and markers that are utilized in clinical procedures. Medical imaging enables more detailed information to be obtained from the body to help in the diagnosis, treatment and clinical follow-up of patients. Equally important is the creation of large clinical databases as well as the efficient storage and handling of these databases.

The problems facing the traditional methods of imaging - i.e. magnetic resonance imaging (MRI), computed tomography (CT), ultrasound (US), and single photon emission computed tomography (SPECT) - have always challenged biomedical engineers to provide appropriate solutions for diagnosing and treating different diseases. In addition to these imaging applications, various branches of surgery are especially interested in the precise 3D localization of the place of intervention, the targeted location and the methods used to reach that level. Image-guided surgery and neurosurgery use tracking systems to utilize the access of the instrument to the lesion by navigating its tip and displaying it on a computer screen overlaid on CT or MRI images. In recent years the term Minimally Invasive Surgery (MIS) has been introduced where surgeons, operating through a relatively small incision, need to know the exact position and orientation of their instruments. Although endoscopes are being widely used for internal imaging, a tracking system can considerably enhance the ease of the procedure.

In all these cases the accuracy and spatial resolution of the measurement have been key factors in providing enough detail and ensuring the usefulness of the method. Video cameras have been widely used for a variety of applications and, as early as the 1970s and 1980s,

computers became an inseparable part in the configuration of such systems. During the last two decades the level of technology has advanced rapidly with the introduction of new image processing techniques as well as algorithms, resulting in the development of new instruments. This has made it possible to increase the number of cameras used in the systems, which thus provide a higher spatial resolution and performance speed.

In order to measure points in 3D space continuously, at least two video cameras are needed to capture images of a given object. Different points of interest of the relevant object are projected onto the camera sensor and their horizontal and vertical positions, known as image coordinates, are measured. The image coordinate values are then passed through a reconstruction algorithm which calculates the space XYZ coordinates. This is a simplified description of the procedure of 3D coordinate reconstruction. This procedure can be repeated for numerous camera captures whereby any change in the position of the points can be registered. The latter describes a process known in theory as target tracking.

The task becomes more complex when the measured target moves in space. In addition to the spatial resolution, the speed of the sequential captures becomes another important factor. It also requires that the data transfer speed and the calculation of the target positions are sufficient to provide an accurate temporal resolution of the motion. Standard video cameras are unable to provide an efficient solution for fast motion tasks such as object tracking in sport events and in various other industrial applications. This must be taken into account as well as the data transfer speed and the calculation speed of the other components, irrespective of whether or not they are computers or any other type of hardware. To provide good technical characteristics, the researchers often have to come to a compromise because spatial and temporal resolution are linked by an inverse proportional relationship. High resolution cameras have a greater number of elements (pixels) which affects the processing speed as more data needs to be handled.

Another problem arises from the fact that there may be multiple targets to be tracked, not merely single ones. In such instances there are two possible ways of solving the problem. Firstly, the hardware components can be made more complex, e.g. increasing the number of cameras, a solution used by many commercial manufacturers. Secondly, a solution is found by adding more control functionality to the objects, e.g. frequency or time-division identification (TDI) is used in conjunction with an external control system which sends signals to the object. It requires wiring or uses radio frequency (RF) or optical link control. Colour identification is another option, useful only with colour sensors or colour filters, which may also need additional signal processing capabilities.

Detecting systems based on ultrasound or a magnetic field can use both ways of multiple target identification: simultaneous and sequential. Their targets are active and can be

controlled by modulated signals or can be activated in sequence. When detection is based on cameras, the identification approach to be followed depends on the nature of the targets. With active targets such as light emitting diodes (LEDs), working in the infrared (IR) or visible light spectral range, the sequential approach applies. However, when passive targets made of reflective material are being tracked, only the simultaneous detection approach will be effective because all targets reflect at the same time.

The second difference between the systems is related to the visibility of the source from the detector. The optical and the ultrasound detectors require a direct line-of-sight view of the target so that the light or sound waves can propagate without interference by interposed objects. The method which uses a magnetic field does not need a direct target view but it can be severely affected by external ambient magnetic fields or the presence of ferromagnetic materials. On the other hand, ultrasound devices are sensitive to temperature and humidity changes and optical sensors can pick up the reflection from materials which are not subjected to tracking.

From the various problems listed above one can focus attention on the reason why one dimensional (1D) sensors may be preferred to two dimensional (2D) rectangular cameras for some applications. This study will be concerned mainly with the problem of tracking multiple targets by linear Charge Coupled Device (CCD) sensors in real time, thereby expanding the objectives into several other areas. These main objectives of the project can therefore be formulated as follows:

1. To explore the literature related to tracking multiple targets in 3D space;
2. To compare the linear versus rectangular sensors and the cylindrical versus spherical lenses by considering the major theoretical and practical aspects of their use;
3. To measure the distortion of a cylindrical lens using simulation and physical measurements and to suggest a procedure for its compensation;
4. To find out the existing advantages and disadvantages when using either active or passive targets and to measure the resolution and positions of both types;
5. To analyse and test different algorithms for reconstructing 3D positions based on 1D sensor measurements using simulated and real data from different targets' arrangements;
6. To explore methods for target identification and tracking;
7. To support the theoretical considerations and concepts with the results obtained from the simulations and the tests; and
8. To review the future possibilities for using one-dimensional sensors for some medical applications.

9. To combine the outcomes from the above objectives in an original research work in both theoretical and experimental aspects.

The difference between linear and rectangular CCD sensors lies in the number of pixels they each contain. The pixels of the linear type are positioned along a line which is referred to as a 1D sensor, while the pixels of the rectangular type form an array which is referred to as a 2D sensor. In most cases the total number of CCD elements in the 1D sensor are fewer, but compared to each direction of the 2D sensor, it can have a significantly higher number of pixels. The standard size of a rectangular CCD could be 512 x 512, 640 x 480 or 768 x 576 which means that there are between 250 000 and 450 000 pixels in total on the chip. Modern technology has recently been used to design arrays with up to several thousand pixels on each side. Linear sensors typically range between 512 and 10 000 pixels along the line and may consist of several lines, the pixels of which are always lower in total compared to the rectangular CCD. If one byte of digital data represents one pixel, the difference in the data flow to be processed for each of the two types of sensors can be appreciated. Linear sensors have greater spatial resolution and they are usually cheaper in price. One specific detail about using these CCDs to track a target in 3D space is that they require a cylindrical instead of a spherical lens, which converts the light from a point source into a straight line.

The procedure of finding the target coordinates in 3D can be divided into two major subsections: data acquisition and coordinate reconstruction, each consisting of various stages which have a specific impact on the final result. These stages define the structure of the project.

The initial stage in the data acquisition for a 3D tracking system based on linear sensors is related to physics when the use of a cylindrical lens in combination with the sensor is considered. This type of lens converts the light from a point source into a straight line and therefore the sensor should be positioned orthogonaly to the lens in order to intercept the light (Figure 1.1).

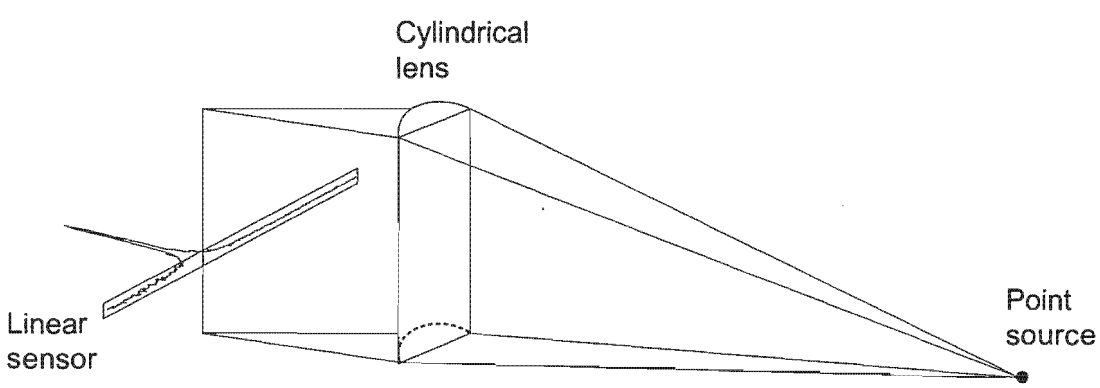


Figure 1.1 Light conversion by a cylindrical lens from a point source into a line orthogonal to the linear sensor showing the peak formed from the active pixels.

Similar to the spherical lens, all optical effects influence the position of the image on the sensor. Distortion, which is one of the five achromatic aberrations, is considered an important factor that could change the exact position on the sensor. It is therefore essential that this research determines the character of the distortion that emanates from a cylindrical lens and finds an effective way to compensate for this distortion. The problem to be addressed here will be how to focus the light and also obtain a good signal-to-noise ratio.

Collecting data from the sensor is another part of the acquisition process. The emphasis will fall on obtaining a reading of the output signal after the light has activated an area of adjacent pixels and has formed a peak. In conjunction with the problem of noise reduction from the ambient light, this section will also compare the difference in intensity between active and passive targets. Another obstacle that could arise when using multiple targets is how to distinguish them individually, especially in the case of two closely situated points. For this purpose general signal processing tools could be very useful and additional algorithms could help find an appropriate solution.

The second major subsection treats the coordinate reconstruction problem. Using algorithms already developed for rectangular cameras, a new system of equations for 1D sensors is suggested. The equations include various parameters, which form part of the calculation process and play an important role in achieving a good spatial resolution. These parameters are obtained in a step called the "calibration step" where known data from various points in 3D space provide the necessary values for the sensors. In the calibration and measurement steps, the use of more than three sensors is recommended, which is the minimum number of equations in the system. This creates a redundancy, which improves the accuracy and also helps when the view to a certain target is obscured from one or more sensors. Redundancy in the number of equations transforms the system into an over-determined problem, when the solution can be found by simply using the Least Squares Method (Hager, 1988). This method, which exists in matrix form, is a convenient and speedy approach for solving problems with multiple solutions.

A combination of three 1D sensors is the minimal configuration for a system using 3D space measurements (Dainis, 1980; Macellari, 1983). In this project the initial number of four linear sensors completing a single device has been chosen and further optimisation of this number presents another objective. Increasing the number of the sensors changes the behaviour of the system parameters, thereby affecting their accuracy and spatial resolution. One method described by Abdel-Aziz and Karara (1971), called Direct Linear Transformation, has been successfully applied to rectangular cameras and could also be converted for use by linear sensors. Another more recent approach is used in the so-called Bundle Adjustment (Gruen, 1996). From this, a system of non-linear equations can be derived and used in the calibration step.

With regard to moving targets,these are tracked by means of a series of reconstructions of their (spatial) coordinates, whereupon the trajectory for each target can be drawn by connecting the corresponding temporal positions. From these data some time-varying parameters, such as velocities and accelerations, can be calculated. If multiple targets are involved in the process, the major problem arises from intercepting trajectories and solving the correspondence problem for the different targets. A variety of techniques, using position prediction and statistical methods, are available from the literature on the subject. Most common are the ones developed for radar applications in marine and aircraft research (Bar-Shalom *et al.*, 1990).

During the last number of years, the performance of tracking systems improved considerably but there has not been much change at a conceptual level. The classic methods have been adapted for the new technology and new applications for tracking multiple moving targets have emerged. The general characteristics of some of the experimental and commercial systems based on linear sensors are summarised in Table 1.

Author (Manufacturer)	Model	Year	Sensors (pixels)	Type of markers	Max. N markers	Light convertor	Resolution	Sampling rate
Experimental systems								
Mitchelson		1971	3	IR laser		Doublet of cylindrical lenses	0.5 mm	1 kHz
Leo and Macellari		1981	2x2 (1728)	IR LEDs	8 per side	Toroidal lens	1.2 mm	2.9 kHz
Macellari		1983	2x3 (2048)	IR LEDs	8 per side	Toroidal lens	4.4 mm	9 kHz
Bianchi et al.		1990	3 (2048)	IR LEDs	50	Toroidal lens		2.5 kHz
Dainis		1990	3 or 4			Cylindrical lens		
Owen		1995	3 (2048)	IR LEDs		Cylindrical lens		2 kHz
Commercial systems								
Optotrak (Northern Digital, Canada)		1990	3 (2048)	IR LEDs	256	Cylindrical/ Toroidal lens	0.15 mm	5 kHz
Coda mpx30 (Charnwood Dynamics,		1998	3	IR LEDs	28	Bar-code mask	0.6 mm	5.6 kHz
Pyxsis Flashpoint (Image Guided Technology, USA)		1995	3	IR LEDs		Cylindrical/ Toroidal lens	0.5 mm	
Optical Tracking System (Radionics, USA)			3	IR LEDs		Cylindrical/ Toroidal lens		

Table 1.1 Some examples of experimental and commercial tracking systems based on linear sensors.

The present project uses these achievements as well as suggests some new techniques and models. As part of the project, the theoretical considerations, as described above, are followed by extensive experimental work. The main contribution could be found in the various tests performed throughout of the research to verify and prove the ideas and concepts and to minimise the error from the measurements and the calculations. The spatial resolution is

compared to that obtained from other similar tracking systems or systems based on different principles such as ultrasound and magnetic fields. Software simulations are also used to complete the picture of the "cylindrical lens – linear CCD" combination. This is done in all physical and signal processing aspects before an effort towards constructing a prototype is undertaken.

Linear sensors are preferred to the rectangular sensors, especially in the cases: when the tracking system uses active targets, as they are easier to detect; and if the speed performance should be a key factor in the cost-effect trade off. In addition their 1D resolution superiority could be a contribution to more accurate 3D position determination. Their use, however, could be limited for tracking simultaneous targets, capable of executing fast maneuvers, because of difficulties in the identification process.

Chapter Two

Literature Review

Cameras have been applied for many years to track different objects in three dimensional (3D) space. The first examples were used in the robotic vision and machine intelligence fields, and subsequent applications have been designed for medical purposes. With the advance of technology these systems have become more sophisticated and currently all are based on the latest computer platforms, although not much has changed in their main structural components. Apart from the multiple applications that are used in robotics such as tracking systems for robotic arms and manipulators, safety systems, instruments for process control and machine vision, a wide variety of applications are now available in biomechanics, radiotherapy and surgery.

The classic combination of two video cameras has been taken from stereo photogrammetry and successfully adapted to solve the problem of object tracking of moving rigid structures as well as of moving flexible structures. In such cases markers, visible to the cameras and describing the most important lines and parts of its shape, are attached to the object, and then geometrically related to each other to follow accurately the motion of the object. In ongoing research it has also been observed that when more cameras are focused on the object from different angles, the accuracy is considerably improved and the problem of ambiguity avoided when distinguishing separate targets. This is also useful when a large volume or long distances have to be monitored as in the example of some sport related activities.

Target tracking has been done not only by optical methods with cameras, but also using ultrasound and magnetic field devices (Bucholz and Smith, 1993; Leotta *et al.*, 1997). The first technique transmits an ultrasound signal from the source target, which is then detected by microphones set at certain positions. The distance to each microphone is measured and the target position in three-dimensional space is calculated. The second technique is based on detecting the magnetic induction from a source target by the number of coils in the space at known positions and then, using the value of the generated current, the distance from the detecting coil to the target is calculated. These two methods have a similar degree of accuracy, which is acceptable for various applications, although the optical method using cameras as described above could provide an accuracy of at least one order of magnitude higher.

As the terms *resolution* and *accuracy* are widely used in all literature sources, they will be clarified for this project to be consistent with the terminology used by different authors. *Linear resolution* will be understood as the capability of a linear sensor to distinguish between two close positions of the same target in space. If the target has moved and two samples from its position have produced images on the sensor, they will be represented by their pixel values,

so that each of them will correspond to a certain distance in the object space. For a sensor consisting of 2048 pixels elements the resolution would be 1:2048. In the instance of an additional centroid algorithm, a sub-pixel resolution can be achieved, thereby improving this value to 1/10 of a pixel, and increasing the linear resolution to about 1:20 000.

Accuracy is another characteristic of the different systems for 3D measurements. This will be related to the standard deviation of the error after the reconstruction of multiple targets and will be given in millimetres. There might also be other measurements of the accuracy of such as mean, absolute mean or absolute standard deviation, which are all related to the system error. Allard *et al.*, (1995) suggested that it could also be expressed through the object-to-base distance for consistency when experiments are carried out at different ranges.

2.1 Algorithm development for 3D coordinate reconstruction using rectangular CCD sensors

The algorithms for camera calibration and for the calculation of the position in 3D space form a substantial part of the data processing in photogrammetry. One such approach is referred to as the "normal case of photogrammetry" which uses the triangulation technique (Marzan and Karara, 1976).

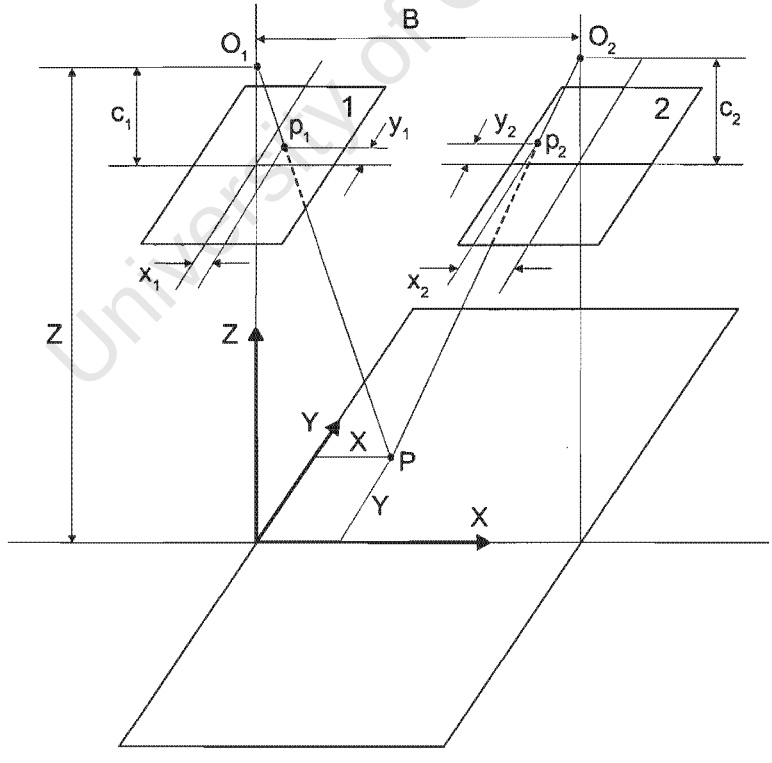


Figure 2.1 The normal case of photogrammetry representing a point P projected onto two photographic planes at O_1 and O_2 (not shown on the picture) where the projected lines pass through another two planes located at distances c_1 and c_2 called principal distances. The two planes pass through the centres of the camera lenses and contain the principal points (Marzan and Karara, 1976).

Figure 2.1 shows the configuration of two projection image planes and an object at point P. Using similar triangles, a relationship can be found to calculate the XYZ coordinates of P. The first can be expressed as follows:

$$x_1 = \frac{X}{Z} c_1$$

$$x_2 = \frac{X - B}{Z} c_2$$

$$y_1 = \frac{Y}{Z} c_1 \qquad [2.1]$$

$$y_2 = \frac{Y}{Z} c_2$$

where

x_1, y_1	=	coordinates of the point P on the photographic plane 1
x_2, y_2	=	coordinates of the point P on the photographic plane 2
X, Y, Z	=	object space coordinates of P
B	=	base distance between the image planes
c_1, c_2	=	principal distances of the left and the right image planes

From equations [2.1], dividing the first by the second equation we obtain

$$X = B \frac{c_2 x_1}{c_2 x_1 - c_1 x_2} \qquad [2.2]$$

Further, adding the last two equations yields

$$Y = Z \frac{y_1 + y_2}{c_1 + c_2} \qquad [2.3]$$

and subtracting [2.3] from [2.2] gives

$$Z = X \frac{c_1 - c_2}{x_1 - x_2} + B \frac{c_2}{x_1 - x_2} \qquad [2.4]$$

In the general case $c_1 = c_2 = c$ and thus equations [2.1] become

$$x_1 = \frac{X}{Z} c$$

$$x_2 = \frac{X - B}{Z} c$$

$$y_1 = \frac{Y}{Z} c \qquad [2.5]$$

$$y_2 = \frac{Y}{Z} c$$

and from there it can be found

$$X = B \frac{x_1}{x_1 - x_2}$$

$$Y = Z \frac{y_1 + y_2}{2c} \qquad [2.6]$$

$$Z = B \frac{c}{x_1 - x_2}$$

This algorithm describes a perfect projection to the image planes and does not take into account the different non-linearities due to optical effects and lens distortion that normally occur with real cameras.

To calibrate a real system of cameras, two general methods have been implemented. The iterative method was applied to the nonlinear system where a solution is obtained after applying a well-known linearization to the base system of equations (Karara, 1980). The second method is the Direct Linear Transformation (DLT) method described by Abdel-Aziz and Karara (1971). It uses the basic projective relationship:

$$\begin{bmatrix} x - x_p \\ y - y_p \\ -c \end{bmatrix} = \lambda \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X - X_o \\ Y - Y_o \\ Z - Z_o \end{bmatrix} \qquad [2.7]$$

which can be rewritten as

$$x + \Delta x - x_p = -c_x \frac{r_{11}(X - X_o) + r_{21}(Y - Y_o) + r_{31}(Z - Z_o)}{r_{13}(X - X_o) + r_{23}(Y - Y_o) + r_{33}(Z - Z_o)}$$

$$y + \Delta y - y_p = -c_y \frac{r_{12}(X - X_o) + r_{22}(Y - Y_o) + r_{32}(Z - Z_o)}{r_{13}(X - X_o) + r_{23}(Y - Y_o) + r_{33}(Z - Z_o)} \qquad [2.8]$$

where

x_p, y_p	=	coordinates of the principal point
x, y	=	coordinates of the object on the image
X, Y, Z	=	space coordinates of the object
r_{ij}	=	orthogonal rotation matrix
c_x, c_y	=	principal distances of the system
λ	=	scale factor
X_o, Y_o, Z_o	=	coordinates of the perspective centre
$\Delta x, \Delta y$	=	systematic errors of the principal point coordinates

After simplifying [2.8] it yields to

$$x + \Delta x = \frac{L_1 X_i + L_2 Y_i + L_3 Z_i + L_4}{L_9 X_i + L_{10} Y_i + L_{11} Z_i + 1}$$

$$y + \Delta y = \frac{L_5 X_i + L_6 Y_i + L_7 Z_i + L_8}{L_9 X_i + L_{10} Y_i + L_{11} Z_i + 1} \qquad [2.9]$$

which is known as the basic formula for DLT (Marzan and Karara, 1976). To solve the equations [2.9] for $L_1 - L_{11}$, a system of at least 11 equations is needed. If each camera provides a pair of x,y data values then a minimum of 6 points in the object space with known X,Y,Z coordinates has to be measured. As long as the DLT is considered to be a linear solution to a nonlinear problem, then some approximations should be applied. One such approximation is to consider the parameters $L_1 - L_{11}$ in [2.9] to be independent. In this case, if an exact solution is required, two of the parameters need to be eliminated which leads to the system of nonlinear equations [2.8]. Another approach is to introduce additional constraint conditions (Bopp and Krauss, 1978):

$$(L_1^2 + L_2^2 + L_3^2) - (L_5^2 + L_6^2 + L_7^2) +$$

$$+ \frac{(L_1 L_9 + L_2 L_{10} + L_3 L_{11})^2 - (L_5 L_9 + L_6 L_{10} + L_7 L_{11})^2}{L_9^2 + L_{10}^2 + L_{11}^2} = 0 \qquad [2.10]$$

$$(L_1L_5 + L_2L_6 + L_3L_7) - \frac{(L_1L_9 + L_2L_{10} + L_3L_{11})(L_5L_9 + L_6L_{10} + L_7L_{11})}{L_9^2 + L_{10}^2 + L_{11}^2} = 0 \qquad [2.11]$$

These constraints have to be further incorporated into the system to be used in the least squares algorithm.

The examples so far have considered images from two cameras because a single camera does not provide sufficient data for calibration. This is also known as a "stereomodel" in photogrammetry and is widely used in a variety of applications. However, if sufficient redundancy is needed to measure more complex objects, more than two cameras must be used (Gruen, 1996). Expanded with more equations the system becomes over determined. The Least Squares Method in a matrix form is one convenient way in which to find an approximate solution and achieve a high level of accuracy.

2.2 System error and distortion

The accuracy in the calibration process, as well as the calculation process to determine 3D position, is determined by the system error. This error consists of a two part systematic error and a random error (Marzan and Karara, 1976). The systematic error comes from the instrument and its physical characteristics and can be described mathematically. The random error is due to instrumental interferences or the influence of the operator. The magnitude of the system error is considered acceptable when the first component is compensated for and the second component is minimized. The systematic component includes non-linearities of the instrument as well as lens distortion. The distortion itself contains a symmetrical part, which is the major part of the radial distortion, and has been best described by the formula:

$$\Delta r = K_1r^3 + K_2r^5 + K_3r^7 \ldots \qquad [2.12]$$

where

$$r^2 = (x - x_p)^2 + (y - y_p)^2$$

x, y	=	measured coordinates on the image
x_p, y_p	=	coordinates of the principal point
K_1, K_2, K_3	=	symmetric distortion coefficients

For some lenses it is also found that the distortion results in a better description by using all the terms in the polynomial:

$$\Delta r = K_1 r + K_2 r^2 + K_3 r^3 + K_4 r^4 + K_5 r^5 + \ldots \qquad [2.13]$$

However, for the majority of the commercial types of lenses, investigators have shown that the polynomial [2.12] is sufficient and in this sum the term $K_1 r^3$ is the most significant.

The mathematical model to account for the asymmetrical or decentering lens distortion was developed by Conrady (1919), further modified by Brown (1966) and cited in Marzan and Karara (1976). It includes the rest of the radial and the tangential (axial) distortion. For most of the close range applications it can be expressed as follows:

$$\begin{aligned} \Delta x &= p_1(r^2 + 2x_s^2) + 2p_2 x_s y_s \\ \Delta y &= p_2(r^2 + 2y_s^2) + 2p_1 x_s y_s \end{aligned} \qquad [2.14]$$

where

$$\begin{aligned} r^2 &= x_s^2 - y_s^2 \\ x_s &= x - x_p \\ y_s &= y - y_p \end{aligned}$$

x, y are the image coordinates, x_p, y_p are the coordinates of the principal point and p_1 and p_2 are the coefficients of the asymmetrical lens distortion. According to some other researchers, for close range applications the variation of the decentering distortion within the photographic field appears to be too small to be of practical significance (Brown, 1971).

2.3 Concept development for the use of linear sensors for 3D measurements

A more recent approach for camera-based systems was the application of linear instead of rectangular sensors. The first publications taking into account the definition of the true position of targets in space using the linear detection of transmitted light appeared in the early 1970s. In 1971 Mitchelson proposed a design for an optical instrument that could simultaneously measure Cartesian coordinates in 3D space. On the basis of this instrument, Mitchelson went further to develop a method called Cartesian Opto-electronic Dynamic Anthropometer CODA (Mitchelson, 1975). It used the light transmitted from gallium arsenide laser sources that were infrared (IR) and detected by an optical system of three cameras lying in a straight line. The cameras were designed in such a way that they could detect only one direction of the moving target. The two side cameras were positioned about 1 m apart and they were sensitive to a horizontal direction change. The third camera in the middle was sensitive to

changes of the target position related to the vertical direction (Figure 2.2).

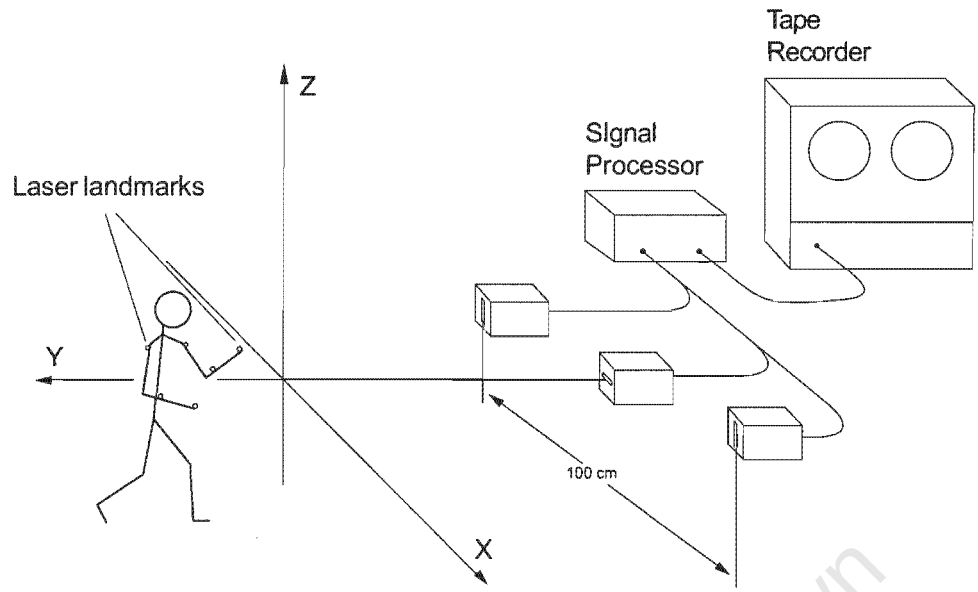


Figure 2.2 Physical arrangement of the cameras in relation to the subject in the system developed by Mitchelson (1975).

For this purpose a doublet of cylindrical lenses was placed in front of each camera and by converting the light from the point target into a straight line, the doublet projected the line onto the photo detector behind, which was placed at a focal distance as shown in Figure 2.3. The line was projected over the entire field of the camera's receptive area, which is covered by a mask of transparent and opaque zones corresponding to the digital Grey code. This technique, together with additional analogue-to-digital conversion, gave an output equivalent to a 12 bit resolution of the instrument. A single camera device could have a field of view of 26° in a horizontal direction and approximately the same in the vertical direction, resulting in the covered volume of space having a solid trapezoidal shape.

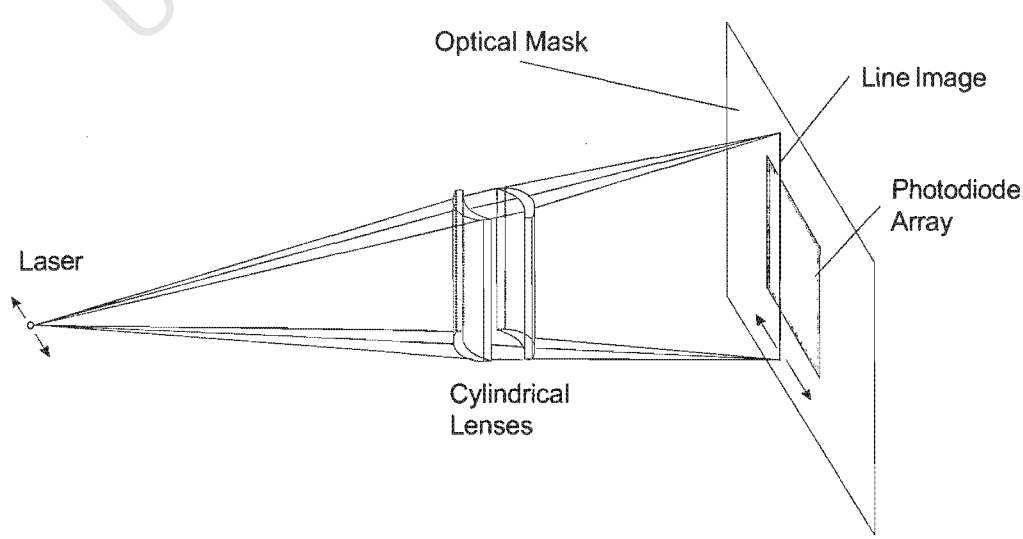


Figure 2.3 Two cylindrical lenses convert the light from a point source and project it onto the camera (Mitchelson, 1975).

The actual field of view for the camera, at a distance of 3 m in this case, was about 1.5 m and the reported spatial resolution was 0.5 mm. The system of cylindrical lenses, as it was designed, had a combined focal length of about 20 mm and width (aperture stop) of 4 mm. This geometry data, together with the photodetector size of 10 mm, defined the field of view of 26° mentioned above. Apparently the best linearity that could be obtained by this system was approximately 0.1%. In order to calculate the space coordinates from the digitized camera values, a technique using the proportionality of the distances between the measured target, the lens centre and the image on the photodetector was described. When this was applied to the two outside detectors it gave the distances in the X and Y directions following the same principle as the human eye. The third detector in the middle, sensitive to the vertical direction, provided the Z direction of the coordinates.

As the main applications for 3D measurements are closely related to the principles of human movement, by the mid 1970s and early 1980s the existing video and TV equipment could ensure frame rates of up to 50 – 60 frames/second. These frames represented a 2 dimensional picture of the object space and contained a large amount of unnecessary data to be handled at each sampling period. Using linear sensors for direct detection of the prominent markers could provide far higher frequencies with very good resolution. Similar to the instrument above described by Mitchelson, another concept was reported using linear CCD sensors by Reymond and Hidalgo (1980) in an application to determine direction using three sequentialy flashed targets. Later, Leo and Macellari (1981) used a system for tracking human movement based on the concept represented in Figure 2.4 and Figure 2.5.

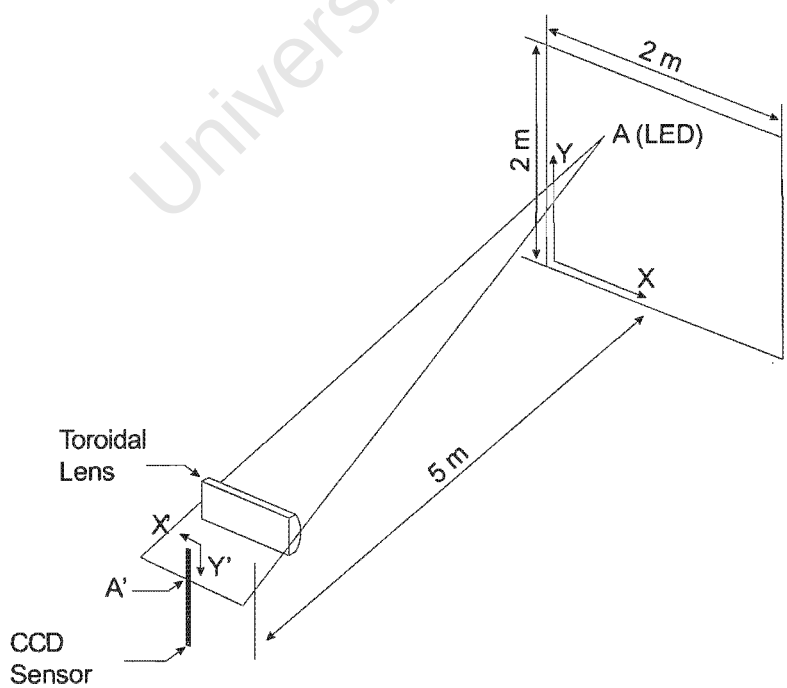


Figure 2.4 The positioning of a toroidal lens and CCD sensor in 3D space (Leo and Macellari, 1981).

The image sensor (Fairchild CCD 121H) consisting of 1728 sensitive elements was positioned in the image plane, parallel to the fromt surface of a toroidal lens (+5 diopters vertical, +15 diopters horizontal power), which converted the light into a line perpendicular to the longitudinal axis of the sensor. The principal idea of this design was to reduce the amount of incoming data and to increase the speed of the marker registration. The system used the 11-bit digitized output from the sensor and, at a clock frequency of 5 MHz, could provide a maximum sensor sarnpling rate of 2.9 kHz and a spatial resolution of 1.2 mm. The largest error on the sides of the measured field was reported to be about 10 mm before compensation. The markers that were used for this application were infrared light emitting diodes (CQX19 Telefunken) in order to avoid interference from the ambient light. The radiated output power of 0.4 W in pulse mode made detection easy at a distance of 5 m. In this case they were sequentially fired for easier identification from a remote controlled timing circuit, so that no wiring was linked to the subject. For this configuration a total of 16 markers, 8 per side, were attached, which was considered to be the minimum for the description of the whole body motion. The detecting system was organized with two sensing devices each consisting of 2 linear sensors (Figure 2.5).

The entire data acquisition module was then connected to a microcomputer. This allowed the system to track 8 markers in the object space while operating at 359 Hz, which was found to be reasonable by the study group compared to the standard of 100 Hz typically used for biomechanical studies (Figure 2.5).

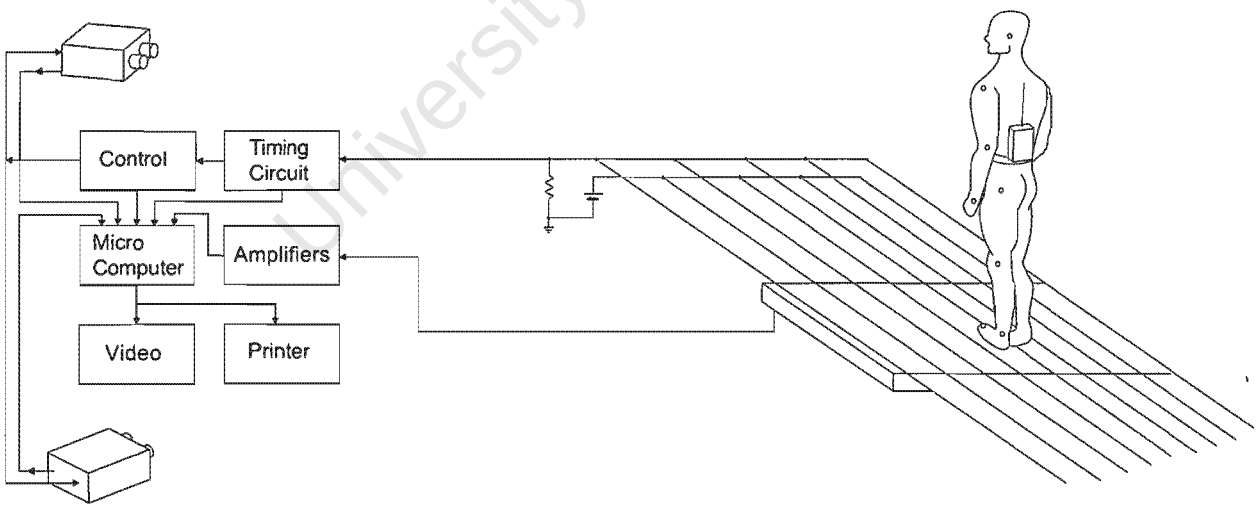


Figure 2.5 Two sensitive units consisting of two linear sensors each for tracking human motion (Leo and Macellari, 1981).

Later in the 1980s linear sensors were again suggested for applications in the tracking of human movement (Macellari, 1983). The system design in this example was changed to use

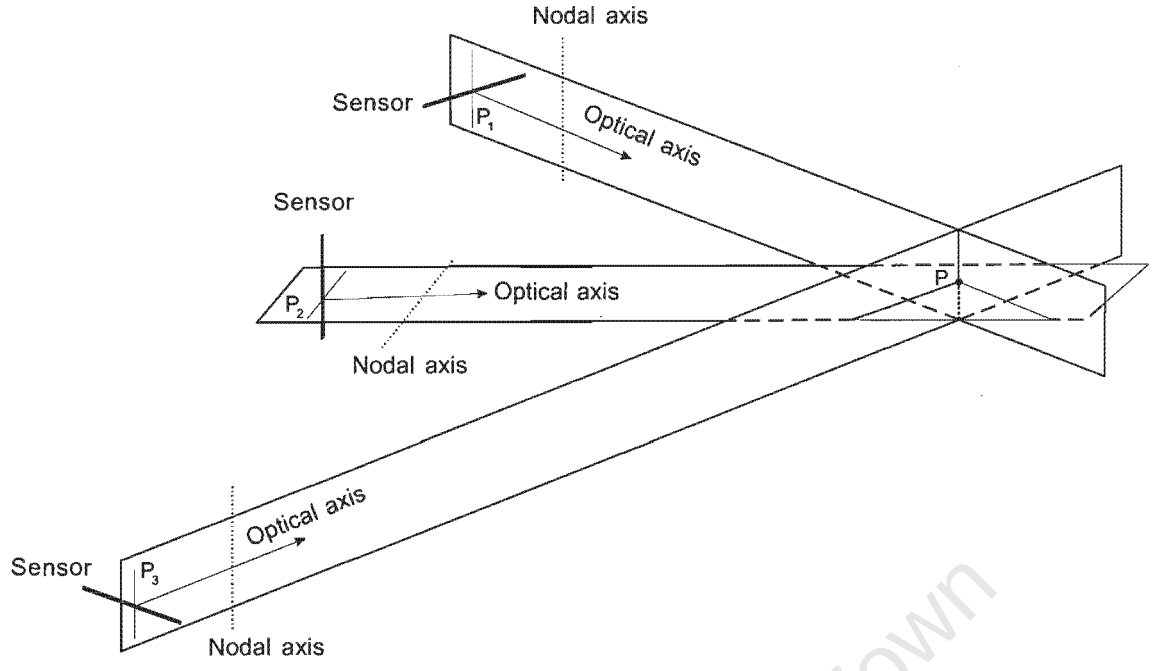


Figure 2.6 Possible arrangement of three 1 dimensional sensors into a 3 dimensional unit. The intersections P_1, P_2, P_3 defining the three planes which intersect at the other side in the point P (Macellari, 1983).

3 linear sensors incorporated into a unit, shown in Figure 2.6, that was arranged similarly to the system described by Mitchelson (1975). The minimum number of image coordinates needed to determine the location of a point in 3D space is three. This design had the advantage of speedier automatic acquisition and occupied less memory than the previous system of four linear sensors.

The lines along the vertical axes of each lens, in combination with point P where the source is positioned, form three planes in such a way that they intersect at point P. When these lines are extended further they hit the sensors behind and form the peak described above, from where the pixel value of each sensor source is determined. If it is assumed the lens is at a known location in the space, a plane can be drawn, either by knowing the point where the source is positioned, or by knowing the position of its pixel value on the sensor. In the case above there are three sensors with cylindrical lenses and the points P_1, P_2, P_3 of the source. From there it can be found out where the position of the source target P is by using the geometrical relation of the intersecting planes already explained. It is also possible to solve the inverse problem by knowing the exact position of the source and the lenses in order to predict the pixel values for each sensor.

By increasing the number of pixels up to 2048 after a more recent sensor (Fairchild CCD

143) was used, the linear resolution was improved up to 1/4000 (Macellari, 1983). This number was calculated after the smallest distance was defined to be 6.5 μm, or half the space between two adjacent pixels. This corresponded to a linear interval of 0.6 mm at a distance of 5 m. Targets that had been used previously were again the same type LEDs (CQX19 Telefunken), sequentially fired for easy identification. The maximum sampling frequency for these sensors was 9 kHz, which resulted in a 1.1 kHz sampling rate per target when 8 targets per side of the subject were tracked. The system consisted of two units for space tracking on either side of the subject, *i.e.* six linear sensors and a separate control unit to activate the markers through a frequency modulated (FM) signal transmitted from the main unit to a receiver positioned on the back of the subject. The reported results showed an improved level of accuracy with maximum errors for the x and z direction (which was the plane parallel to the plane of the sensors) of 1.7 mm and for the y direction (which was normal to that plane, or sometimes called the image depth) a maximum error of 4.4 mm on the side of the field of view. Lens distortion was pointed out as playing a significant role (Macellari, 1983) as a possible reason for these errors together with the system non-linearities of utilising a physical origin.

The same research group presented another example of human motion study based on linear sensors a few years later (Bianchi *et al.*, 1990). Active markers such as IR LEDs were used as in the past. In this application 3 cameras were attached to a vertical pole with side cameras separated by a distance of 2.1 m and the other camera positioned in the middle, as is shown in Figures 2.7 and 2.8.

Figure 2.7 Positioning of the camera in the stereometric unit used by Bianchi *et al.* (1990).

Figure 2.8 Outlook of a single camera from the stereometric unit.

Sequential identification of the targets was applied. The resolution was increased to 1/16000 based on the analogue-to-digital conversion which had been extended to 14 bits. The speed of the system was limited to 2.5 kHz, which defined a certain number of markers that could be detected at a given sampling rate *i.e.* from 5 markers at 500 samples per second, up to 50 markers at 50 samples per second.

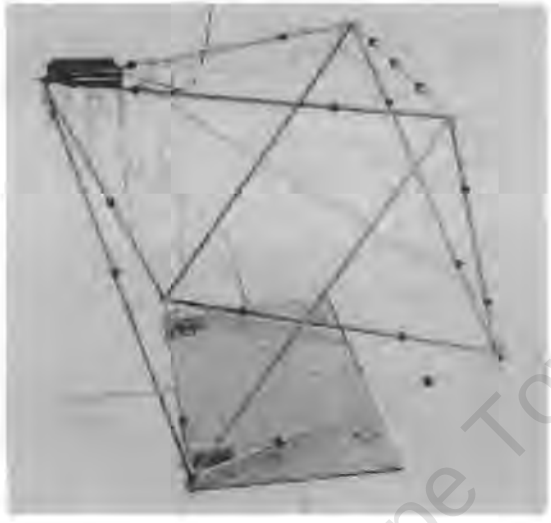


Figure 2.9 Calibration frame with 20 control markers (Bianchi *et al.*, 1990).

The algorithms to reconstruct the 3 dimensional X,Y,Z coordinates were based on the formulae for Direct Linear Transformation (Abdel-Aziz and Karara, 1971) and an additional polynomial function was used to model the lens distortion. The system needed initial calibration to determine the camera parameters by measuring 20 points in the object space with known X,Y,Z coordinates (Figure 2.9). The difference in the calibration step was that instead of a pair of equations used for 2 dimensional cameras with 16 parameters, the method used was of one equation for each linear sensor with 7 parameters. For this procedure optical axes of the superior and inferior cameras were directed to form a convergent angle of 25°, which differed from the normal 90° angle considered to be optimal for such tests (Bianchi *et al.*, 1990).

Three or more linear sensors can be used to identify a single target in a space although 3 sensors are not sufficient to identify multiple targets simultaneously (Dainis, 1990). The two configurations suggested in the patent by Dainis (1990) represent detectors for measuring positions consisting of 3 and 4 linear sensors respectively. In the first example (Figure 2.10), sensors are positioned in a single plane orientated at 120° to one another, a so-called "triaxial linear sensor camera".

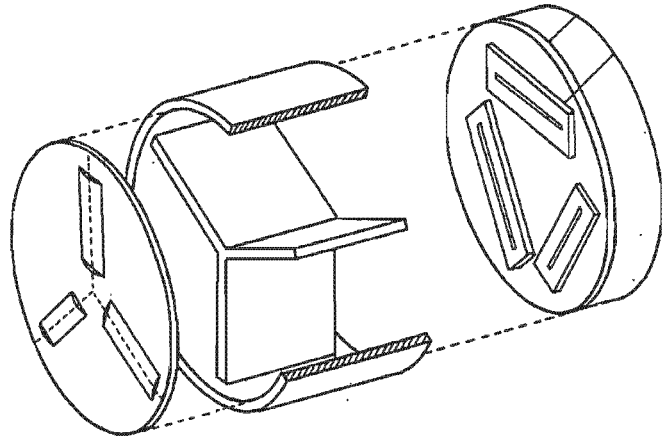


Figure 2.10 Position and orientation of the lenses and sensors in the "triaxial linear sensor camera" (Dainis, 1990).

Each sensor has a cylindrical lens in front of it, orthogonal to its longitudinal axis (Figure 2.11). An additional separator was placed between the lenses and the sensors to ensure that each sensor only received a signal from its matching lens (*cf.* Figure 2.10).

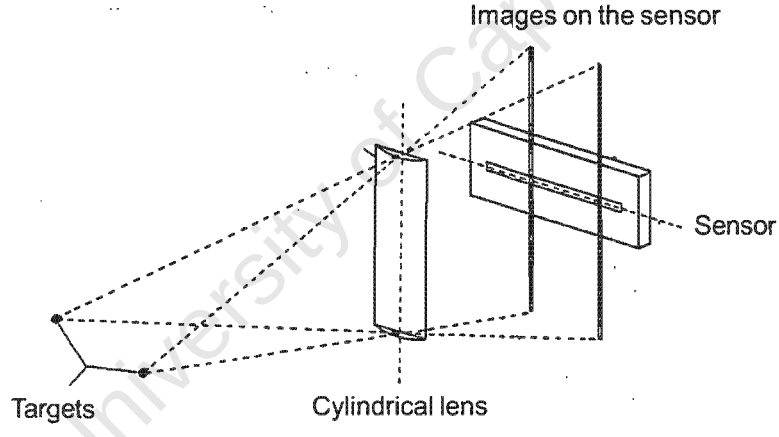


Figure 2.11 Images on the linear sensor received from targets at different positions (Dainis, 1990).

When a target is positioned in the space in front of this detecting device there is a plane formed by the optical axis of each lens and the target itself. Thus two of the planes intersect in a line containing the target. The third plane also intersects this line exactly at the point where the target is. If the planes are extended towards the sensors behind, they will intersect at another three points on each sensor. These points represent the sensor values and are measured in pixels from the beginning of the sensor line. From these values their planes can be reconstructed by using the axes of the lenses, thereby making it possible to find their intersection point where the initial target is positioned. If the sensors' values are known the planes constructed through these points and the lenses' vertical axes will intersect in a line containing the target and orthogonal to the image plane. In case of measuring multiple targets

with the 3 linear sensors from Figure 2.10, the planes passing through the pixel values will intersect in multiple lines. Some of these lines would be defined by two crossing planes, but the targets would be found on those lines formed by three crossing planes. There is a basic condition according to which the number of intersecting lines could be determined. If two linear sensors with non-collinear orientation are used to measure N targets, the number of the intersecting lines will be N^2. This information will be insufficient to determine which line exactly contains the targets. If a third sensor is added, however, then the N^2 lines will be intersected from another N planes. In this case as a result of the mutually crossed lines and planes a set of intersected points will be formed would be projected on the image plane. The number of these points could be found from the expression:

$$N\_intersecting\_points = 3N^2 - 2N \qquad [2.15]$$

Using this formula it could be found that for 2 targets N_intersecting_points would be 8 and only 2 of them would lay on those lines containing the targets. Following [2.15] with 3 targets the number of the intersecting points would be 21 and only 3 of the lines passing through these points, obtained from three crossing planes, would contain the actual targets. Increasing the number of targets further would increase proportionally the number of crossing planes and intersecting points. As a result some of the crossing planes would become so close, that a possible confusion may occur in which 3 planes intersect in a single point instead in 3 different points. It should be noticed that in the case of orthogonal projections towards the image plane all crossing planes are parallel to one direction and so are all intersecting lines. The targets in this case lay on these lines and their positions could be calculated as a distance from the image plane using the appropriate algorithm. In other cases when the optical effects are considered the crossing planes would be directed towards the image plane at different angles and the resulted intersections woud produce points containing the targets, which positions could be calculated using transformation algorithms.

The detecting system consisting of 3 linear sensors described above is effective mainly when the targets are activated sequentially. If additional sensor is used in combination with the other three, the condition to determine the targets' position will become dependant on four planes intersecting in a single point. This redundancy would decrease the possibility of a confusion due to close intersecting planes. The second type of detecting system suggested by Dainis (1990) is referred to as a "three dimensional position measurement device" for measuring targets in space simultaneously and consists of 4 linear sensors according to Figure 2.12. Two of the sensors at the ends are mounted horizontally, so that vertical planes from each target form intersecting lines parallel to the image plane. The central sensors positioned at 45° to one another form other sets of intersecting lines. The actual target positions are found where the two planes intersect a vertical line, so that each target position is defined by four intersecting planes. The number of intersecting points in this case does not follow the condition implied by formula [2.15].

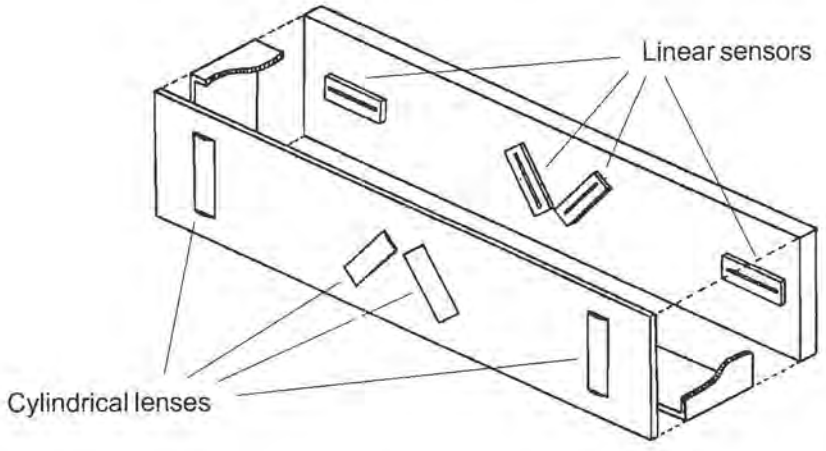


Figure 2.12 Configuration of the "three dimensional position measurement device" (Dainis, 1990).

A prototype of a 3-sensor camera system was designed at the University of Virginia with the linear sensors set up accordingly as is shown in Figure 2.6. Each linear sensor consisted of 2048 pixels and was equipped with a cylindrical lens. The outlook of this device is shown in Figure 2.13.

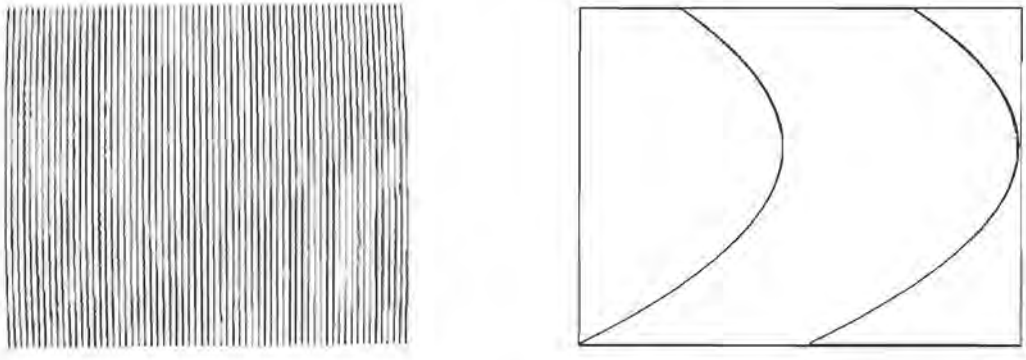
Figure 2.13 Three linear sensor camera system called Triclops, designed at the University of Virginia (Owen, 1995).

This system was capable of detecting single targets in 3D space, or multiple targets, when they were fired sequentially. The target control was implemented using a radio link and the targets (IR LEDs) were activated over a period of several milliseconds. The linear resolution tests produced errors of about 0.25 mm (20 pixels) on the sensors, positioned on the sides, and 0.07 mm (6 pixels) on the sensor at the centre. The distortion from the cylindrical lens was suggested to be the major cause for these errors. The value of the error, after the 3D reconstruction of multiple points in a volume with the following size: X = 600 mm; Y = 600 mm; Z = 1000 mm; at 1 to 2 m from the system, was found to be about 8 mm in X, 4.5 mm in Y and 40 mm in Z direction (Owen, 1995).

2.4 Commercial instruments for tracking targets in 3D space

A commercial company called Northern Digital Inc. presented its product called OPTOTRAK at the First World Congress of Biomechanics over ten years ago (Crouch, 1990). It consists of a rigid construction with 3 linear sensors mounted on a metal bar 1.1 m long as shown in Figure 2.14. The two sensors on the sides were horizontally aligned and pointed slightly inwards at 9°. The third sensor was positioned in the middle orthogonally to the other two. The optics consisted of a system of anamorphic (cylindrical) lenses that stretched the incident light into a line orthogonal to the sensor, providing a field of view of 35 degrees. It was equipped with the necessary hardware and software to drive the targets (up to 256 IR LEDs), to implement the data processing and to provide the interfacing to a host computer. It was claimed by the company in the technical specification, that the 2048 element linear CCD sensors could apparently provide a resolution higher than 1/200 000, a high sampling rate of 5 kHz and could operate at distances from 1 to 8 m (Crouch, 1990).

Figure 2.14 OPTOTRAK 3D Bar Camera (Crouch, 1990).



a b

Figure 2.15 (a) Plot of the data showing characteristic barrel lens distortion after number of points at the same and different horizontal positions were scanned for calibration. (b) Magnified view of the rightmost curves illustrating the difference between acquired data and mathematical model .

A precise factory calibration determined 19 camera parameters including compensation for symmetrical and asymmetrical lens distortion, so it could be used in the field without any prior calibration. The distortion was defined for each linear sensor in the following manner: multiple positions at one height were measured and then plotted (Figure 2.15a). The magnified curves on the extreme right side, together with the mathematical mode, are shown in Figure 2.15b.

A more recent measuring device the called CODA mpx30, designed for motion analysis, was constructed by the company Charnwood Dynamics Ltd (Figure 2.16). It used the same arrangement of linear sensors as the system designed at the University of Virginia, but the identification of the markers on the sensor was based on a different concept.

Figure 2.16 CODA mpx30 device from Charnwood Dynamics Ltd for 3D measurements.

The cylindrical lens was replaced by a flat window with a pseudo-random mask pattern, similar to a bar code (Figure 2.17). The light from the marker projected a shadow onto the sensor and when it was moved the shadow pattern also changed relative to the marker's position. The output signal from the sensor was further cross-correlated with the software model of the mask.

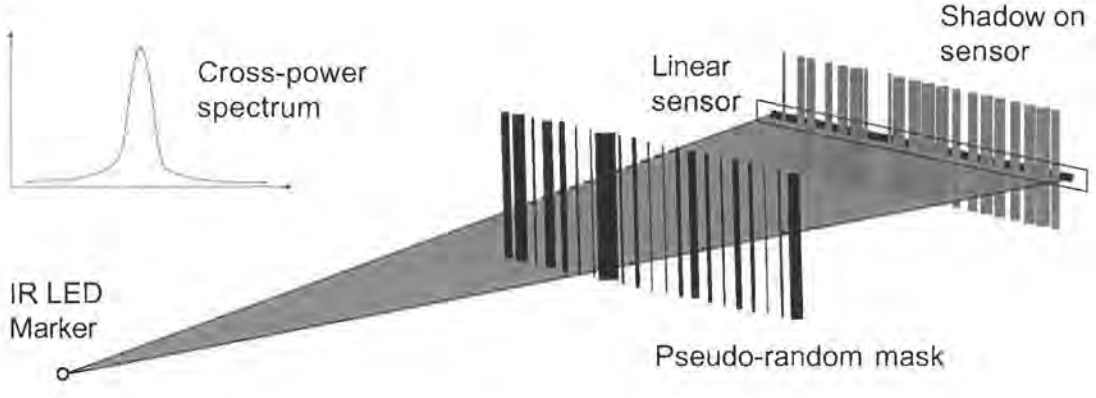


Figure 2.17 Light transformation by the mask in front of the linear CCD sensor according to CODA scanning technique and the obtained power spectrum of the cross-correlated signal.

If x (n) is used to denote the signal pattern from the software model, and y(n) denotes the signal pattern from the shadow obtained from the marker in space, the cross-correlation function can be expressed as:

$$R_{xy}(m) = \frac{1}{N}\sum_{n=0}^{N-1-m} x(n)y(n+m) \qquad 0 \leq m \leq N-1 \qquad [2.16]$$

where m is the displacement of the signal y(n) given by a finite number of samples (pixels in the case of a linear sensor). According to signal processing theory, the z-transform of the cross-correlation function gives the cross-power spectrum of the two signals (Rabiner and Gold, 1975). Following the relationship between the z-transform and the Fourier transform for the signals, these can be expressed as:

$$S(z)|_{z=e^{j2\pi f}} = S_{xy}(f) = \sum_{m=0}^{N-1} R_{xy}(m)e^{-j2\pi mf} \qquad [2.17]$$

and the cross-correlation function can be found from the inverse Fourier transform of the cross-power spectrum:

$$R_{xy}(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{xy}(f)e^{j2\pi fm}\,df \qquad [2.18]$$

The cross-correlation function is important when the signals x (n) and y(n) are related by means of a filtering operation and a delay. In radar systems, for example, y(n) could be a delayed and filtered version of x(n). It follows that the peak of the cross-correlation function is a good estimate of the delay (Rabiner and Gold, 1975). Similarly, in the case of the linear sensor, the mask acts as a filter and the linear displacement of the marker creates a displacement of the shadow on the sensor corresponding to m (Figure 2.17). Further, the obtained cross-power spectrum provides the peak frequency on the spectral curve, which after applying the inverse Fourier transform, gives the spatial (linear) displacement. This approach is efficient for noisy signals and contributes to a greater linear resolution when single targets are detected.

The above system operated with sequentially flashed IR LEDs used as markers. They were attached to the skin together with a small controlling device which synchronized them, receiving an infrared signal from the CODA camera system (Figure 2.18). Its performance was to track 6 markers at 800 Hz sampling rate or up to 28 markers at 200 Hz sampling rate. The errors were minimized to a range of 0.1 mm in the X and Z directions and to 0.6 mm in the Y direction at a distance of 3 m to the object.

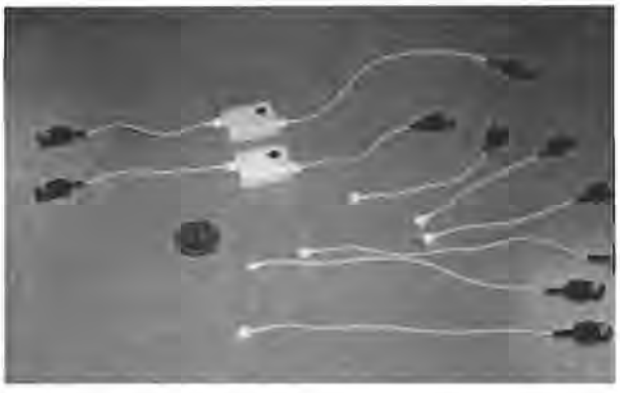


Figure 2.18 Active markers used by the CODA mpx30 divice for 3D measurements.

2.5 The problem of tracking targets

Target tracking has been the essential task in many procedures related to observing or analyzing different kinds of motion. This task has been used to determine the positions of stars, to track vessels and aircraft in naval research and space and has various applications in the robotic industry (Murray and Basu, 1994), biomechanics (Whittle, 1990; Greaves, 1995) and medicine (Maciunas, 1993, 1998). When an object is being tracked in 3D space there are two general methods to determine its momentary position. One uses the projections onto 2 dimensional planes (CCD cameras and video cameras). The first step in this approach is to define the trajectory of the object on the image, in other words to carry out 2D tracking. After this step the information from all the 2D images has to be transferred into 3D space. This procedure is called 2D →3D tracking (Greaves, 1995). The tracking process applies computation of the predicted position of the target using statistical methods whereupon it finds the residual error from the actual measured position and the predicted one. Subsequently algorithms to minimize this error are used. Another approach for object tracking is based on one-dimensional data from different detectors where typical examples can be found in most radar applications (Bar-Shalom *et al.*, 1990). This method determines distances and angles in relation to the object and updates them at each step of the process. From there the 3D position can be determined and the process could be referred to as 1D →3D tracking. Similar to the method above, differences and errors are estimated in 3D space from the predicted and measured values.

When focusing on 3D target tracking, the type of sensor used to detect the targets will be of primary importance because it determines what type of algorithm should be developed. Most video based systems are built up using 2D video cameras. Other systems for long range applications use radio signals (radar applications), ultrasound (sonar applications for naval research and navigation) or infrared light (at close range and long range as in the application of imaging sensors for surveillance). The latter are considered as 1D sensors, but their image space could be mapped in 2D for a more expedient presentation. In the case of video cameras

the tracking problem is formulated as 2D →3D tracking. It is first done in 2D, using sequential scans of the specific target, determining the target's position on each scan through the local x and y image coordinates and representing its track in 2D. These coordinates from multiple cameras are then transformed by a 3D reconstruction algorithm, allowing the target's XYZ position in the global coordinate system to be calculated at any time and its track to be updated in 3D.

Radar and sonar applications operate in a similar way and often use 2D polar coordinates for the local coordinate system because the detector position coincides with that of the source transmitting the signal. These coordinates give the range and the azimuth angle and their calculations are based on the time necessary for the signal to travel to the object and back. The term "scan" is used with these systems and it denotes one complete turn of the transmitter-receiver module, where the angular speed determines the period of the scan (Kurien, 1990). This term could be used with linear sensors as well where it would be equivalent to a single cycle reading of all pixel values.

In both applications, when either an image-based or a signal-based system for tracking is used, the algorithm can generally be divided into two major parts: single sensor tracking and multiple sensor tracking. On the other hand, tasks involving multiple targets are usually related to their unique identification on the sensor. This problem can best be solved when sequential activation of the targets is possible, so that for each scan one target at a time is reported. The method is usually based on external control of the target activation where a modulated signal is used to control this process. Each target is sequentially activated during every scan, so that the system could identify them independently. If the control signal for each target is modulated at a separate frequency then the identification is unique and there is no need of additional information from the tracking process. The identification could take place either at the stage when the signal is received from the sensor in case of an analogue approach, or at a latter stage after the signal is converted to digital form. In practice for these two applications either frequency or pulse modulation is used to solve the problem.

In the case of simultaneous target detection, which occurs with most passive targets, the identification will be solved together with the tracking problem, at the first stage of the algorithm development, *i.e.* single sensor tracking. Often at this stage, a problem may arise from targets located close to one another on the sensor, causing difficult identification. Although there exist some algorithms to separate such targets, the problem could be efficiently resolved at the next stage of the tracking process - the multiple sensor tracking, where calculations of the 3D coordinates are performed.

Another aspect of the above discussion on target tracking relates to the state and the motion parameters of the targets. The term "state" will denote the situation where the target either

belongs or does not belong to an existing track, while a "track" denotes a trajectory defined by a certain number of successive positions of the same target. Here, the state parameters include the identification and the association of the target to a specific state on each scan, while the motion parameters are all kinematic characteristics such as trajectories, velocities and accelerations. Many algorithms use a model-based approach where the two types of parameters are expressed in a mathematical form. In these algorithms the parameters complement each other, so that kinematic evaluations often utilize the identification process and in turn the correct identification of the target's state may help to obtain the kinematic evaluations (Kurien, 1990).

To implement tracking and identification of multiple targets by linear sensors in this research, a dynamic model is suggested to solve the first problem and a probabilistic model is suggested for the second problem. The dynamic model includes equations that calculate current position and velocity of the target and works for each target individually. It also uses a quadratic curve approximation on the data collected from previous scans to update the information of the tracks. This model could work both with sequential and simultaneous targets, as in the second case additional criteria should be applied to distinguish targets that cross their tracks. The probabilistic model is one possible instrument to solve this problem especially for multiple (passive) targets. It uses probability density function to characterise the behavior of each target in relation to the existing tracks, i.e. the probability of whether the target belongs to a certain track. This previously noted as target's state is used in the process of identification and only tracks containing targets are considered for evaluation in the process, while those that do not contain targets are terminated.

In the next chapters the focus will fall on the following issues: physical and optical principles of using a linear sensor in combination with a cylindrical lens; target detection, signal processing and linear resolution; three dimensional coordinates reconstruction and accuracy of the calculations; and multiple targets identification and tracking using linear sensors.

Chapter Three

Linear Versus Rectangular Sensors and Cylindrical Versus Spherical Lenses

This chapter will compare rectangular and linear Charge Couple Devices (CCD) and consider their advantages and disadvantages as well as some technological aspects and details about the necessary optics required for each of them. Rectangular CCD cameras have been used over the last two decades to determine the position of targets in space. Linear CCD sensors have often successfully replaced the rectangular CCDs in some of the applications because of recent advances in technology. The main problems to be considered here are the combinations of a linear sensor with a cylindrical lens and a rectangular sensor with a spherical lens and how both combinations affect the quality and intensity of a signal. Part of the chapter will be devoted to solving the problems related to compensation for the lens aberrations (i.e. lens distortion) on the image when different lenses are used. The resolution and data rate performance of the different sensors will also be studied.

3.1 Some technological aspects of the image sensors

The manufacturing technology of the image sensors is based on the creation of a semi-conductor structure consisting of discrete photosensitive elements. The photosensitive silicon layer on top of the element absorbs light and converts the energy of the photons from the spectral range of 400-1100 nm to liberate electrons from the silicon. These electrons create a charge which is stored in a capacitive element constructed within the silicon material by adding impurity components, thus forming zones of negative-electron abundance and positive-"hole" abundance. The charge of the capacitor is then converted to current and transferred out of the sensor. Technologically these elements differ in the structure of their silicon layers and the way they store the charge. There are two types of imaging elements, based on photodiodes, which contain a positive bias region under the photosensitive layer and Metal Oxide Semiconductor (MOS) capacitors. The latter have electrodes to isolate each element and to prevent the excessive charge from strong irradiation overflowing to the neighbouring element.

Following the principle of technology, according to which the charge is transferred out, two main types of sensors are known, the Charge Couple Device (CCD) sensor and the Complementary Metal Oxide Semiconductor (CMOS) sensor. The CCD transfers the charge from element to element in a sequential manner and each element read out destroys the image. The CMOS sensor is capable of accessing the charge at random and reading separate areas of the image independently. When its elements are based on MOS capacitors the image is destroyed by the read out as in the case of the CCD, although there is another tech-

nology called Charge Injection Device (CID), which is capable of a non-destructive read out, and where the image can be captured repeatedly.

From the point of view of internal chip topology most control functions of the CCD take place outside the chip on the printed circuit board of the camera. This allows for more flexible design depending on the requirements of the application. Unlike the CCD, the CMOS sensor has its control functions integrated into the chip, which makes the functional design less flexible.

Evaluating different factors contributing to the use of the CCD and the CMOS sensors, it can be concluded that CCDs have more advantages in resolution and image quality. CMOS sensors require less on board circuitry and system space and can offer more integration power (Litwiller, 2001). Both sensors are considered equally reliable and their cost is approximately the same and depends on which of the above characteristics are preferred by the customer. These considerations determine the main usage of the CCD sensors, which is found in the field of digital photography, broadcast television, high performance industrial imaging, and most of the scientific and medical applications. CMOS sensors are more suitable for high volume and restricted space applications with lower image quality requirements such as security cameras, PC video conference systems, bar-code scanners, fax machines, consumer scanners and some in-vehicle uses of cars (Litwiller, 2001).

3.2 Advantages and disadvantages of using linear instead of rectangular sensors

3.2.1 Physical characteristics of the sensors

First of all, it is necessary to cast some light on the conversion characteristics of systems based on rectangular and linear sensors. The above sensors are sometimes also called two dimensional (2D) and one dimensional (1D) CCDs. The rectangular sensors provide information in the x and y directions of a plane, so that landscape images or other objects can be projected from their 3D real space into the 2D image space. The conversion occurs when the incident light, from the object, passes through the lens being refracted from the glass material and is projected further onto the sensor's plane. The camera lens must be aligned parallel to the plane of the sensor. In theory the term "principal plane" describes the plane crossing the optical axis orthogonally, and which is parallel to the normal plane of the front surface. The obtained projections on the sensor are very similar to what can be seen with the human eye.

Secondly, linear sensors can obtain 1D information from points in space or 2D information from objects. In the case when a point has to be imaged on the sensor an intersecting plane, containing the point has to be constructed and directed to the sensor at certain angle. It is done either by emitting light from the point through a slit mask which forms a plane, or using

a lens, which converts the light into a straight line and crosses the sensor at the focal point. When using masks mainly active targets (laser or light emitting diodes) could be detected, while lenses allow both active and retroreflective targets illuminated by light to be used.

Linear sensors are used in an extremely large number of applications, especially in those obtaining two-dimensional (2D) images. In the process of reconstructing the information from an object (1D to 2D conversion), the linear sensor has to be moved in a direction perpendicular to its axis. This is a simple description of how the scanners operate acquiring information from the object line by line. Similar principle of operation follow the facsimile machines and the photocopiers. These devices register the information by illuminating the object from one side and detecting the patterns by linear sensors from the other. In all these cases there is a motion orthogonal to the axis of the sensor. The sensor moves in the case of the scanners and photocopiers, while in the case of the facsimile machines this is the object. Some other industrial applications use linear sensors for inspection of production lines by checking the number and the shape of the manufactured items. In these cases a spherical lens is mounted in front of the sensor which allows this sensing device to observe the line from a distance and focuses a large area on a much smaller sensor format providing a good spatial resolution. The sensor is positioned orthogonal to the direction of motion of the line and operates on the same principle as the scanner. The necessary light illuminating the objects on the line should be positioned, so that the sensor receives a clear signal. In such application different colours are often used for easier identification and to increase the number of the tracked objects.

To reconstruct objects in 3D space, rectangular sensors are used to acquire the necessary data. For this procedure there must be at least two sensors. From the pair of 2D images on the sensors a stereometric system can reconstruct every point in 3D using photogrammetric methods (2D to 3D transformation). Thus the object can be presented in 3D with all its shapes and details that are visible to the cameras. Apart from the imaging of objects, such a system can also calculate single points such as targets in space. This is a task that only requires information about the targeted point. Scientists have also used this method to determine the positions of some stars and other celestial objects in relation to Earth. A system based on linear sensors differs from rectangular sensors, in that it is not able to reconstruct shapes and details of distant objects but can only point out targets throughout the process of 1D to 3D transformation.

Last but not least, the size of these two types of sensors is an important factor for image resolution, the transfer of data and computation time. The format of a CCD camera can be rectangular or square. In the early stages of their development, CCD chips with resolutions of 256 x 256 and 512 x 512 pixel elements were released. The classic broadcast camera formats originated from the early video tube cameras, which have an aspect ratio of 4:3. Technology in the late 1980s was able to provide sensor element sizes of about 10-15 μm and later on in

the 1990s, with the advances of the chip manufacturing industry, these sizes were decreased to about 5 μm. For the typical sensor resolution of about 700 horizontal elements by 500 vertical elements, the overall size was reduced from a 2/3 inch along the diagonal to 1/2 and 1/3 inch. High resolution cameras available on the market by this time generally consisted of 3072 x 2048 pixels (Kodak DCS460) where the chip, with one of the highest resolutions of 5128 x 5128 was manufactured by DALSA (Shortis and Beyer, 1996).

Although there is a trend to build more and smaller sensor elements, the meaningful size is limited to 2-3 μm by the diffraction properties of the lenses. Sensors of 20 000 x 20 000 are theoretically feasible to manufacture, and with a pixel size of 5 μm, the format will become 100 x 100 mm. According to Shortis and Beyer (1996), cameras of 12 000 x 9 000 will probably still be the upper limit for video cameras and they will certainly be the base for photogrammetric applications which do not require a real time response. At present most manufacturers on the market are offering CCD chips with sizes from 1024 x 1024 to 4096 x 4096 (Figure 3.1). Their cost has not changed dramatically in the last few years. Both monochrome and colour chips are available as well as CCD and CMOS versions.

Figure 3.1 Monochrome rectangular CCD sensor from Kodak KAF 16801E with a resolution of 4096 x 4096 and physical image size of 36.88 x 36.88 mm.

When using the last mentioned 2D sensor, the amount of data output reaches 16 MBytes per cycle where 1Byte represent one pixel value. This sensor in Figure 3.1 operates at a clock frequency of 10 MHz which means that 1.7 seconds is the time needed to complete the read out of a single frame. As communication in this instance is slow and these type of sensors can be applied in the field of digital photography and astronomy, where the speed limitations are not of primary importance, but where the spatial resolution is the key factor. At the other end of the product range one finds sensors with a small crystal size *e.g.* 240 x 240 pixels and with high speed rates of over 1000 frames per second.

Compared to these 2D sensors, linear sensors have a significantly lower number of *total*

pixels ranging from 1024 to 8192, with some high resolution devices consisting of up to 15 000 pixels (Figure 3.2). They are capable of operating at much higher data rates than the 2D sensors. Typical values of their speed vary from 1000 to more than 20 000 lines per second. For example, a sensor with 8192 pixels from DALSA can provide a line rate of 18 kHz, reaching an output data flow of 147.5 MBytes/second, if every pixel is represented by 1 Byte. This means that with this type of linear sensor the communication line may be unable to support this data rate and some data reduction would be necessary prior to sending the data

Figure 3.2 Colour and monochrome linear sensors from Kodak. Colour linear CCD sensor KLI-14403 with 14 403x3 (RGB) pixels, 72 mm long with pixel size 5 x 5 μm (left) and monochrome linear CCD sensor KLI-8811 with 8800 pixels, 61.3 mm long with pixel size 7 x 7 μm (right).

to the next system module. As long as only points are detected by such a system, and the position of each target has been calculated, the output data rate can be reduced to:

$$\text{Data rate} = N \times r_s \qquad [3.1]$$

where:

N = the number of detected targets per line

r_s = the line rate of the sensor

In this instance the sampling rate of the system will be based on whether the targets are identified simultaneously or sequentially. With the first assumption, which is always the case with passive (light reflecting) targets, each target will be captured just as many times per second as the sensor line rate is:

$$\text{Target Sampling Rate} = r_s \qquad [3.2]$$

If targets are activated and identified sequentially, then the sampling rate for each target can be given by:

$$\text{Target Sampling Rate} = \frac{r_s}{N} \qquad [3.3]$$

where N and r_s have the same meanings as in [3.1]. Therefore, it will depend on the application requirements whether the value of the Target Sampling Rate will be sufficient to describe the kinematics of the motion. In most of the measurements in biomechanics, for example, sampling rates of 100 Hz are usually considered acceptable. For some industrial applications, however, frequencies up to 500 Hz and higher might be needed. From formula [3.3], where the sampling rate is known, one can calculate theoretically the possible number of targets that can be tracked if they are activated sequentially. In the case of simultaneously controlled targets, this number will depend on the resolution of the linear sensor, the number of pixels activated from a single target, the field of view and the distance from the sensor. Here one can mention that one of the advantages of the rectangular sensor to the linear sensor is that it can identify a maximum of N^2 targets simultaneously at the expense of the output data flow, whereas in the same situation the linear sensor can only identify N targets at higher speed.

3.2.2 The combination linear sensor and cylindrical lens

Using the combination of linear sensor and cylindrical lens to image point targets in space is presented on Figure 3.3.

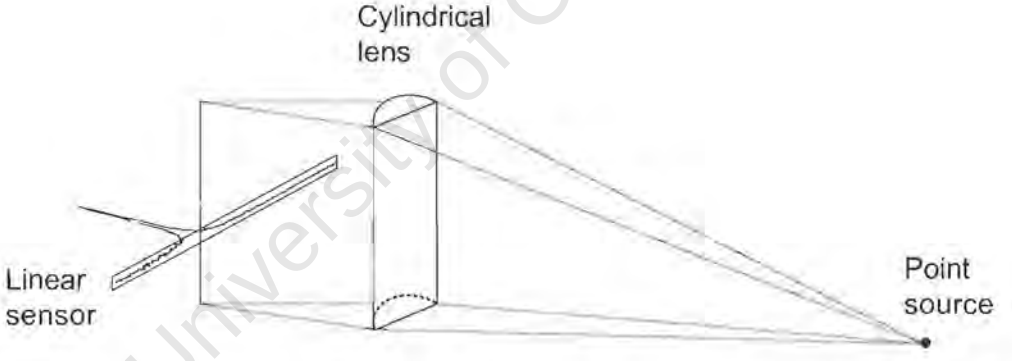


Figure 3.3 Image of a point source through a combination of linear sensor and cylindrical lens.

The lens has optical power along its vertical axis which is also referred to as the "power meridian". This type of lens does not have any optical power along the other horizontal axis and converts the light from a point source into a straight line parallel to the vertical axis. The outgoing light, which forms a line, intersects the sensor further along and activates a group of pixels according to the position where it strikes the sensor after being refracted from the lens. Figure 3.4 illustrates one example of this effect captured by a rectangular sensor. Changing the position of the point source in space will change the position of the refracted light in a direction orthogonal to the vertical axis. It will also change the position of the line in Figure 3.4 in a horizontal direction. This displacement of the line on the image will only relate to the particular component of the target's motion vector in 3D, which is parallel to the sensor.

Figure 3.4 A line image on a rectangular sensor obtained from a point source passing through cylindrical lens.

3.3 Theoretical considerations of some optical aberrations

The following section of this chapter will focus on some of the optical effects of a spherical lens when used with a rectangular sensor. As the subject of this work is focused mainly on the monochromatic detection of targets in the near infrared range, optical aberrations due to the chromatic property of light will not be taken into account. Further in the chapter a parallel will be drawn between the nature of the distortion from a spherical and a cylindrical lens and some supporting measurements will be presented.

With regard to the projection of images, there are two main considerations to be borne in mind when imaging point objects through a lens onto a plane, which is the sensor surface itself: firstly, factors affecting position of the projected image and secondly, factors reducing the quality of the image. Changes in intensity when using spherical and cylindrical lenses will be illustrated at the end of this chapter. From the first group of factors it is believed that lens distortion plays the most significant role in the far and close range photogrammetric applications. Distortion, as described in optical theory, is one of the terms in the equation completing the expression of the so-called Siedel aberrations (Born and Wolf, 1959). Some of these terms influence the quality of the image and the focusing, but distortion is the one that changes the lateral magnification considerably, thereby affecting the positioning on the projected image.

The next section starts with an explanation of the general lens distortion concepts of a spherical lens from the point of view of photogammetry. This is followed by some formulations from the geometrical theory of aberrations and derivation of the aberration functions. Later in the chapter the focus is on the instrumental techniques which can free a lens from optical aberrations.

3.3.1 Distortion of a spherical lens from a photogrammetric point of view

To reiterate what was pointed out in the literature review, the symmetrical or radial distortion is described as a sum of mathematical terms in its Gaussian form

$$\delta_r = K_1 r^3 + K_2 r^5 + K_3 r^7 + \ldots \quad [3.4]$$

where K_1, K_2, K_3 are coefficients of radial distortion

$$r^2 = (x - x_p)^2 + (y - y_p)^2 \quad [3.5]$$

r is the radial distance, x, y are the image coordinates and x_p, y_p are the coordinates of the principal point.

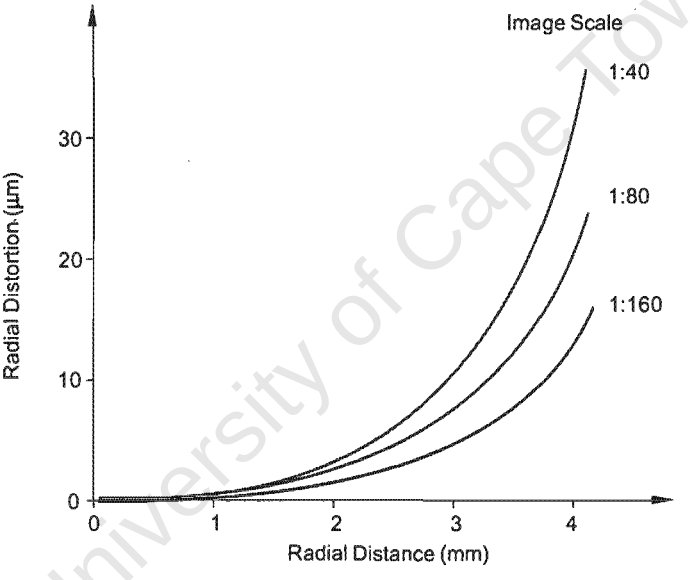


Figure 3.5 Gaussian radial distortion of a 25 mm spherical lens at various image scales (Fryer, 1996).

Generally δ_r is measured in µm and varies with the radial distance, which is given in mm. From Figure 3.5 it can be seen that the distortion is dependent on the focusing. If an object is positioned at a certain distance from the lens and images are taken at various focal lengths, the size of these images on the projection plane would be different and this is related to as image scaling. It also varies with the depth of field, but only in very close range cases. The magnitude of the radial distortion can be as large as 300 µm on the side of the image format and should not be ignored in precise photogrammetric tasks (Fryer, 1996).

Balanced distortion δ_{rb} is a term used where the Gaussian curve is mathematically transformed by changing the principal distance from the lens to the sensor, defined in [2.1] and [2.8] and shown in Figure 2.1, with Δc to meet the following conditions:

- the mean value of the transformed distortion curve extended to a certain radial distance is zero (Figure 3.6), or
- the mean square value of the distortion to a certain radial distance is a minimum, or
- the maximum and minimum values of distortion to a certain radial distance are equal (Brown, 1968).

This is achieved by adding the term $K_o r$ in equation [3.4], and shown in equation [3.6]. The role of this term is to tilt the abscissa axis at a certain angle to satisfy the desired condition.

$$\delta_{rb} = \delta_r + K_o r = K_o r + K_1 r^3 + K_2 r^5 + K_3 r^7 + \ldots \quad [3.6]$$

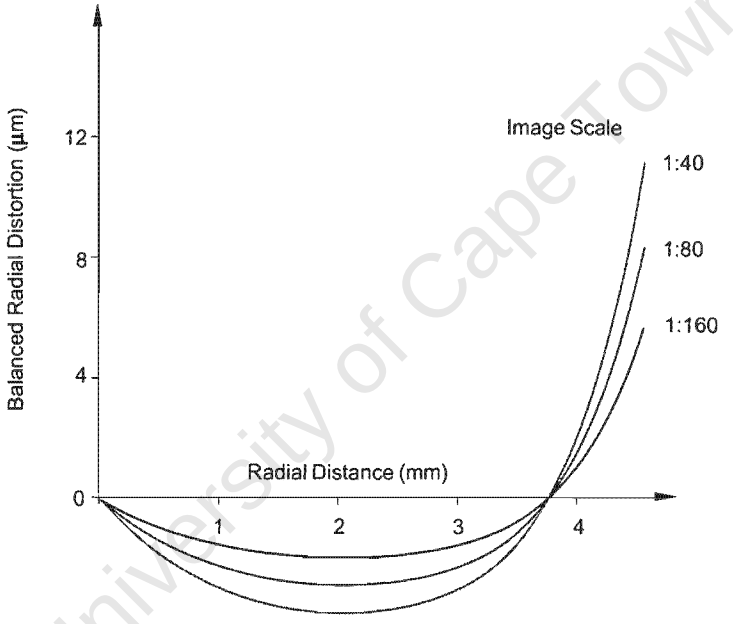


Figure 3.6 Balanced radial distortion of the spherical lens from Figure 3.5 (Fryer, 1996).

For most lenses on non-metric cameras, only the elements K_o and K_1 are significant for describing distortion. For wide angle lenses, however, it might be necessary to include K_2 and K_3 as well. These radial distortion coefficients can be found at two different focus settings, one set close to the camera and the other one set at infinity (Magill, 1955; Brown, 1972).

Decentering or tangential distortion is considered as a displacement from the ideal optical axis of the lens, which causes "off centre" geometrical displacement on the image. The reason for this could be imperfection in the lens mounting in the system where the optical axis is not perfectly orthogonal to the image plane. Its graphical representation is similar to that of the radial distortion, but its magnitude is usually at least one order less (Figure 3.7).

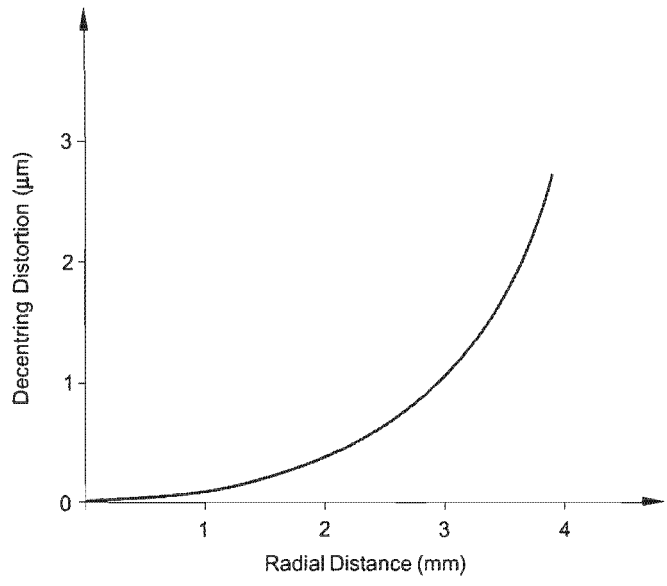


Figure 3.7 Decentering distortion of a 25mm spherical lens from Figure 3.5 (Fryer, 1996).

3.3.2 Optical aberrations of a spherical lens from a general optics point of view

According to general optics theory, lens aberrations can be derived using trigonometric functions. The concept is based on the sine condition for refracted light and the power expansion of the sine and cosine functions:

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots$$
$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots \qquad [3.7]$$

Using these approximations for small angles, which applies in the case of paraxial rays, only the first term in the series would describe the angle sufficiently. This is known as the first order aberration theory and in this instance Gaussian Optics implies exact imaging (Pedrotti and Pedrotti, 1987). When the expansion includes the second term $\theta^3/3!$, it denotes the third order theory. The aberrations were studied and classified by the German mathematician Ludwig von Siedel in 19th century. There are five elements in the expression derived from the aberration function called Primary Siedel Aberrations (Born and Wolf, 1959). In Figure 3.8 a pair of coordinates is shown *e.g.* x_o, y_o, of the point P_o in the object plane, projected coordinates ξ_1, η_1 of the point P_1 in the image plane and the pair of coordinates $x_o\xi_1$, $y_o\eta_1$ at the intersection P' of the incident ray from P_o, with the cross-sectional plane of the lens orthogonal to the optical axis.

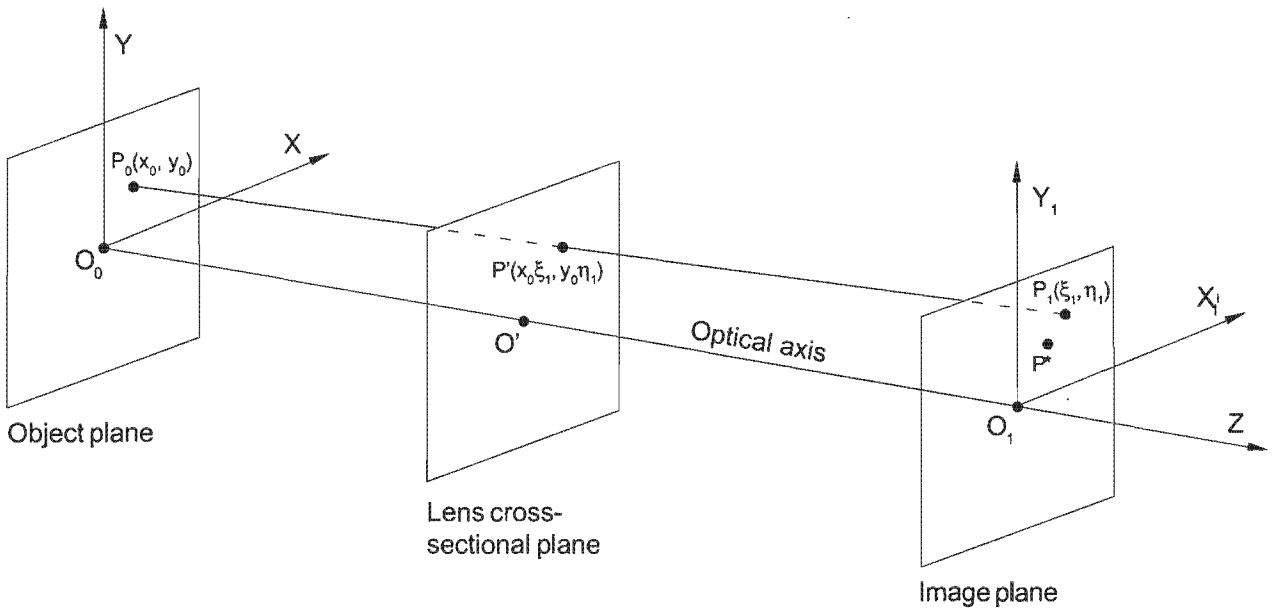


Figure 3.8 Object plane, lens cross-sectional plane, image plane and projections of the point P_o and their coordinates showing optical aberration of the point P_1 from the ideal image point P* (Born and Wolf, 1959).

$$\phi = -\frac{1}{4}B\rho^4 - Ck^4 - \frac{1}{2}Dr^2\rho^2 + Er^2k^2 + F\rho^2k^2 \qquad [3.8]$$

where $\phi(x_0,y_0,\xi_1,\eta_1)$ is the aberration function of four variables x_0,y_0,ξ_1,η_1 entering the following combinations:

$$r^2 = x_o^2 + y_o^2 \qquad \rho^2 = \xi_1^2 + \eta_1^2 \qquad k^2 = x_o\xi_1 + y_o\eta_1 \qquad [3.9]$$

In this description, r and ρ are the radial distances at the object and image plane respectively and the pair $(x_o\xi_1, y_o\eta_1)$ corresponds to the coordinates in the lens cross-sectional plane with their radial distance k. The aberration in the X and Y direction can be described in the form of the first derivative of the function ϕ:

$$x_1 - x_o = -\frac{\partial\phi}{\partial\xi_1}$$
$$y_1 - y_o = -\frac{\partial\phi}{\partial\eta_1} \qquad [3.10]$$

Here x_1 and y_1 are the transformed ξ_1 and η_1 in the image plane. Substituting [3.8] in [3.10]

gives the general expression of the ray aberration components:

$$\begin{aligned}\Delta x &= x_1 - x_o = x_o(2Ck^2 - Er^2 - F\rho^2) + \xi_1(B\rho^2 + Dr^2 - 2Fk^2)\\ \Delta y &= y_1 - y_o = y_o(2Ck^2 - Er^2 - F\rho^2) + \eta_1(B\rho^2 + Dr^2 - 2Fk^2)\end{aligned} \qquad [3.11]$$

This is simplified by choosing the object point to lie in the Y-Z plane (Figure 3.8), so that $x_o = 0$. Introducing further polar coordinates it could be written:

$$r^2 = y_o^2 \qquad \xi_1 = \rho\sin\theta \qquad \eta_1 = \rho\cos\theta \qquad [3.12]$$

and [3.8] becomes:

$$\phi = -\frac{1}{4}B\rho^4 - Cy_o^2\rho^2\cos^2\theta - \frac{1}{2}Dy_o^2\rho^2 + Ey_o^3\rho\cos\theta + Fy_o\rho^3\cos\theta \qquad [3.13]$$

Then for the aberrations in the x and y directions it could be obtained:

$$\begin{aligned}\Delta x &= B\rho^3\sin\theta - 2Fy_o\rho^2\sin\theta\cos\theta + Dy_o^2\rho\sin\theta\\ \Delta y &= B\rho^3\sin\theta - F\rho^2y_o(1 + 2\cos^2\theta) + (2C + D)y_o^2\rho\cos\theta - Ey_o^3\end{aligned} \qquad [3.14]$$

If all coefficients in [3.13] have a zero value the case equals the description by the Gaussian approximation. The following types of Siedel aberrations are known if the mentioned coefficients have finite non-zero values:

1. Spherical aberration ($B\neq0$)

$$\begin{aligned}\Delta x &= B\rho^3\sin\theta\\ \Delta y &= B\rho^3\sin\theta\end{aligned} \qquad [3.15]$$

2. Coma ($F\neq0$)

$$\begin{aligned}\Delta x &= -2Fy_o\rho^2\sin\theta\cos\theta = -Fy_o\rho^2\sin2\theta\\ \Delta y &= -F\rho^2y_o(1 + 2\cos^2\theta)\end{aligned} \qquad [3.16]$$

3. Astigmatism ($C\neq0$), (Figure 3.9)

$$\begin{aligned}\Delta x &= 0\\ \Delta y &= 2Cy_o^2\rho\cos\theta\end{aligned} \qquad [3.17]$$

4. Curvature of field (D≠0)

$$\Delta x = Dy_o^2 \rho \sin\theta$$
$$\Delta y = Dy_o^2 \rho \cos\theta \qquad [3.18]$$

These two components of the aberration error function are best studied together. Then C≠0, D≠0

$$\Delta x = Dy_o^2 \rho \sin\theta$$
$$\Delta y = (2C + D) y_o^2 \rho \cos\theta \qquad [3.19]$$

The astigmatic image contains tangential T and sagittal S focal lines (Figure 3.9). The image is formed about midway between T and S, where the focus would be circular and is denoted as the "circle of least confusion". The curvature of field also contains tangential and sagittal components and they are characterised by the values: 2C + D, usually called the tangential curvature of field, whereas D is called the sagittal curvature of field.

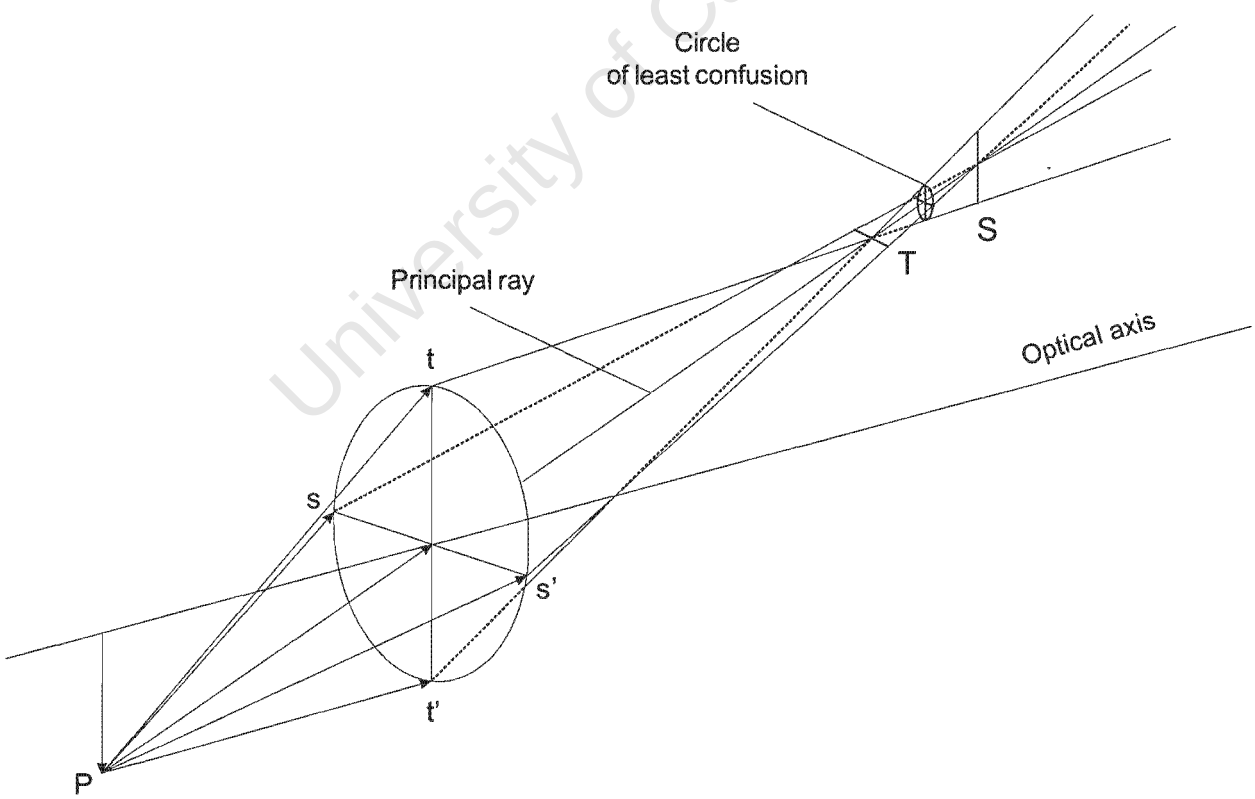


Figure 3.9 Astigmatism of a lens with its tangential and sagittal focal lines (Pedrotti and Pedrotti, 1987)

5. Distortion ($E \neq 0$)

$$\begin{aligned} \Delta x &= 0 \\ \Delta y &= - E y_o^3 \end{aligned} \qquad [3.20]$$

Pincushion distortion ($E<0$) and barrel distortion ($E>0$) are distinguished (Figure 3.10 b,c) by off-axis unproportionality of an image consisting of straight lines (Figure 3.10 a).

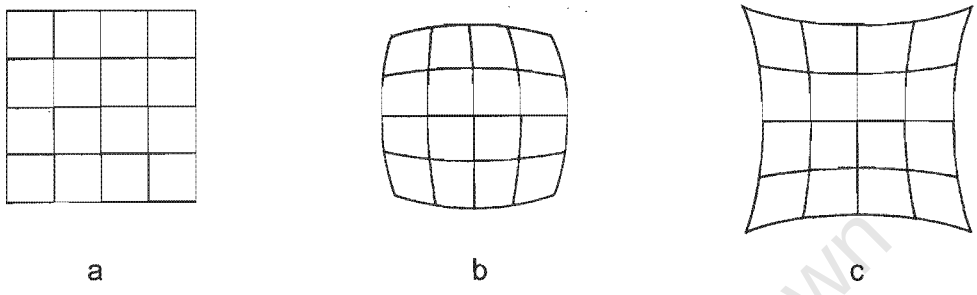


Figure 3.10 Distortion transforms the straight lines from a grid image (a) into a barrel type image (b) or pincushion type image (c).

Three of the described Siedel aberrations, namely the spherical aberration, coma and astigmatism, cause a blurring effect on the image. The other two, namely the curvature of field and distortion, produce changes in the position and the form of the image. In general it is impossible to construct a system which is free from all five aberrations. However, a suitable compromise between their types and magnitudes could provide a solution in the case of a precise optical instrument. Some of the examples in photogrammetry show that non-metric cameras can be free of the first four components, but they will still display signs of distortion, which has to be further compensated for by using mathematical models. Although these equations are not difficult to implement, the calibration might be slightly affected in terms of stability and may require more extensive pre-measurement activity.

For a cylindrical lens it could be assumed the same arrangement from Figure 3.8. By taking into account the fact that optical transformation will occur only along one of the directions of the principal plane (lens cross-sectional plane) the assumption of setting one of the coordinates x_o or y_o to zero will also be valid. Then equations [3.12] through [3.20] will define the changes in the image space due to the explained optical aberrations in the same way as it was shown for a spherical lens. It should be pointed out that if the aberrations affects both directions, as in the case of a spherical aberration, coma and curvature of field, the actual changes in the positioning and the image shape and quality will appear only in the direction parallel to the lens curvature. In the other direction these aberrations will take place at infinity due to the infinite value of the radius of curvature. Distortion and astigmatism according to [3.17] and [3.20] appear only in the direction of the non-zero coordinate. When cylindrical lens is used,

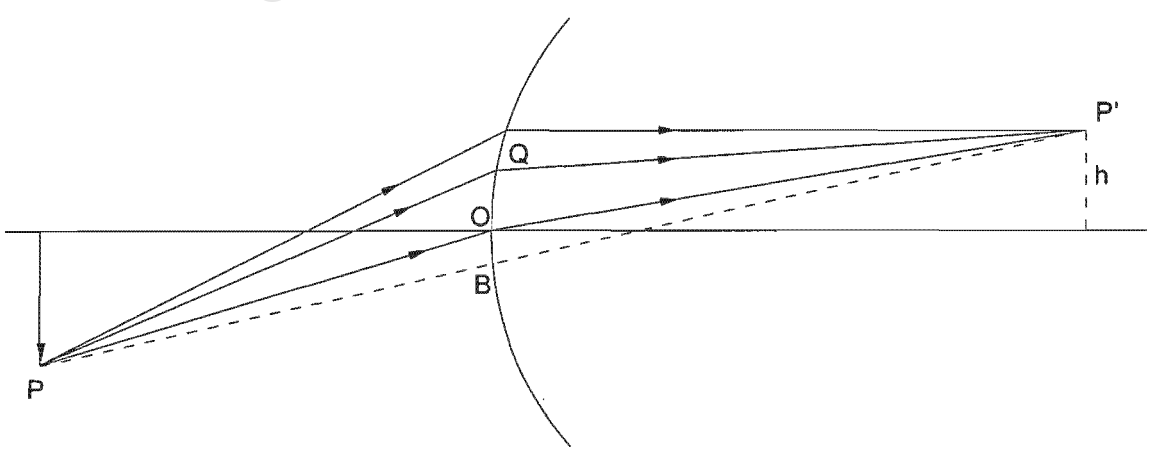
the coordinate axis along which these aberrrations are measured must coincide with the lens power axis.

The first three optical aberrations that affect the image quality could have a strong effect on the linear resolution. Blurring may cause asymmetric shapes of the registered patterns and could extend them over a larger area on the sensor, where the values of the activated pixel elements could be reduced. Subsequent use of algorithms, such as the centroid algorithm, presented in the next chapter, involving the values from these pixels to calculate exact position may produce inaccuracies due to the above aberrations. In such occasion using filters is believed to be one practical solution, because these inaccuracies are close in magnitude to the noise from the ambient light. Distortion, however, is observed to have much higher influence on the positioning and mathematical as well as physical methods are suggested for its compensation.

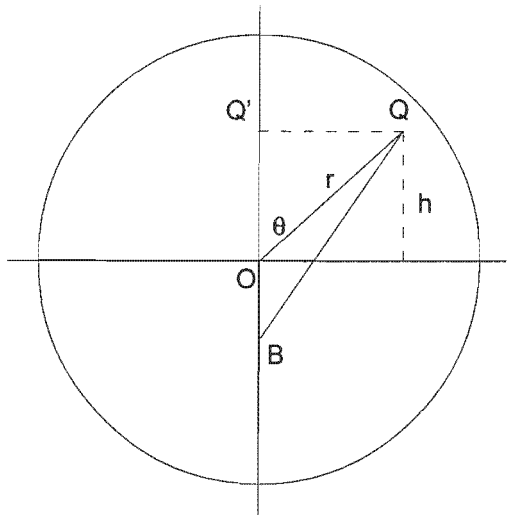
From the previously described theoretical considerations in section 3.3.2 there is another interpretation of the optical aberrations of a spherical lens when the terms are directly taken from the aberration function [3.13] (Pedrotti and Pedrotti, 1987). Thus, the distortion could be described by:

$$\delta = E\, h^3\, r \cos\theta \qquad [3.21]$$

In this equation h and r have replaced y_o and ρ from [3.13], while θ remains the same. The above parameters are shown in Figures 3.11a and 3.11b



a



b

Figure 3.11 Surface of a spherical lens and projection of an off-axis point. Lateral view (a) and frontal view (b).

Then the following transformation for a cylindrical lens could be derived. Taking into account the fact that there is an optical power only along the vertical axis, the point Q will be projected in Q' on the vertical axis, setting $\theta = 0$, $r = h$ and equation [3.21] will be rearranged to:

$$\delta = E\, h^4 \qquad [3.22]$$

Due to the convergent character of the trigonometric functions describing the optical aberrations the mathematical model suggested to compensate the latest uses power series of odd or even terms [3.7], where only the first one or two terms could be significant. In this case equation [3.25] may require the use of the first and the second terms in addition to the third term in the expanssion of the cosine function presented in the original equation for the distortion. As it was pointed out by Marzan and Karara (1976), for some lenses the distortion could be expressed using both odd and the even terms of the power series [2.13]. This expression will be of significant importance when large lateral displacements in the object space are observed and the projected images appear at the end of the image format defined by the sensor.

Another approach to free a lens system from distortion is to use the effect of aperture stop. Such a stop, placed close in front of the lens, produces barrel type distortion and when it is behind the lens leads to a pincushion type. In reality the effect of stop might be created from an opening on the cover of the system or be a part of the lens holder.

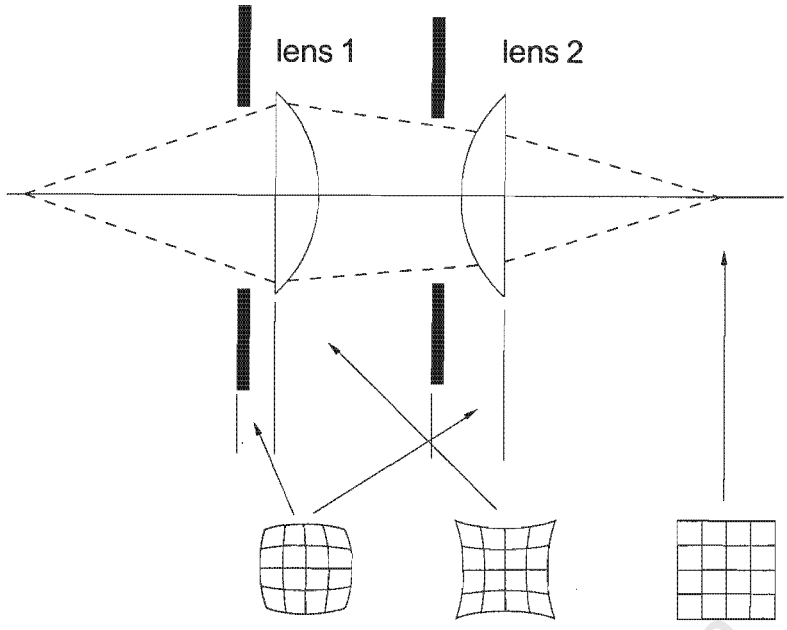


Figure 3.12 One example of a lens doublet and two aperture stops for correcting the initial barrel distortion.

This effect allows for a combination of two lenses (a lens doublet) to be assembled with one opening in between and another one at the front, as shown in Figure 3.12, so that the barrel distortion, created by the first and the second opening in front of lens 1 and 2, is eliminated by the equivalent amount of pincushion distortion created by the second opening behind lens 1. However, even after taking such careful precautions, some aberration due to a curvature of field may occur when the lens system is tested for its lateral resolution. It is also known that a field stop placed behind the lens can change the angular field of view (Figure 3.13).

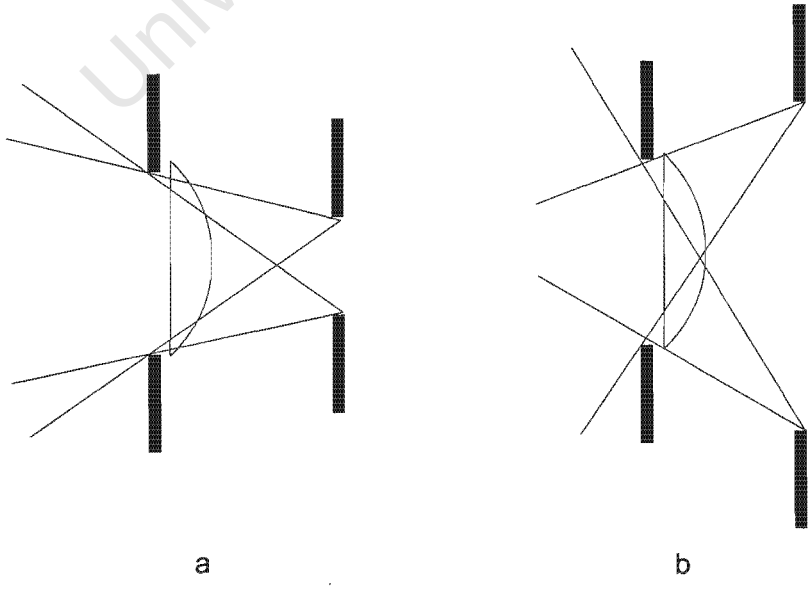


Figure 3.13 The effect of the field stop created from the size of the sensor behind the lens: (a) for a rectangular sensor;and (b) for a linear sensor.

Compared to the size of a rectangular CCD sensor of about 35 mm, which has 4096 x 4096 pixels on each side, a linear CCD with 8800 pixels is almost double that size, being over 60 mm in length. If both of them are equipped with a lens of the same width, both spherical and cylindrical respectively, the field stop created by the size of the sensor in the second instance will enlarge the field of view. This in turn will increase the lateral distance in the object space that can be projected through the lens onto the sensor. According to the definition of the optical aberrations, one can conclude that a longer linear sensor will produce more aberration changes from a lens with the same geometrical features because the image format is extended and can look at a larger object space. Therefore, considering [3.15] through [3.20] where y_o and ρ are involved in the expressions, whenever the range of these variables changes, it is expected that the amount of the related optical aberration will change proportionally.

Lens distortion also depends on the focal distance as shown in Figure 3.5. Following the example in Figure 3.12b, if a point is projected on the linear sensor at a certain lateral distance from the optical axis, it could be shown that moving the sensor farther from the lens will produce narrower angle of view and the point will be projected closer to the end of the sensor, where the image will suffer from more distortion.

3.4 Experimental measurements for distortion by a cylindrical lens

As distortion changes the lateral magnification, equally spaced points projected on to a cylindrical lens will be imaged at different intervals. This has been determined in two ways for a cylindrical lens in order to define how much the linear resolution of an optical system will be affected. With this purpose in mind, the first test was done using a computer simulation model of such a system. Secondly, a real system of a cylindrical lens combined with a linear CCD camera was used to measure targets and both results were plotted and analysed. All measurements were taken along the horizontal axis parallel to the CCD line.

This experimental data was obtained using simulation software (Optical PlayGround), which constructed a system of a cylindrical lens, orthogonal to a linear CCD camera and a light source. The software model of the lens was created following a real lens prototype later used for real measurements to determine linear resolution as well as to capture data with multiple sensors for 3D coordinates reconstruction. This particular lens was a planoconvex type i.e. with a cylindrical curvature only on the one side while on the other side it was plane. The lens was described by the following parameters: height (along the axis of the cylinder) = 42.25 mm; width (along the axis of optical power) = 19 mm; thickness = 4 mm; radius of curvature = 16.54 mm; focal distance ≈ 37 mm; refractive index = 1.52, corresponding to glass material BK7. The CCD sensor consisting of 2048 sensitive elements (pixels) and 24.58 mm in length, was positioned at the focal distance behind the lens.

A light source was positioned in front of the camera at the same level as the linear sensor at a distance of 1500 mm between the source and the optical surface.

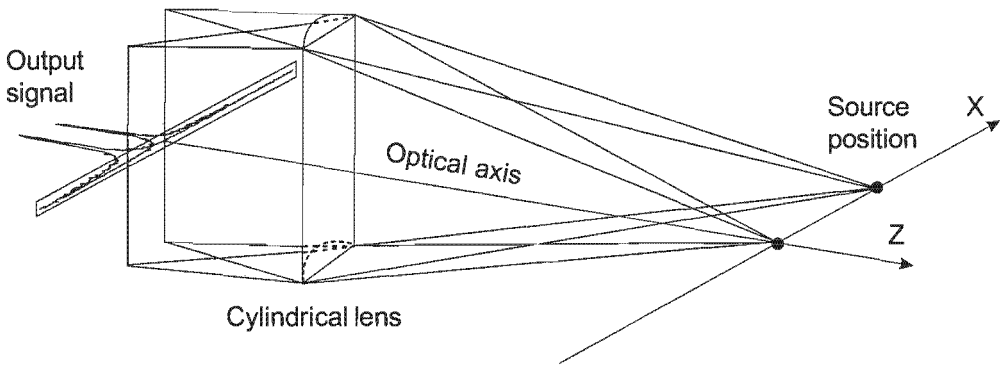


Figure 3.14 Positioning of the source moved along the horizontal axis to measure the distortion of a cylindrical lens produced on a linear sensor. The displacement of the peak on the output signal corresponds to the linear displacement.

The source was moved sequentially at equal intervals along the horizontal direction starting from the optical axis towards the right side of the field of view of the lens-sensor system The total displacement covered about 1/5 of the lateral field of view or 171.5 rnm from the centre to the side. This model produced as output, a peak for each of the 50 points corresponding to their positions on the X axis, starting from the centre at 0 through intervals of 3.5 mm (Figure 3.14). The maximum pixel value from each peak was chosen to represent the position of the source point. Then these values were plotted along the vertical axis versus the distances to show the character of this function. The predicted values of these measured points were also calculated and plotted on the same graph. The data from this graph is summarised in Appendix A.

In Figure 3.15 the camera function is curved at the side section, whereas the section close to the axis equals the line of the predicted values . The exponential shape of the measured line suggests a barrel type distortion, where for those intervals away from the centre of the image space, the corresponding intervals on the sensor become shorter. These differences in the interval length are smaller if they are formed by those points close to the centre.

The measured camera function can be further approximated by a polynomial. Such an approximation follows the theory from section 3.3.1 discussing the distortion from a photogrammetric point of view. This approach also complies with the theory from section 3.3.2 where power series are used to represent the trigonometric functions describing the character of the optical aberrations. Increasing the polynomial order from 1 to 10 minimises the approximation error as shown in Table 3.1 and it becomes considerably lower after the second order curve is applied.

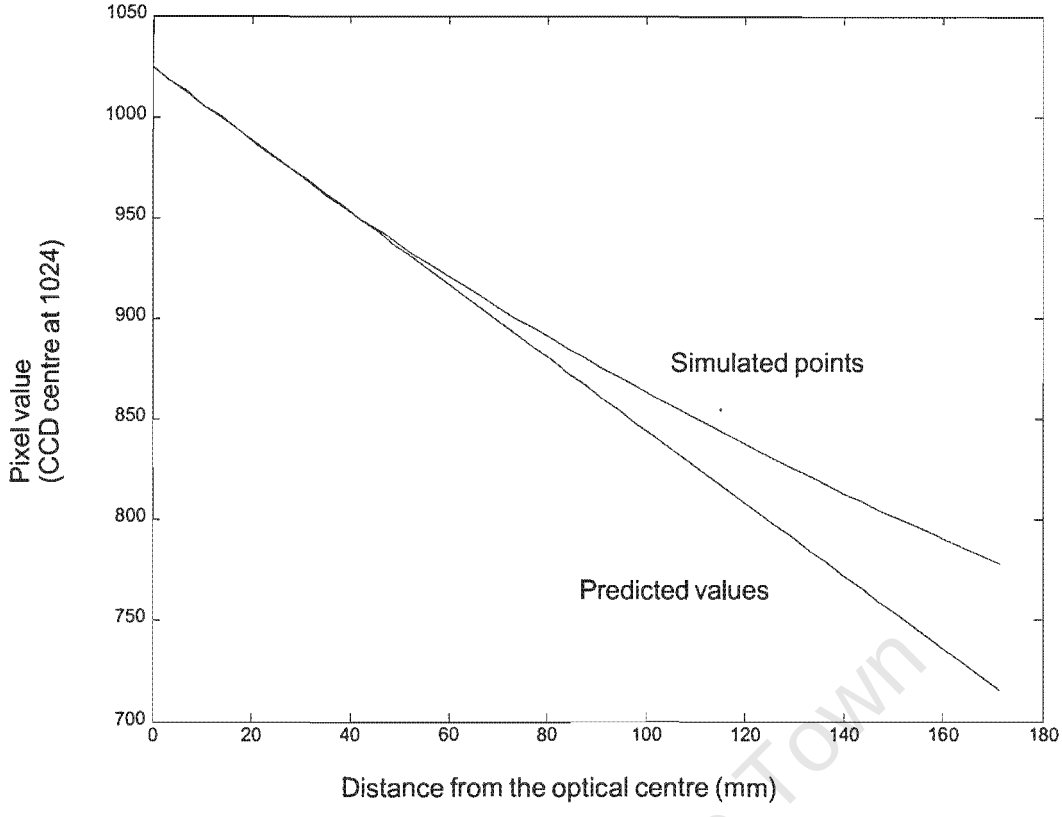


Figure 3.15 Software simulation of a distortion of a cylindrical lens obtained along the horizontal line parallel to the lens power axis. The line in red shows the character of the function from the simulated points and the line in green the line from the predicted values.

Order	Mean pixels	S.D. pixels	R.M.S. pixels	Abs. error pixels
1	0.00	6.00	5.94	12.79
2	0.00	0.62	0.62	1.29
3	0.00	0.35	0.35	1.09
4	0.00	0.35	0.34	1.06
5	0.00	0.29	0.29	0.83
6	0.00	0.26	0.26	0.74
7	0.00	0.25	0.25	0.76
8	0.00	0.25	0.24	0.81
9	0.00	0.24	0.24	0.77
10	0.00	0.23	0.23	0.77

Table 3.1 Mean, standard deviation and absolute maximum error of the differences between measured pixel data from Figure 3.14 and their curve fit for polynomial orders from 1 to 10.

These data also shows that approximation curve follows precisely the character of the measured data producing a negligable mean value and only the standard deviation and the absolute maximum error are the significant factors describing the random error. This error has decreased considerably after the use of the first three power orders in the polynomial, thus confirming the conclusions made by the photogrammetrists that using power terms to model the lens distortion gives significant results up to the term of order 5.

The second test was performed with a real camera lens system according to the setup shown in Figure 3.16. Two infrared (IR) GaAs light emitting diodes (LEDs LD275, Siemens) were used to form linear displacements, one positioned in the centre of the field of view at 0 mm, and the other one moved along the horizontal axis parallel to the CCD sensor. The cylindrical lens used for this test was manufactured by Edmund Scientific (Barrington, USA) and its geometrical parameters were already described in the previous simulation procedure. The linear sensor for the experiment was incorporated inside a commercial camera device Pulnix TL-2048HJ. In this case the camera-source distance was set to 300 mm to linearize precisely the targets according to the lens-sensor system and to receive a good angular view in both left and right directions. The video output of the camera was connected to a digital oscilloscope and the time intervals between the two peaks were measured in microseconds. These measurements were taken at 5 mm steps in both directions. The measured horizontal displacement permitted by the sensor-lens angular view was 135 mm to the left of the object space, while that to the right was 130 mm.

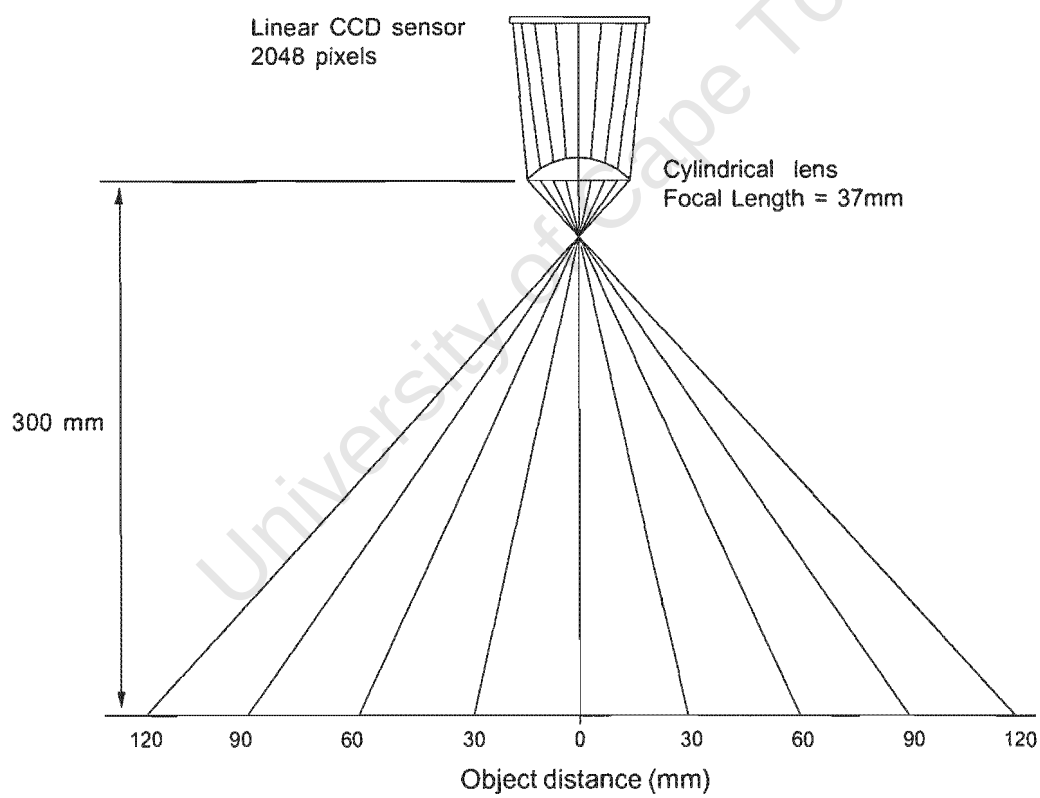


Figure 3.16 A setup of a cylindrical lens and a CCD sensor showing the directions of the light from different angles and refracted through the lens onto the sensor. The projected image points suggest possible distortion of a "barrel" type. This picture does not show an exact scale considering the distance from the source to the lens and the lateral distance.

The alignment conditions of this set up were achieved in the following way. First, the camera, which carried the sensor and the lens was aligned to the ground horizontal direction using a

bubble level, and then in the same way a vernier caliber, which carried the source (an IR LED clamped to the moving arm) was aligned. The letter was used as a direction to move the source. The midpoint of the lens was aligned by positioning three points on the same line: a laser pointer behind the camera, the middle of the lens and another reference point taken between the lens and the source. This reference point was selected on the perpendicular to the directions of the sensor and the source. Next, the source was slid to a position putting it on the line defined by the laser beam. The laser beam was assumed to be parallel to the optical axis, both defining a plane orthogonal to the sensor. Inaccuracies in the above set up could have been due to hand positioning and instrumentation tolerances.

Prior to the alignment procedure the lens was fixed to the camera with its main axis orthogonal and its flat surface parallel to the direction of sensor. From this fixing inaccuracies in positioning the lens could have come from a slight rotation of the lens around the optical axis and around the vertical axis (orthogonal to the sensor).

The intervals from the image space on the sensor, recorded from the oscilloscope, were then converted from microseconds into pixels using the time/space ratio to read out the entire CCD sensor, and plotted versus the intervals corresponding to the horizontal displacement. The predicted values of these points were also calculated and plotted on the same graph. The comparison of the measured pixel intervals and their predicted values, as seen in Figures 3.17 and 3.18, showed a systematic decrease in length towards the ends of the image space.

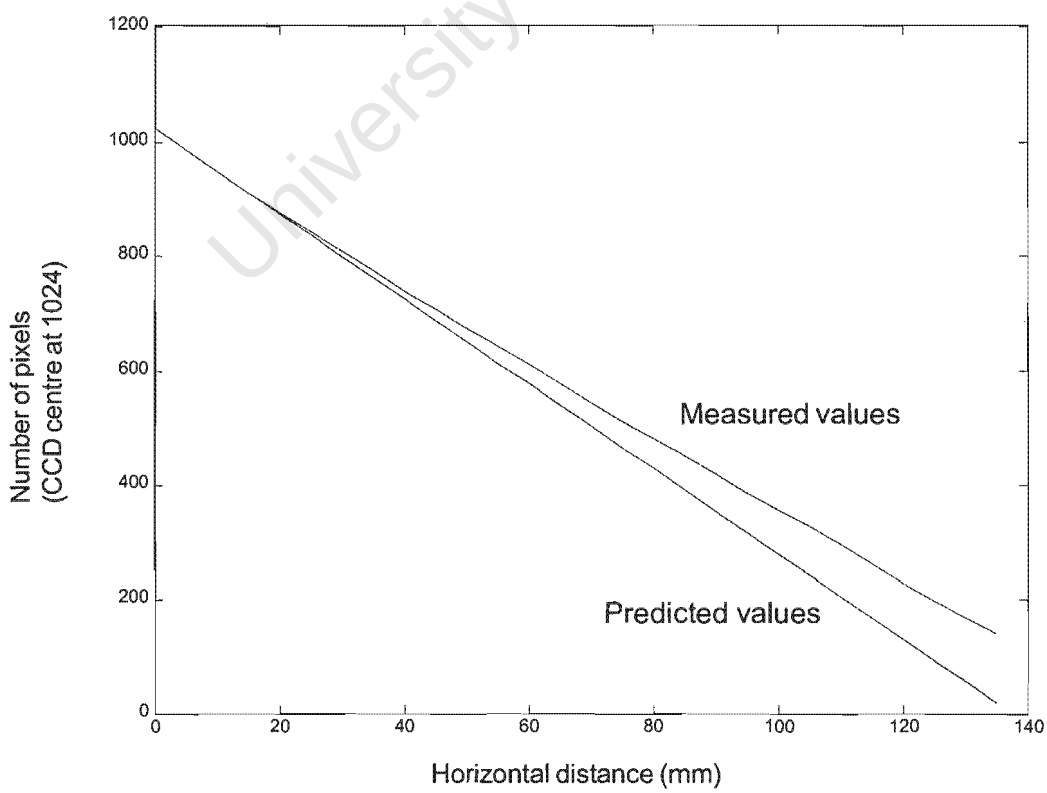


Figure 3.17 Measurements in pixels, taken from a linear CCD and cylindrical lens and their predicted values from the left hand side of the object space on Figure 3.16.

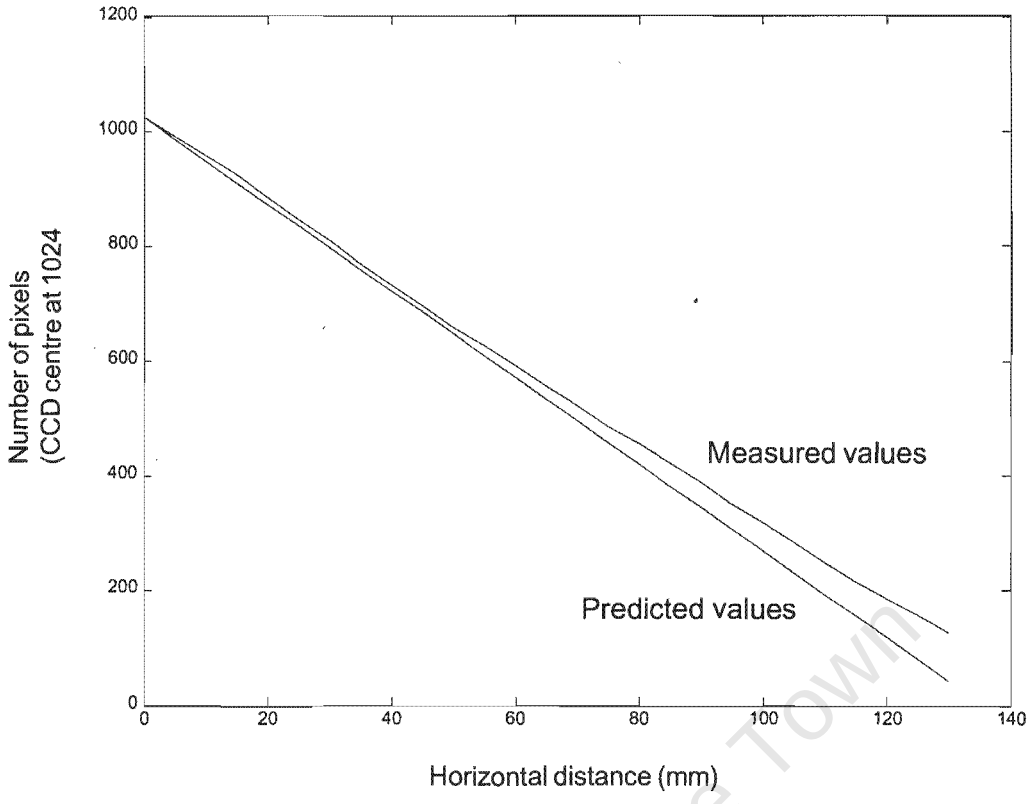


Figure 3.18 Measurements in pixels, taken from a linear CCD and a cylindrical lens and their predicted values from the right hand side of the object space on Figure 3.16.

The data from these two graphs is summarised also in Appendix A.

The next step was to calculate in pixels, the difference between the measured and the predicted values representing the error and to plot them against the horizontal displacement along the sensor. These data points were approximated by a 3rd order polynomial to reduce the random component of the error from the hand positioning of the source when estimating its value and shown on Figures 3.19 and 3.20. This error increases showing an exponential character away from the optical axis, suggesting a possible "barrel" type distortion, similar to that of a spherical lens, but in this case observed only in the direction in which the lens has an optical power.

In Figures 3.19 and 3.20 the horizontal zero line should correspond to the line of the predicted values from Figures 3.17 and 3.18 respectively. The distortion error for the point at 130 mm, for example, using the data from Appendix A could be estimated to be 112.7 pixels for the left side and 125.2 pixels for the right side of the lens. These divided by the total number of pixels on the linear sensor, which is 2048, amount to 11% distortion, for the left side and 12.2% for the right side of this cylindrical lens.

The differences in the measurements between the left and the right side in Figures 3.17 and 3.18 and Figures 3.19 and 3.20 are seen. The reasons for these could be various. In Figure3.20

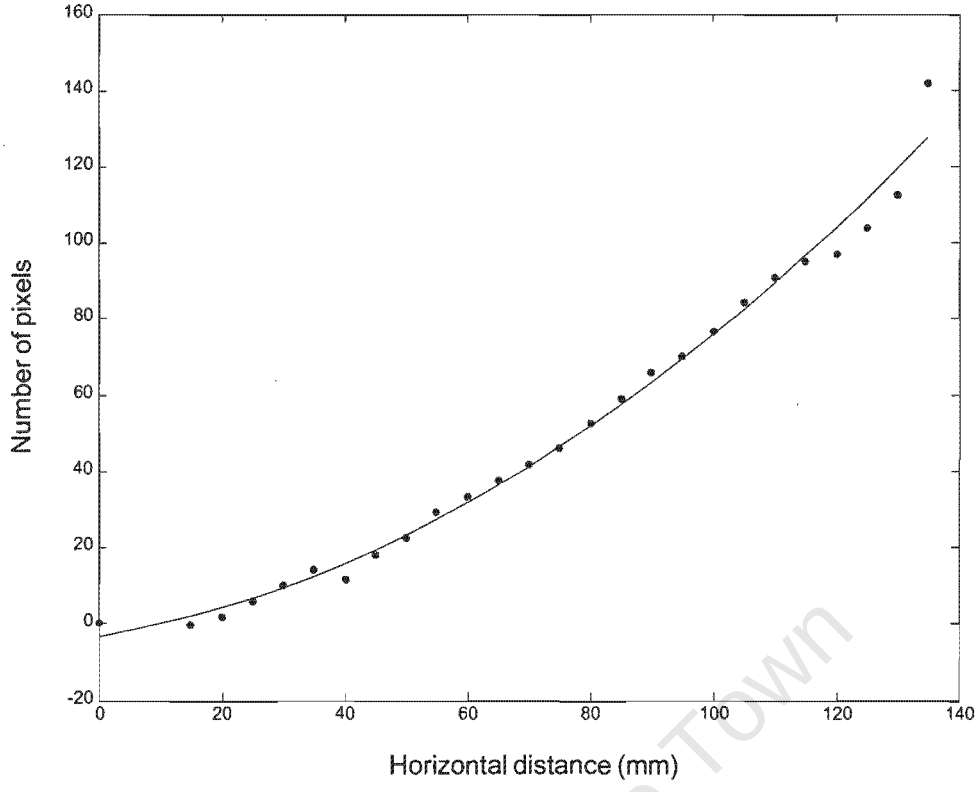


Figure 3.19 The error function of the distortion of a cylindrical lens (in magenta) and its 3rd order approximation (in red) from the left hand side of the object space on Figure 3.17.

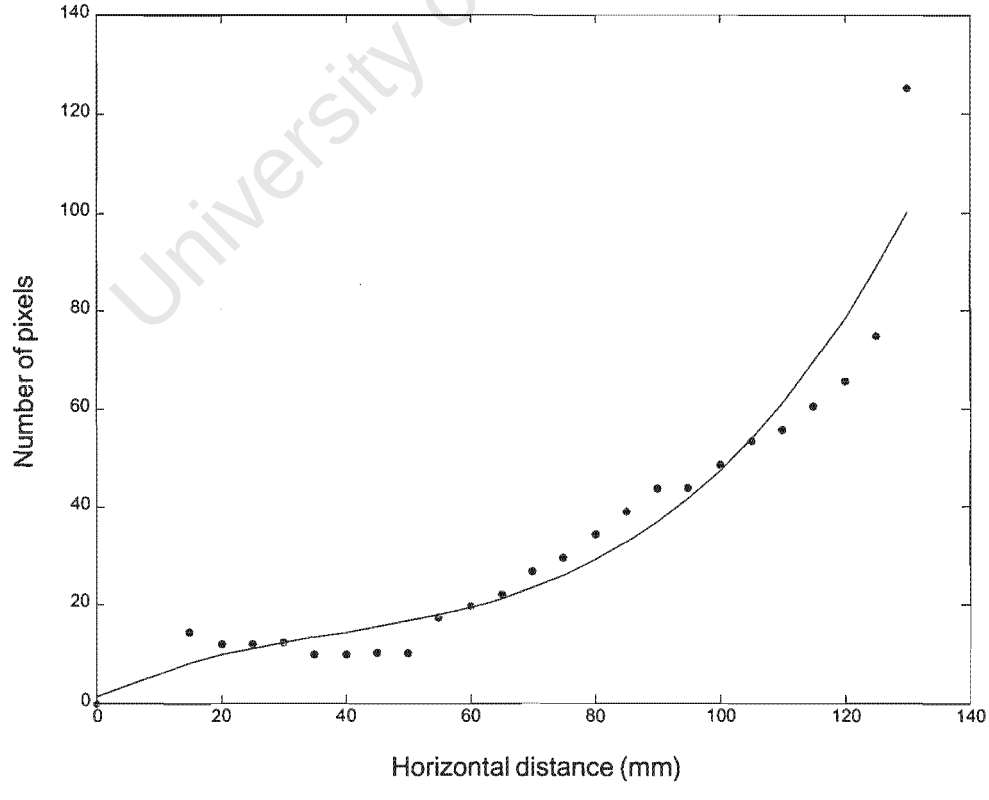


Figure 3.20 The error function of the distortion of a cylindrical lens (in magenta) and its 3rd order approximation (in red) from the right hand side of the object space on Figure 3.18.

where measurements were taken from the right hand side of the lens, the curve in the central area starts with a jump and then behaves as if no distortion occurs. These could be interpreted as hand positioning inaccuracies. This section of the curve is followed by a section corresponding to the side area, where the differences start to grow approaching the shape demonstrated from the curve in Figure 3.19.

The overall displacement of the curve in Figure 3.19 is also bigger compared to that of the other curve in Figure 3.20. This could be due to an imprecision, when the lens was mounted in front of the sensor. A rotation and a translation along the vertical axis combined with another rotation around the optical axis could cause an offset of the whole image to the one side of the sensor. This should be considered a systematic, rather than random error.

Left side					Right side				
Order	Mean pixels	S.D. pixels	R.M.S. pixels	Abs. Error pixels	Order	Mean pixels	S.D. pixels	R.M.S. pixels	Abs. Error pixels
1	0.00	8.28	8.12	27.50	1	0.00	12.85	12.59	50.83
2	0.00	4.12	4.04	13.90	2	0.00	8.58	8.41	32.87
3	0.00	4.12	4.04	14.12	3	0.00	7.57	7.42	24.93
4	0.00	3.61	3.54	10.22	4	0.00	7.18	7.03	20.21
5	0.00	3.17	3.11	7.38	5	0.00	4.00	3.92	13.13
6	0.00	2.25	2.21	5.26	6	0.00	3.25	3.19	10.76
7	0.00	1.76	1.73	4.21	7	0.00	2.45	2.40	7.41
8	0.00	1.21	1.18	3.20	8	0.00	1.71	1.68	4.17
9	0.00	1.20	1.18	3.36	9	0.00	1.30	1.27	2.66
10	0.00	1.20	1.18	3.36	10	0.00	1.19	1.17	2.41

Table 3.2 Mean, standard deviation and absolute maximum error of the differences between the measured pixel data and the approximated data in Figures 3.18 and 3.19, for polynomial of power orders from 1 to 10, for the left and the right hand side of the lens respectivelly.

Similar to the example with the simulated lens, the polynomial approximating the error was tested for orders higher than three. Table 3.2 shows the mean, standard deviation, RMS and maximum absolute error of this approximation for power orders from 1 to 10. From the numbers above it could be seen that the errors are larger in the beginning and decrease with the order of the approximation curve. The process of minimising the error by increasing the order of the polynomial, however is more gradual compared to the test with simulated data The above presented results cover the entire field of view defined by the lens and the sensor, unlike the results from Table 3.1 which are related only to the central part of the lens close to the optical axis.

It is also seen in both curves that the two positions on the sides of the field of view (last two points on the curves), produce considerably more distortion than the rest in the data set, which gives the large initial values of the absolute error on Table 3.2. If these two points are

eliminated from the evaluation, better results would be achieved at lower order of the polynomial approximation.

In order to show consistency from both tests with simulated and measured data their accuracy should be compared using the object-to-base distance related to the lateral distance. Then the differences between measured and theoretically predicted data should appear in the same range. Considering the data from Table A.1 it is seen that for lateral distance of 171.5 and object-to-base distance of 1500 mm the angle of view is 6.52 degrees. The same angle for object-to-base distance of 300 mm in the second experiment will give a lateral distance of 34.3 mm and the closest measured point is at 35 mm. From Table A.1 and A.2 the differences between the measured and predicted pixel values could be found and these will be the estimates for lens distortion. Then the accuracy presented as a ratio related to the object-to-base distances for both tests (according to Allard *et al*, 1995) could be found using the formula:

$$\text{Accuracy} = \frac{\text{object\_to\_base\_distance}}{(\text{D x pixel\_size})} \qquad [3.23]$$

where D is the difference in pixels between measured and predicted data and the pixel_size is 0.014 mm. Thus for the test with simulated data the ratio will be 1:1712, and the test with measured data will give ratios of 1:2143 on the left hand side and 1:1488 on the right hand side. These values for the second test will give an average of 1:1816 and a standard deviation of 1:463.

All these observations are in support of the suggestion from section 3.3.2 that with a longer sensor the distortion on the sides would be greater. With more distortion on the sides of the sensor the random component of the error would increase proportionally, following the non-linear character of the systematic error.

For practical experiments, such as the one targeting to determine lens distortion, the measurements on both sides of the lens should be repeated for certain number of times. Then the data should be averaged prior to the analysis for minimal random errors

These measurements could also be repeated for lenses with different geometry to find out how exactly the distortion would change the shape of these curves. From the theory this aberration is not dependant on the focal length, only on the lateral distance and wider lenses are expected to produce greater position differences on the sides. It is also expected that in these cases higher order approximation might be necessary to compensate the aberration. A further investigation on this problem could provide valuable information on how to work out efficient correction algorithms and how are they going to correlate with the already established methods for distortion corrections for spherical lenses.

3.5 Discussion on the first order theory applied to a cylindrical lens

An example, which considers the angle of the projection of the incident light, is given as additional proof that distortion has to be expected when using a cylindrical lens. This example uses the same sensor-lens and target arrangement from Figure 3.16, where the distance to the target is 300 mm. Since the distortion affects the lateral magnification, the image of the object displaced at 130 mm to the right side, in the horizontal direction, and refracted by the cylindrical lens, will be projected with a slight error to the opposite side in the same direction on the sensor.

The sine function of the angle at which the incident ray falls onto the lens surface, following equation [3.7], could be presented by its power expansion terms. Thus, the value of the sine of the angle θ (Figure 3.21) will vary according to the number of terms used in the approximation. Using Snell's Law,

$$\frac{\sin\theta'}{\sin\theta} = \frac{n}{n'} \qquad [3.24]$$

where θ and θ' are the angles of the incident and the refracted ray respectively, and n and n' are the refraction indices at both sides of the lens surface. It can be assumed that this angle variation will be valid for each refraction of the ray as it makes its way until it leaves the lens.

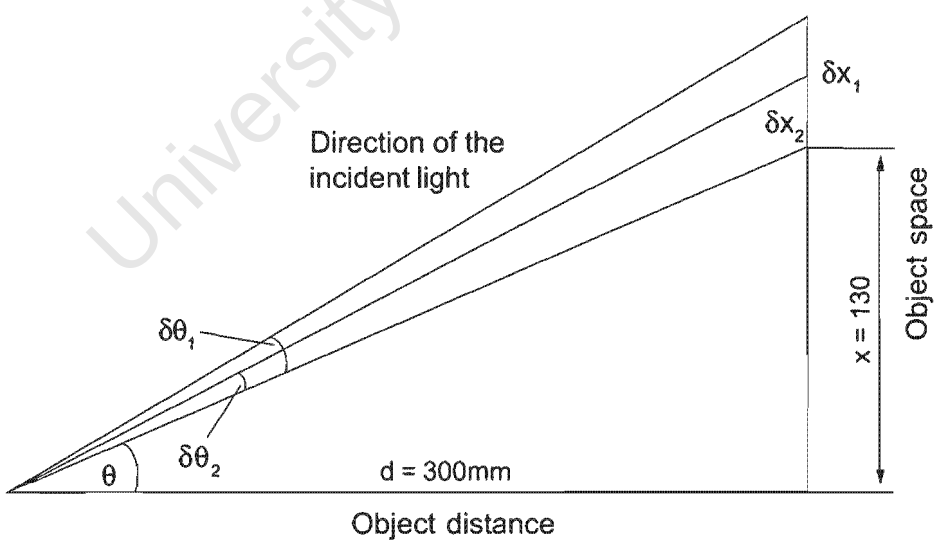


Figure 3.21 Projection of an object horizontally displaced and the error angles δx_1 and δx_2 obtained from the sine function represented by its terms of power expansion.

The introduced error from the angle variation would ultimately affect the image position on the sensor. It is also possible to calculate what the difference in the lateral direction x would be (Figure 3.21) as in the case of Gaussian approximation, where only the first term of the power

series is used, and in the case of the third order theory as well, where the first and the second terms are used. This angle is theoretically derived by:

$$\theta = \text{arctg}\frac{x}{d} = 23.43^0 = 0.13\pi \qquad [3.25]$$

and

$$\sin\theta = 0.3976$$

Then, using the power expansion for $\sin\theta$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \ldots \qquad [3.26]$$

the following approximation for $\sin\theta$ could be enforced using only the first term in the series:

1st case

$$\sin\theta_1 \approx \theta = 0.13\pi = 0.4084$$

$$\delta\theta_1 = \arcsin\theta_1 - \arcsin\theta = 0.67^o$$

The approximation according to the Third Order Theory with the first and second terms will be:

2nd case

$$\sin\theta_2 \approx \theta - \frac{\theta^3}{3!} = 0.13\pi - \frac{(0.13\pi)^3}{3!}$$

$$= 0.4084 - 0.0046 = 0.4038$$

$$\delta\theta_2 = \arcsin\theta_2 - \arcsin\theta = 0.39^o$$

With these angular errors, the displacement errors and using the following relations:

$$\delta x_1 = \frac{d\sin\theta_1}{\sqrt{1-\sin^2\theta_1}} - x \qquad \delta x_2 = \frac{d\sin\theta_2}{\sqrt{1-\sin^2\theta_2}} - x \qquad [3.27]$$

for x = 130 mm, at a distance of d = 300 mm, δx_1 = 4.22 mm, δx_2 = 2.42 mm, and for a longer distance d = 2000 mm, where x = 867 mm, δx_1 = 27.8 mm, δx_2 = 15.8 mm respectively. From these examples it is seen that using two terms in the power expansion instead of one, reduces the error by a factor of two and one has to bear in mind which type of distortion is present to adjust the correction accordingly. The earlier measurements from a 19 mm wide cylindrical lens revealed that the average distortion over one side of the 25 mm long CCD sensor was about 12% . This was compared to the 300 μm radial distortion on the side of the image format of a 10 mm long sensor produced by a 25 mm lens for a conventional video camera, which was 6%. Thus the size of the field stop, created by the sensor length, is a significant factor which influences the systematic error described in section 3.3.2, as well as the measured random error presented in section 3.4.

3.6 Intensity change when using cylindrical instead of spherical lens

Another important characteristic when using cylindrical instead of spherical lenses, besides the effect of optical aberrations, is the loss of intensity due to the conversion of light. A test to detect passive retroreflective targets was performed to find out what change is produced in the signal-to-noise ratio when using a different type of lens. For this purpose two passive targets were placed about 4 m away from the camera and illuminated with a 6 W infrared (IR) source consisting of multiple light emitting diodes (LEDs). The test of the intensity from passive targets used two different types of lenses (cylindrical and spherical), mounted on a commercial video camera JVC TK1270 with a sensor measuring 768x576 pixels and a diagonal of 0.5 inch. The images were recorded on a videotape and later transformed into a 640x480, 16bit grayscale digital format, using a frame grabber Matrox Meteor II.

First a spherical lens was mounted onto the camera and an image of the targets was captured (Figure 3.22a). Then the lens was replaced with a cylindrical one and an image captured once again (Figure 3.22b).

a b

Figure 3.22 Images from two passive targets captured with the same camera with: (a) spherical lens; and (b) cylindrical lens.

There was a clear difference between the intensity levels in both pictures and it was necessary to apply further contrast enhancement to the second image to make the visual recognition of the target patterns possible. The next step was to perform a quantitative analysis of the images. Cross lines were taken from both images and plotted together on Figure 3.23, presenting the pixel values as 16 bit numbers which showed grey scale levels between 0 and 65535.

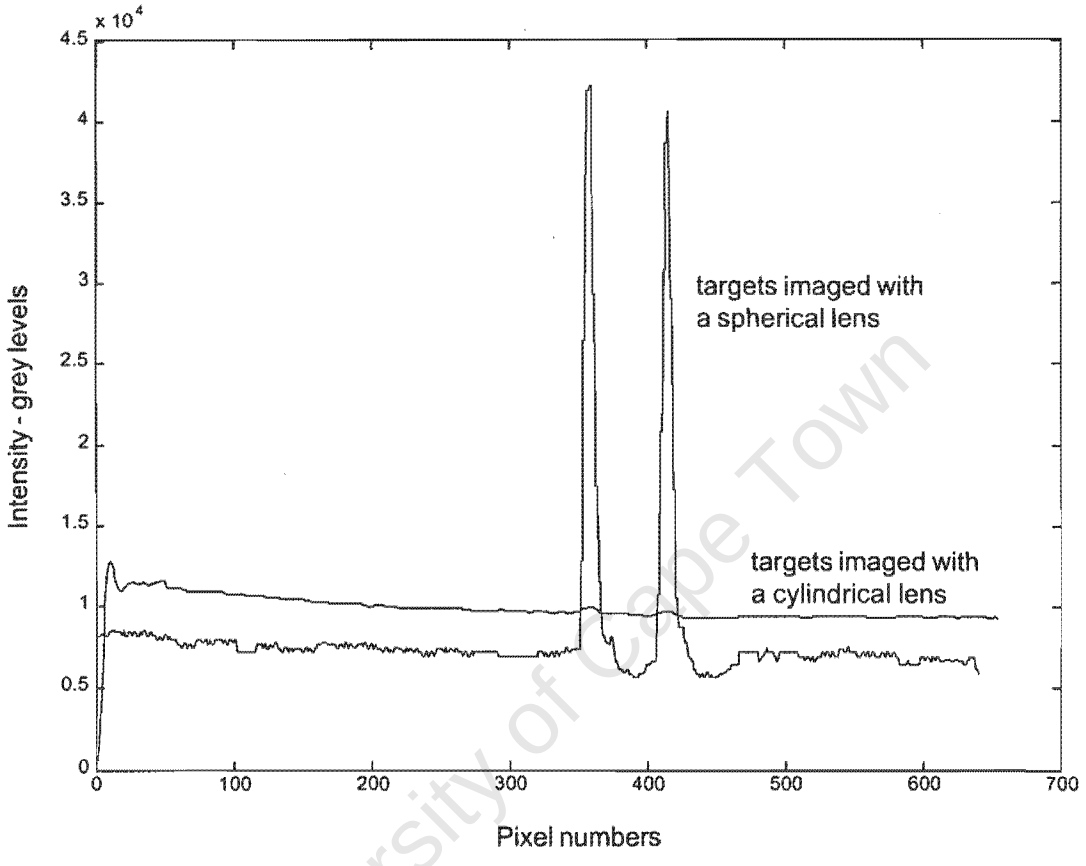


Figure 3.23 Line profiles from the images from Figure 3.22 showing the difference in the intensity levels. On the horizontal axis the pixel numbers along the image are given from 0 to 640, while on the vertical axis the intensity grey scale levels from 0 to 65535 are plotted.

The two lines on the graph showed the positions of the targets along the line and their intensity values. The two peaks from the first image are very clearly distinguishable with intensity levels of over 40000. For the second image it was required to calculate the sum of as many as 300 lines and average their values in order to recognize the targets as small peaks appearing above the line of the background. The elevation of the line representing the background at the extreme right on the graph, showed a very strong ambient light reflection, which was stronger than the reflection from the targets. After the effect of this reflection was eliminated it was found that the grey scale level, corresponding to the intensity of the targets captured with a cylindrical lens, was slightly higher than 9700.

Subsequently the background from both images was removed and the intensity of the pure signal calculated. The averaged peak values in grey scale levels were 33918 and 334 for the

spherical and the cylindrical lenses respectively. From this the coefficient of attenuation in the patterns from the two lenses was found to be:

$$K_{att} = \frac{I_s}{I_c} = 101.55 \qquad [3.28]$$

where

I_s - intensity of the pure signal from the image taken with a spherical lens
I_c - intensity of the pure signal from the image taken with a cylindrical lens

Another comparison was done using the signal to noise ratios:

$$SNR_s = \frac{I_{ts}}{I_{bs}} = 5.546$$

$$SNR_c = \frac{I_{tc}}{I_{bc}} = 1.035 \qquad [3.29]$$

where SNR_s is the signal-to-noise ratio obtained from the spherical lens and SNR_c is the signal to noise ratio obtained from the cylindrical lens, I_{ts} and I_{tc} are the intensities of the signal from the targets from each lens, I_{bs} and I_{bc} are the intensities of the signal from the background from each lens.

The attenuation coefficient between the two types of lenses showed considerable loss of light intensity due to the conversion of the light from a point into a line. The amount of light on the image with the cylindrical lens was distributed along the two vertical lines from the targets. It was still possible to acquire this converted power, with the rectangular sensor used in this experiment by averaging multiple lines. However, if it had been replaced by a linear sensor, only a small portion of the area equivalent to a single line would have been scanned.

The signal-to-noise ratios revealed that the spherical lens performed better on noisy signals such as the reflected signal from passive targets. It could also be argued that it has better focusing capabilities compared to the cylindrical lens due to its two-dimensional curvature which provides more optical power.

These quantifications have pointed out the main advantages of using spherical lenses to detect passive retroreflective targets. If cylindrical lenses are to be used, the need for a strong

illumination, parallel to the optical axis, to provide the best reflection from the targets, has to be considered. Additional signal processing techniques should also be applied to reduce the noise. In spite of this, target disappearance would still be possible, which would make the eventual identification and tracking process more difficult.

A combination between a spherical and a cylindrical lens which has optical power in both horizontal and vertical direction is called a toroidal lens or sometimes astigmatic lens. This type of lens has different radii of curvature along the two axes. The important optical effect of projecting a point source through such a lens comes from the two optical aberrations: astigmatism and curvature of field. The two orthogonal curvatures define two different focal distances and the images obtained at each focal point would be straight lines orthogonal to one another. The actual image would appear as shown on Figure 3.9 in the middle between the focal points and will be deformed according to the optical power along each axis. Thus, a point source would be projected into an ellipse with its major and minor radii corresponding to the optical power in that direction. All the rest of the five Siedel aberrations will also appear when using such a lens and their effect will be stronger compared to that obtained from a cylindrical lens. The major benefit of combining a toroidal lens with a linear sensor is that more light will be focussed at the image location which is intersected by the sensor and this will increase the intensity. From this effect a better signal-to-noise ratio would be obtained which ultimately could solve the intensity problem when passive targets are detected. Toroidal lenses also reported as anamorphic lenses were used in the applications by Leo and Macellari (1981), Macellari (1983) and Bianchi *et al.* (1990).

Considering the main physical characteristics of the sensors in 3.2.1 as well as the requirements of using lenses it could be pointed out that linear sensors compared to rectangular sensors have number of advantages such as better resolution and faster output. However, significant loss of intensity occurs when linear sensors are combined with cylindrical lenses. Linear sensors could contribute from alternative methods to image targets, such as the application that uses a projection mask on Figure 2.16. In the case of a lens conversion of the light, these sensors would prefer active targets rather than passive targets, unlike the rectangular sensors equipped with spherical lenses, which could work fine with both targets. Lens distortion is a factor that should be accounted for when both types of sensors are used and further studies might be necessary to determine the exact amount of distortion for cylindrical and toroidal lenses with different geometry.

Chapter Four

Passive versus Active Targets and Data Processing

This chapter will consider several issues related to the nature of linear sensor signal detection and data processing. Firstly, the effect of light conversion from the cylindrical lens on the signal will be presented as well as the method of determining the pixel position over the sensor. This will be followed by a linear resolution test and some explanations about filtering and other signal processing techniques. The next subject will be a comparison between active and passive targets, determining their respective use. The final part of the chapter will focus on the methods to separate multiple targets from one another and to improve intensity when the quality of the signal is poor.

4.1 A technique for signal acquisition using a cylindrical lens

Light conversion from a cylindrical lens was initially tested with a rectangular CCD camera with 512 x 512 pixels, where an active marker was imaged then digitized by a frame grabber and analyzed. The signal was taken from a single cross line from the rectangular image shown in Figure 4.1 and plotted. The graphic in Figure 4.2 shows a profile of the marker, which was an Infrared Light Emitting Diode (IR LED), where the intensity is plotted versus the length of the CCD sensor which is given as the number of pixels. The intensity is represented by an 8 bit grey scale, thus providing 256 grey levels.

Figure 4.1 Image from an IR LED taken with rectangular camera and a cylindrical lens, showing the light conversion from a point into a line.

The signal contains noise due to light reflection from the background and errors due to the analogue-to-digital conversion. Filtering is therefore suggested as a necessary step at the data processing stage. The picture also shows that with a well focused beam of light, the detected marker creates a peak over 20-25 pixels. A linear sensor consisting of a few thousand elements will provide very good linear resolution, which is better than the rectangular CCD

with only 512 pixels along one side. In this example it is assumed that the intersection point between the CCD sensor and the line from the focused light is found at the maximum of the profile, which is the position of the marker on the sensor.

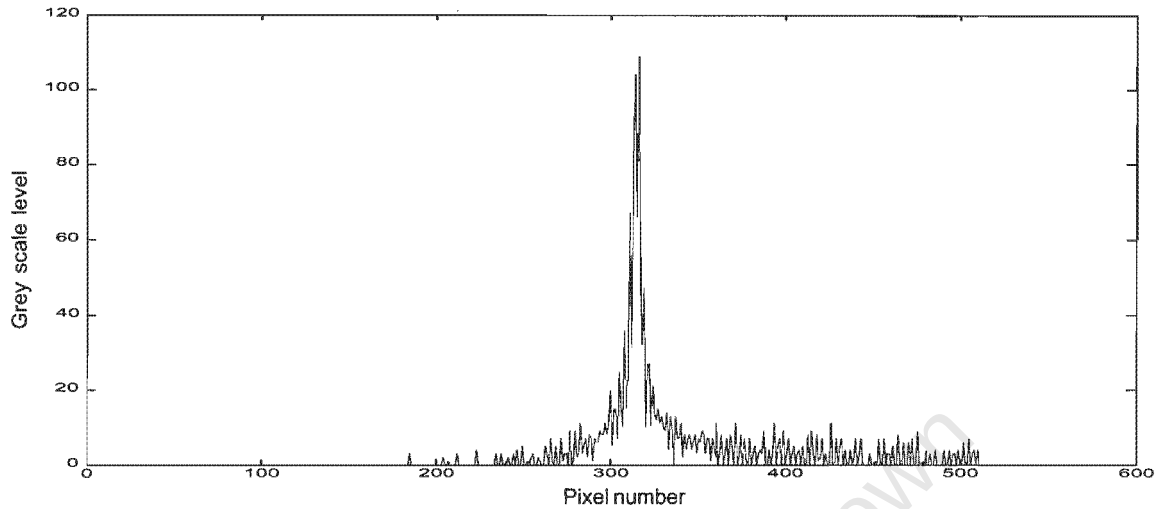


Figure 4.2 Profile of the image from Figure 4.1 taken from an active target positioned about 3 m away from the camera.

4.2 Filtering and centroid algorithm to improve linear resolution

4.2.1 Requirements of the different types of filters for noise reduction

It is reported in many literature sources that subpixel resolution can be achieved if the centre of gravity (centroid) value of the peak is calculated. Such an algorithm can improve the linear resolution to 1/50 of a pixel. Another way to improve the resolution is to apply filtering which reduces the noise followed by the calculation of the centroid. There is a large variety of filters that can be used for this type of signal such as the Finite Impulse Response (FIR) filters, Infinite Impulse Response (IIR) filters and some filters based on arithmetic operations. "Statistic order" type filters, such as the average and the median filter, compared to the convolution types (FIR, IIR), are more efficient when the objective is to reduce the noise without a delay. The median filter, compared to the average filter, is less sensitive to rapid changes in the amplitude of the signal .

Convolution type filters, unlike the "statistic order" filters, produce phase shift. The effect is a change in the positions of all points, including those of the peak, in a direction which moves the signal backwards as a result of the delay. This phase shift has to be compensated for further in the process of evaluating the peak position by moving it back by a number of pixels equivalent to the initial phase shift. The definition of this parameter is given in Appendix B where FIR and IIR filters are discussed. These two types of filters compared to the "statistic

order" type, require more computational power but perform better on high frequency noise. Their magnitude response, phase response and group delay are shown in Figure 4.3 and Figure 4.4. It is desirable that the phase response of the filter is linear to facilitate finding the compensation in sample points. FIR filters could be designed so that they have a linear phase response and a constant group delay (Figure 4.4b). On the other hand, a Butterworth filter, which is an IIR filter, is not stable when the condition for linear phase response is implied in the design. This is derived from the fact that under such conditions, for any pair of poles of its transfer function falling within the frequency unit circle, there exists a pair of zeros outside the circle which makes the system unstable.

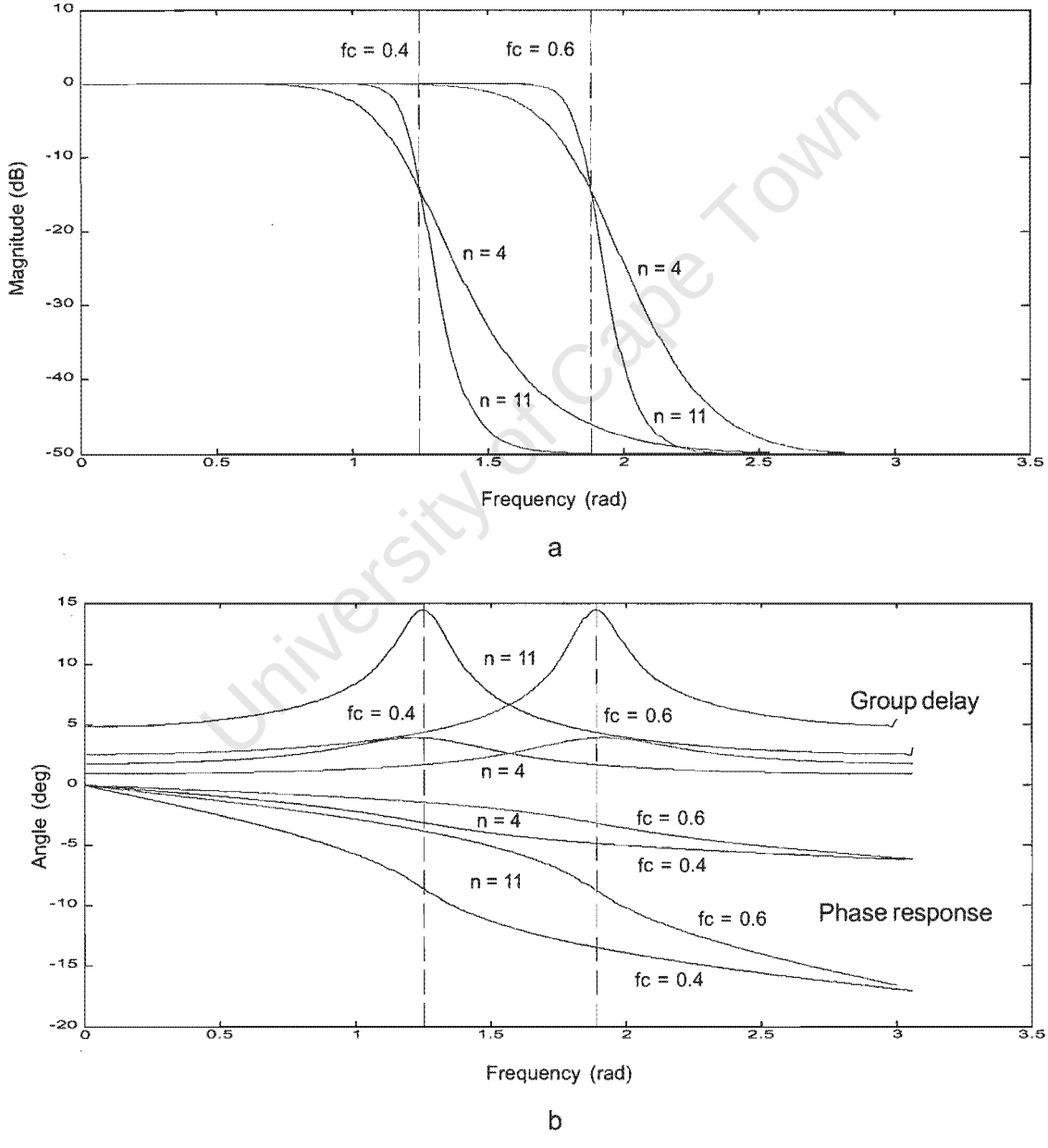
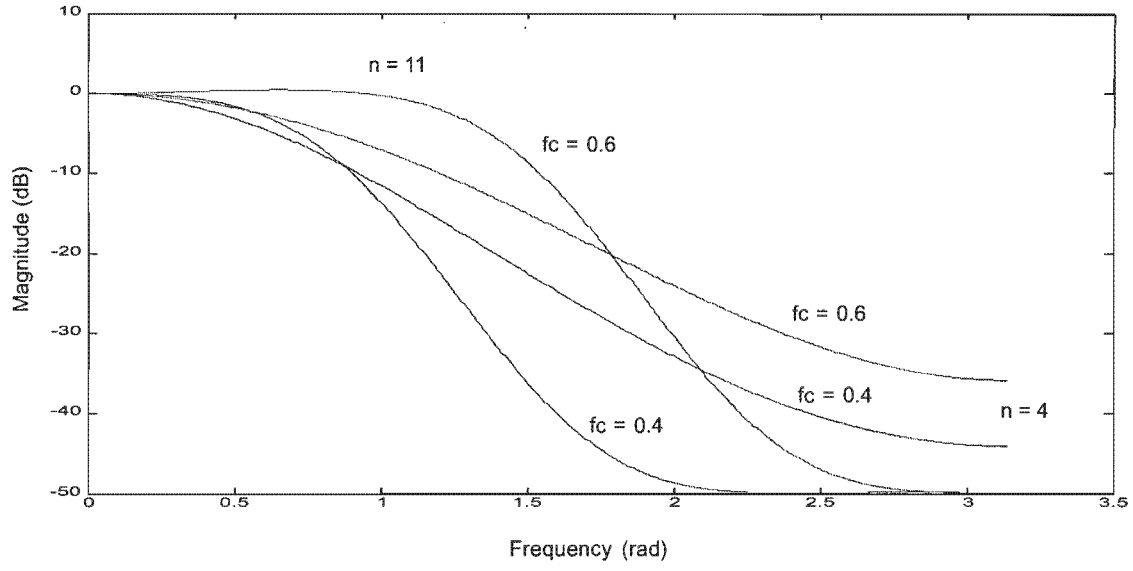
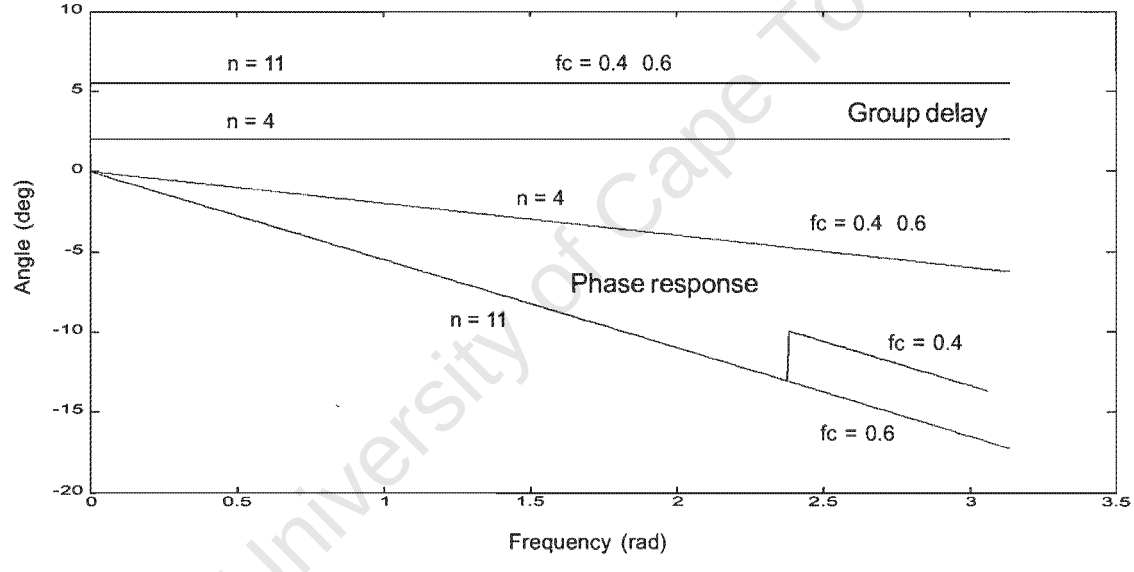


Figure 4.3 Magnitude response (a), phase response and group delay (b) of a Butterworth filter of order 4 and 11 and cutoff frequencies 0.6 and 0.4.



a



b

Figure 4.4 Magnitude response (a), phase response and group delay (b) of a Finite Impulse Response (FIR) filter of order 4 and 11 and cutoff frequencies 0.6 and 0.4.

On Figure 4.4 the phase response of the filter of n=11 and fc=0.4 presents a point of singularity. However, the working range of this filter falls inside the pass-band, which is for the frequencies below 1.3 rad. One practical rule of choosing the width of the pass-band is to calculate it for attenuation of 3 dB.

4.2.2 Centroid algorithm

The linear resolution significantly improves when the centre of gravity (centroid) of the peak, rather than its maximum value, represents the position along the sensor. The formula to calculate the centroid is described as follows

$$C_n = \frac{\sum n.P_n}{\sum P_n} \qquad [4.1]$$

where,

n = pixel number on the sensor
P_n = value (grey level) of pixel number n

In this case only the pixels above a certain level should be chosen for the calculation to give a true representation of the centre of gravity. This threshold technique, illustrated in Figure 4.5, eliminates the background where the level at which to cut the signal is usually recommended to be a fraction of the maximum peak value (Gush, 1996). One effective threshold level for active targets would be 1/3 of the peak maximum and this value was used further in this research, as well as other different proportions of the peak maximum when passive targets were tested.

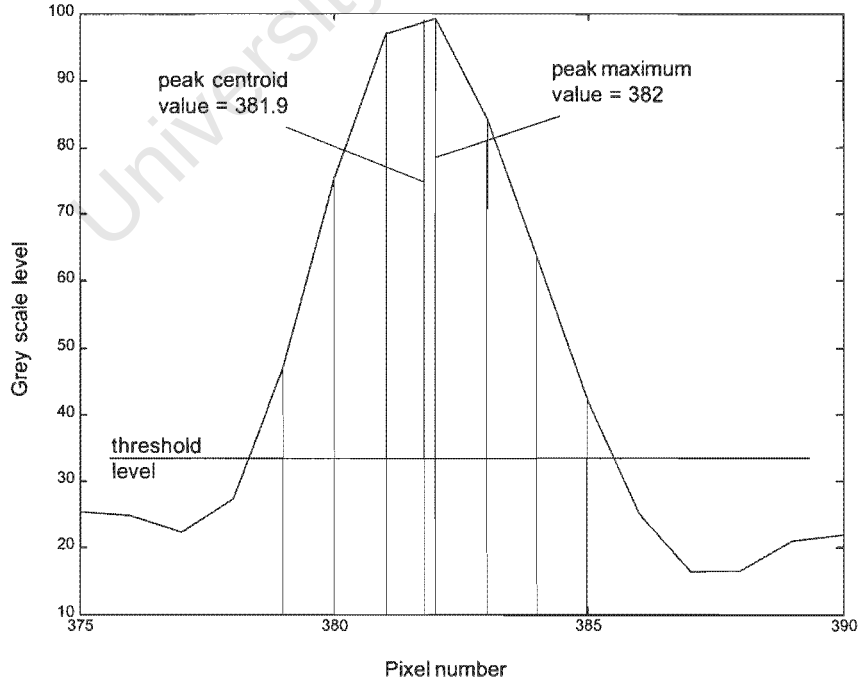


Figure 4.5 Peak shape of an active target showing the maximum peak value and its centroid value calculated from the pixel values above the threshold level.

To check the effectiveness of the described data processing techniques, which includes filtering followed by calculation of the centroid value of a target's position, linear resolution tests were performed. A rectangular CCD sensor mounted in a commercial camera JVC TK1270 and equipped with a cylindrical lens with geometrical parameters described in section 3.4 was used to measure one active marker (infrared LED) positioned 2 m away from the camera. The marker was moved in a direction parallel to the horizontal axis of the camera and images similar to that shown in Figure 4.1 were captured by the sensor. Space intervals were set at 5 mm, 1 mm, 0.5 mm and 0.1 mm to prove that the system could distinguish the positions of these points. For this test three different filters were used: (a) a median filter, which is a "statistic order" type, with apertures of 3, 4, 7 and 8 points, so that there were two small and two large sample apertures of odd and even number of points each; (b) a Butterworth filter of order n=4 and n=11, at two different cutoff frequencies fc=0.6 and fc=0.4; (c) a FIR filter also of the same order and cutoff frequencies. The marker was aligned to the camera in a similar way described in section 3.4, where the distortion of the cylindrical lens was measured. A setup of metal bars was constructed to provide the direction along which the marker was moved and another direction orthogonal to this, coinciding with the optical axis. Then the marker was clamped on a vernier caliber and aligned to the bar parallel to the lens surface and moved along the preset intervals.

After the marker positions were measured the data sets forming the peaks of the markers were filtered and plotted to show the shift of the marker along the horizontal line of the sensor. Figure 4.6 and Figure 4.7 show the positioning of these peaks for displacement step of 0.5 mm and 0.1 mm respectively, where line profiles from all marker positions were plotted together. The horizontal axis on both figures shows the absolute pixel number on the sensor, where 0 is the left most pixel and the total number of pixels is 512.

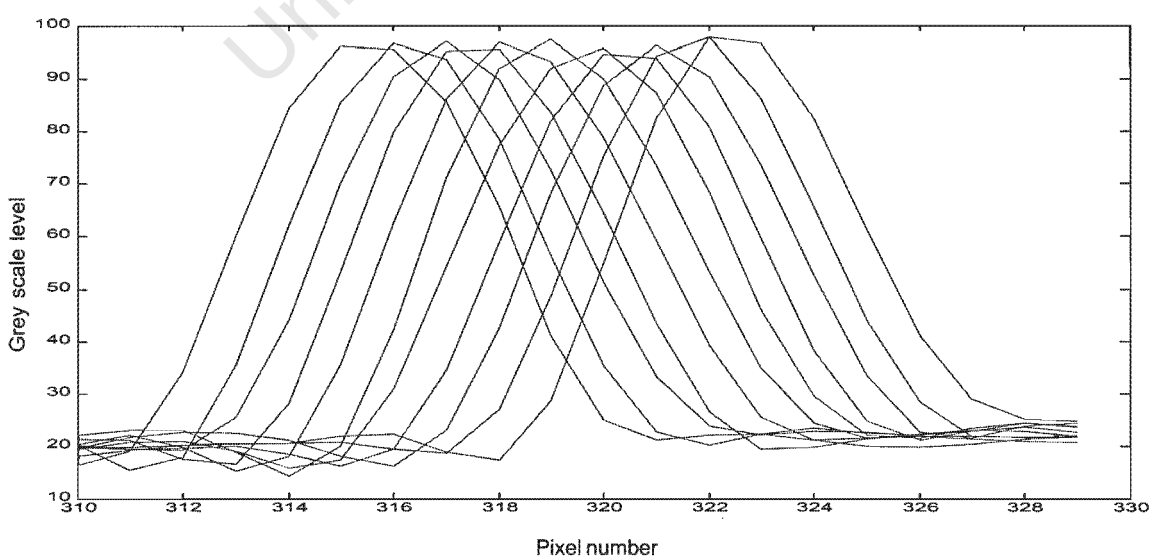


Figure 4.6 Marker positions at 0.5 mm interval (0 - 5 mm) pixel value = 315 - 322, applied 11th order FIR filter (fc = 0.4).

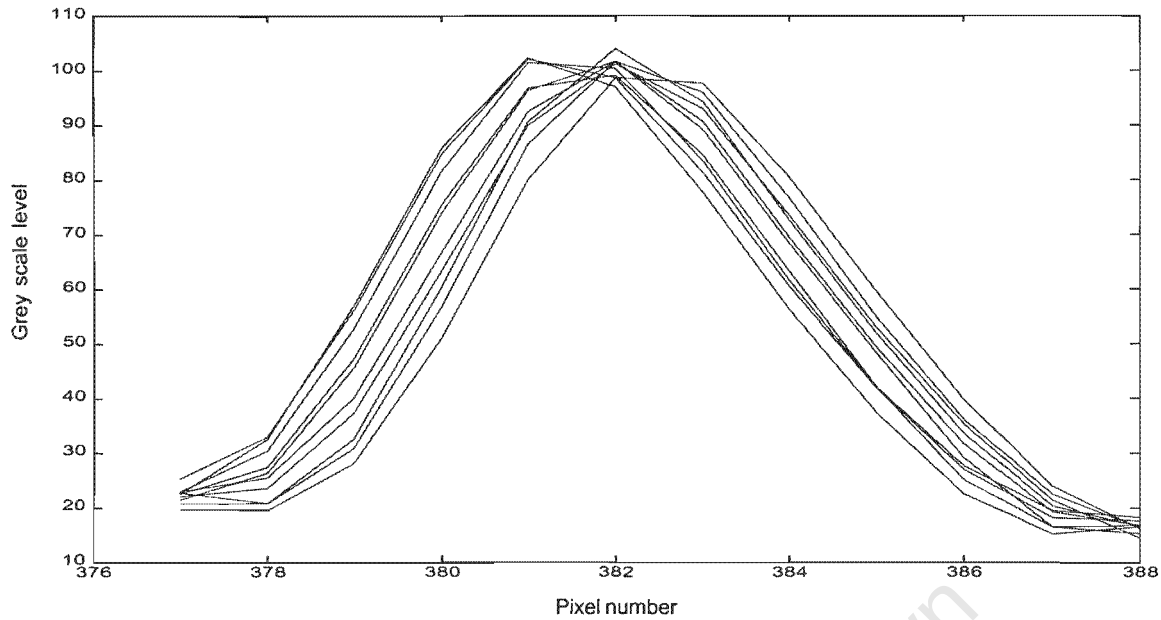


Figure 4.7 Marker positions at 0.1 mm interval (50 - 51 mm) pixel value = 381 - 382, applied 4th order Butterworth filter (fc = 0.4).

Due to a hardware functionality reason, the image has a total shift to the right hand side, so that the centre of the camera lens system appeared at pixel number 310 instead of at 256. The vertical axis on each graph shows the intensity of the signal from the marker in grey scale levels, where the maximum value was 255 corresponding to the 8 bit analogue-to-digital conversion of the signal intensity.

The corresponding peaks for each measured position are clearly distinguished in Figure 4.6. In Figure 4.7, however, the displacement step was too small and the peaks corresponding to several successive positions are grouped on top of one another, causing their maximum values to occupy the same pixel position. In this case the small peak displacements have produced changes mostly in the peak shape, rather than an actual position change along the sensor. This example clearly shows the necessity of more accurate position determination such as the centroid algorithm.

After filtering, a threshold of 1/3 of the peak maximum value was subtracted from the signal, leaving only these pixels with high grey scale values for the calculation of the centroid. The results from the 0.1mm linear resolution test are shown in Tables 4.1 through to 4.3. In Table 4.1 the linear displacement measured in millimetres is given in the first column. In the second column the maximum pixel values are shown and in the third column are given the centroid values calculated from the raw pixel data. Then in the same order the results are presented after a median filter with an aperture of 3, 4, 7 and 8 sample points was applied to the raw data. First, the maximum pixel values from the filtered data are shown, followed by the centroid values.

Raw data			median filter							
0.1mm			n = 3		n = 4		n = 7		n = 8	
mm	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
50.1	379	379.88	380	379.29	380	380.07	378	379.30	379	380.08
50.2	379	379.57	380	379.90	380	380.12	379	379.27	379	379.78
50.3	379	379.94	379	379.67	380	380.17	379	380.25	379	380.45
50.4	379	380.00	379	379.72	381	380.51	379	379.72	379	380.45
50.5	380	380.38	380	380.03	380	380.21	380	380.02	380	380.24
50.6	380	380.41	380	380.37	380	380.25	380	379.75	380	380.26
50.7	379	378.16	379	380.73	380	380.95	379	380.79	379	381.04
50.8	380	380.43	380	380.41	380	380.63	380	380.46	380	380.97
50.9	380	378.57	379	380.43	380	381.26	379	380.78	379	381.32
51.0	381	380.54	380	380.49	382	381.37	380	380.79	380	381.35

Table 4.1 Linear space intervals at 0.1 mm, the corresponding pixel values and the centroid values of the target moved along these intervals, after a median filter was applied to reduce the noise on the signal. The pixel and centroid values are given in ascending order for each filter aperture.

From this data it can be seen that for each gradual 0.1mm displacement of the marker in the first column, there are values corresponding to each peak maximum. These pixel values do not follow an exact ascending order. The same situation occurred after the centroid values were calculated from the raw pixel values in the third column. From the filtered data it can be seen that no phase shift was produced by the median filter. The filtered pixel values were moved equally to the left and the right of the position giving the middle of the measured distance due to the smoothing effect. This was also observed with the centroid data calculated after each filtering using a different aperture. The differences between the pixel values from each step considering the raw and the filtered data mainly occur because of the different shapes of the peaks, causing the maximum pixel value to move one pixel either to the left or to the right hand side. After the centroids were calculated, these values were arranged in a certain order. The expectation was that after filtering, the noise from the data would be removed and the centroid data would be sorted in ascending order. However, none of the median filters that were used could sort the data in the exact order.

Table 4.2 shows the maximum pixel values and centroid values after a Butterworth filter was applied. The phase shift has moved these values from their initial positions (Table 4.2, column 2) at a certain number of pixels, to the right along the main direction of the sensor. The number of pixels in the phase shift depends on the order of the filter and the cutoff frequency and it is related to the nonlinear character of the phase response explained earlier. This has to be compensated for in such a way that the displaced pixels in the direction of the shift are subtracted back. This compensation will be different if different filters are used, but has to remain constant for all the points in the entire data set. From this table it can be seen that the Butterworth filter of the lower order (n = 4) is able to sort the values for both frequencies, although, the higher order filter (n = 11) could only solve the problem at the lower frequency (fc = 0.4).

Butterworth filter							
n = 4				n = 11			
fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
380	380.46	381	381.46	382	382.01	385	384.89
380	380.50	381	381.51	382	382.39	385	384.93
380	380.84	381	381.85	382	382.40	385	384.97
381	380.90	382	381.90	382	382.80	385	385.03
380	380.95	382	381.95	382	382.84	385	385.07
381	381.32	382	382.01	383	382.88	385	385.14
381	381.37	382	382.06	383	382.94	386	385.50
381	381.40	382	382.07	383	382.93	386	385.52
381	381.45	382	382.13	383	383.00	386	385.88
381	381.82	382	382.84	383	383.41	386	385.95

Table 4.2 A Butterworth filter was applied to the raw signal and the pixel and centroid values are given in ascending order for each filter order n and cutoff frequency fc.

In Table 4.3 the data after filtering with a FIR filter is presented. This filter is more convenient because it has a linear phase response and a constant group delay. The effect of this filter is similar to the Butterworth filter, where again the lower order (n = 4) solves the linear displacement problem at both cutoff frequencies and the higher order (n = 11) is only efficient at fc = 0.4. Additional tables with complete results from all these resolution tests are printed in Appendix C.

FIR filter							
n = 4				n = 11			
fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
381	381.52	381	381.51	385	385.29	385	385.31
381	381.54	381	381.54	385	385.33	385	385.35
381	381.59	381	381.58	385	385.36	385	385.38
381	381.97	381	381.64	385	385.43	385	385.44
382	382.35	382	382.33	385	385.49	385	385.50
382	382.38	382	382.38	385	385.54	385	385.56
382	382.41	382	382.41	385	386.23	386	385.93
382	382.43	382	382.42	385	385.90	386	385.93
382	382.46	382	382.47	386	386.26	386	386.29
383	382.55	383	382.56	386	386.37	386	386.38

Table 4.3 A Finite Impulse Response (FIR) filter was applied to the raw signal and the pixel and centroid values are given in accending order for each filter order n and cutoff frequency fc.

In Figure 4.8 three sets of points are shown: the raw data, the raw data after filtering with an 8 point median filter and the raw data after filtering with a low order (n = 4) FIR filter. All the data was taken from the 0.5 mm resolution test.

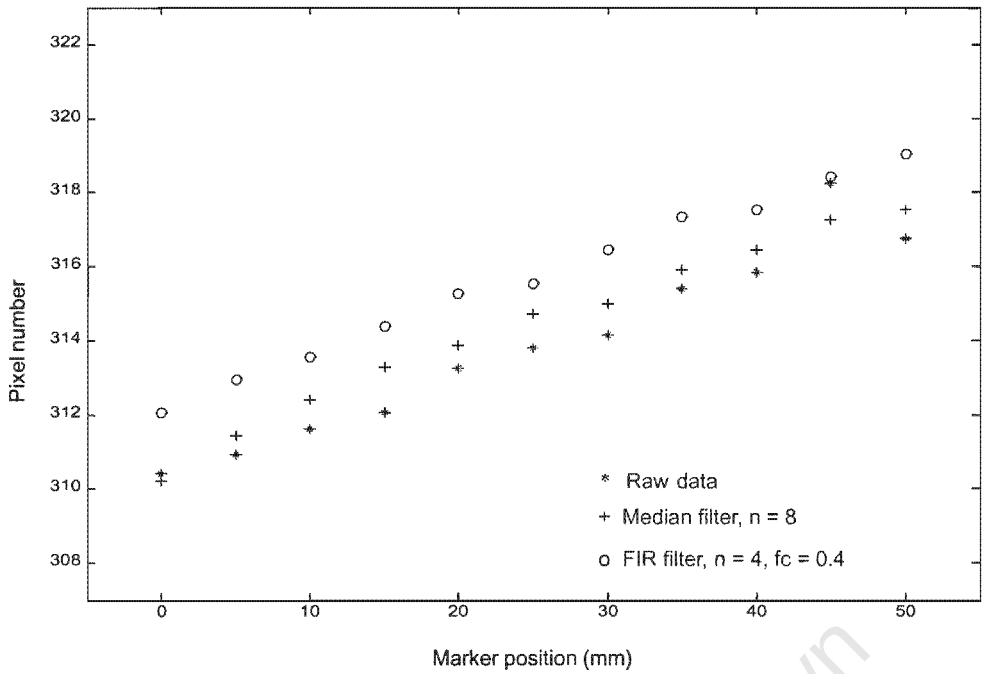


Figure 4.8 The raw and filtered data with two different filters from the 0.5 mm resolution test plotted as marker position vs centroid pixel position.

In a similar way the raw data from Table 4.1 (columns 1 and 3) are plotted together in Figure 4.9 to show the linear displacement in the object space (horizontal axis) and in the sensor space (vertical axis). It is noticeable that some of the points do not follow the order, bringing discontinuity in the linear relationship. However improvement can be seen after different filters have been used: poor results after median filtering, slight discontinuity after high order (n = 11) Butterworth filter and full linearity achieved after low order (n = 4) FIR filtering. The phase shift from the convolution filters (Butterworth, FIR) is well distinguished for the two upper data sets.

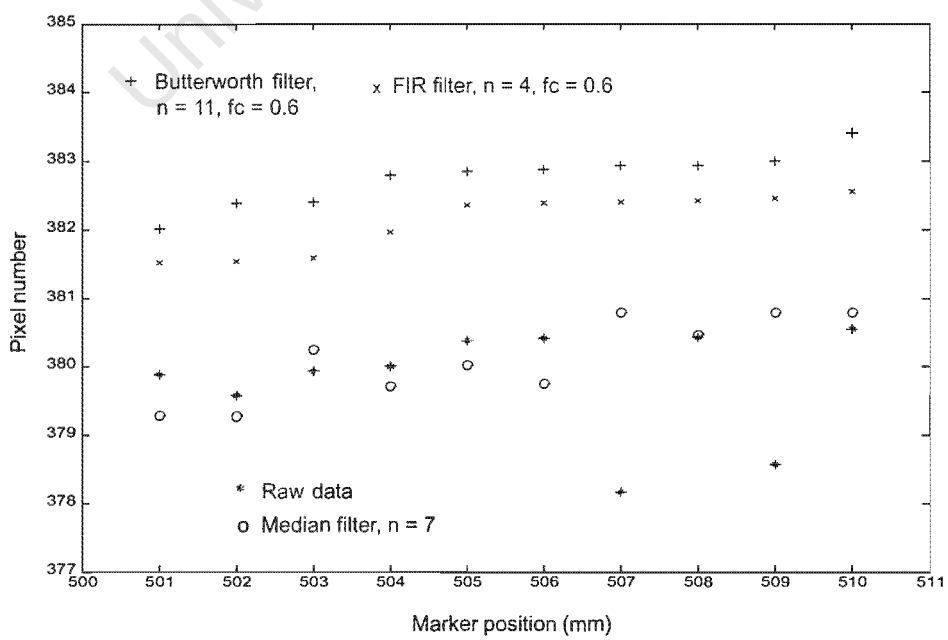


Figure 4.9 The raw and filtered data with three different filters from the 0.1 mm resolution test plotted as marker position vs centroid pixel position.

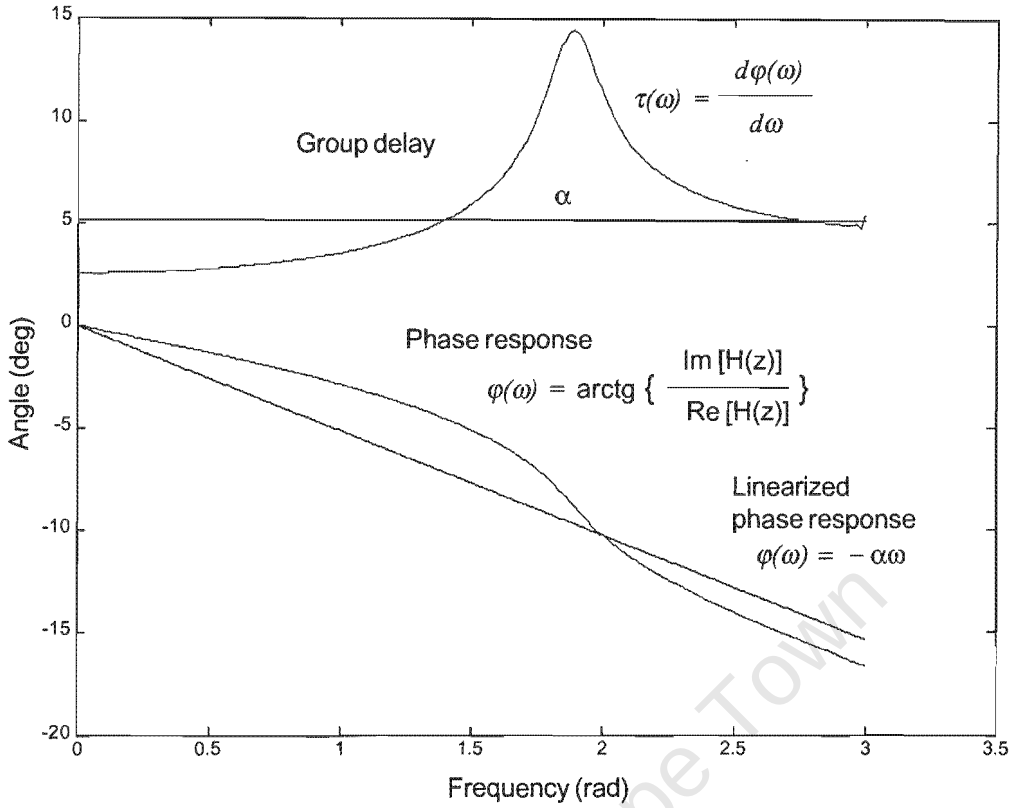


Figure 4.10 The actual and the linearized phase response and group delay of a Butterworth filter.

According to the theory in Appendix B in the case of FIR filter the phase shift can be easily determined for filters of order n = 4 and n = 11 using [B.3]. Then using [B.8] it can be found that for n = 4 it will be 1.5 sample points, while for n = 11 it will be 5 sample points. The number of points is equivalent to the number of pixels which must be subtracted from the calculated centroid value to obtain its true position. For a Butterworth filter, first the linearized phase response has to be obtained, as shown in Figure 4.10, and from there its first derivative according to [B.13] will give a constant number of points to be used further for compensation of the centroid position. Following this approach it can be found that for a filter of n = 4 and cutoff frequencies fc = 0.6 and fc = 0.4, the phase shifts are 1.9 and 2.3 points or pixels respectively. For a filter of n = 11 and cutoff frequencies fc = 0.6 and fc = 0.4 these shifts will be 5.1and 6.3 points or pixels respectively.

When compensating for the phase shift, all mentioned non-linearities should be accounted for to measure the position accurately. For some IIR filters the phase response may need to be approximated by not one, but two or more straight line sections to avoid large linearization errors (shown by the curve and the straight line of the phase response in Figure 4.10). Using a single line approximation could lead to miscalculation of the correct position of the target on the sensor. For multiple targets (or for sequential measurements of a single target) these compensations will contain the same displacement errors which would not disturb the target

identification and the tracking process, but the definition of the 3D coordinates could be significantly affected in terms of accuracy.

It could be concluded that for real time applications the statistic order filters, such as the average and median filters are more convenient and faster. They could be recommended in cases when the linear resolution is required to be in the range of 1 to several millimetres. For more precise measurements the convolution type filters should be used and the advantages of FIR filters especially for their linear phase response should be emphasised. The convolution types filters compared to the statistic types, however, are slower if their response is to be implemented by algorithms. There could also be expected more time delay for the phase shift compensation. One faster solution to this problem is to implement these filters in hardware using digital circuits.

Using an approach that entails filtering and centroid algorithm to remove the noise at the signal acquisition stage, the linear resolution of the sensor is considerably improved. The position differences of the test with the finest displacements vary from a few tenths of a pixel to one or two hundredths of a pixel, which confirms the observations from earlier theoretical reports. The experimental systems described by Macellari (1983) and Bianchi *et al.* (1990) in Chapter 2 report resolutions from 1/4000 to 1/16000. These systems were based on a 2048 element linear CCD, which for the first value corresponds to an improvement of half a pixel, while for the second value it is one eighth of a pixel. It is necessary to note that these numbers were also related to the performance and the range of the analogue-to-digital converters, which provided the grey scale levels of the measured signal and the range of the transferred numbers. The devices used were 12 bit and 14 bit respectively. The later commercial system for 3D position measurement, OPTOTRAK from Northern Digital Inc., claimed a linear resolution of 1/200 000. Considering the same 2048 element sensor this means that an improvement of about one hundredth was achieved by using sub-pixel resolution techniques.

4.3 Advantages and disadvantages of using passive and active targets

There are two main types of targets that can be used for tracking: active targets, where for the purpose Light Emitting Diodes (LED's) are fired in front of the detecting sensor, and passive targets made of retro-reflective material, illuminated by light and reflecting back to the camera. In both cases infrared (IR) light is highly recommended, because it does not easily interfere with the ambient light and does not require a special environment. In the past, experiments were conducted with ultraviolet (UV) light for reflective targets, but detectors at this wavelength are expensive and apparently have not justified the total cost of a system. Some recently developed systems use visible light in the red wavelength region and passive targets imaged from the cameras through filters passing this particular wavelength. Systems using active

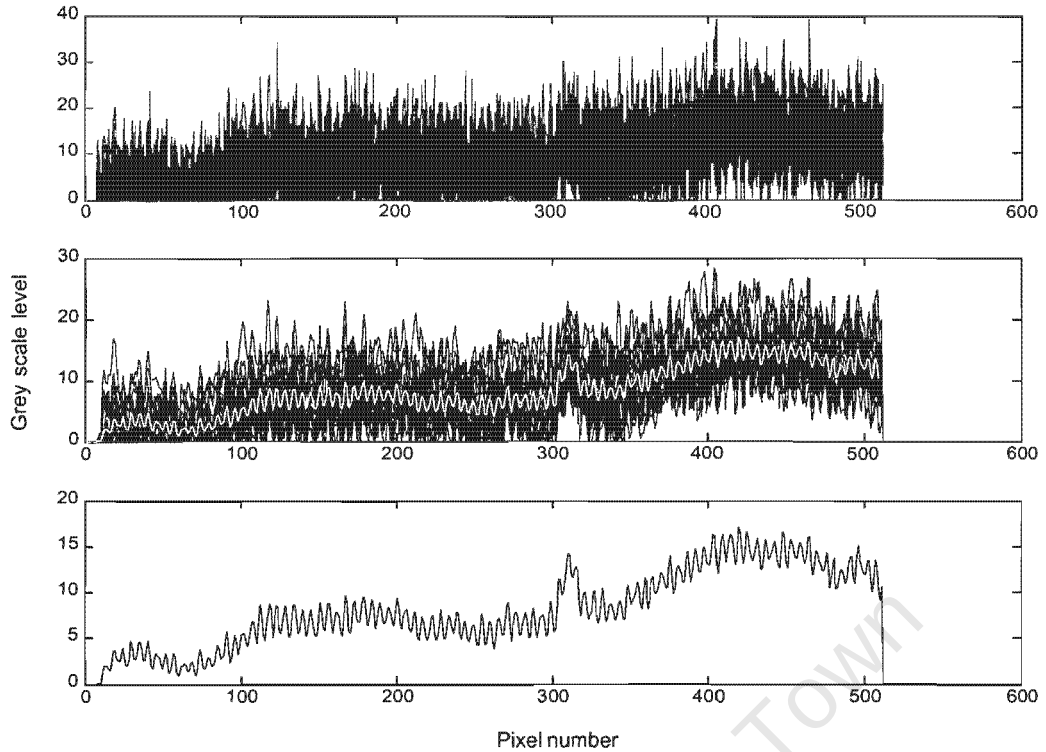
targets have numerous advantages. If they operate within the visual light region, colour identification of the targets is one of the useful solutions to the target identification problem. Another solution to this problem, also applicable to the infrared range, is to activate targets sequentially, one at a time. This technique is not possible with passive targets because they all reflect at the same time.

The last but not least advantage is that active targets, compared to passive ones, can provide a much higher intensity to the camera. This fact should be taken into account when using linear sensors and cylindrical lenses, because of the loss of intensity already discussed in Chapter 3. System applications based on rectangular cameras and spherical lenses can work far more easily with both types of targets. There are some other factors to which either type is sensitive and this could be an obstacle to many applications. LED markers have a limited angle of light radiation, so that large angle emitters lose up to 60% of their output intensity at 120°, and others with a narrower radiation angle can only provide about 20% of the total intensity at 50°. This calls for the use of a more complex detecting system *i.e.* an increased number of cameras, large distances and various angles of positioning, to cover the required volume.

Using passive markers for tracking an object in a specified space, rather than active markers, is more convenient for several reasons. First of all, they can be easily attached at any spot on the object, whether this is an instrument, part of a machine or a human body. Secondly, the wiring of the electrical circuit is avoided, which allows the tracking of free limb motion in human movement analysis. And finally, their number and geometrical positions can be changed for better object-space detection. There are some drawbacks to using targets made of retro-reflective material: the intensity of the reflected signal coming back to the sensor is several times lower than that from an active marker. Another important consideration is that multiple targets cannot be temporally resolved (they all reflect at the same time), unlike active targets, which can be flashed sequentially. When retro-reflective markers are used the direction of the illuminating light has to be as close as possible to the optical axis to produce maximum reflection. The problem in this case may occur from the strong reflection received from side objects that are not targeted for tracking.

4.4 Signal processing to obtain linear coordinates of multiple targets

In this section several experiments were performed both with active and passive targets. First of all, a passive target was imaged using the same rectangular camera and cylindrical lens with a strong reflection from the ambient light. To "uncover" the target from the background a number of lines were filtered and averaged to see their actual shape and position (Figure 4.11a). Afterwards the background had to be removed to make the peak appear above the noise. For this purpose a polynomial approximation was applied to the signal (Figure 4.11b).



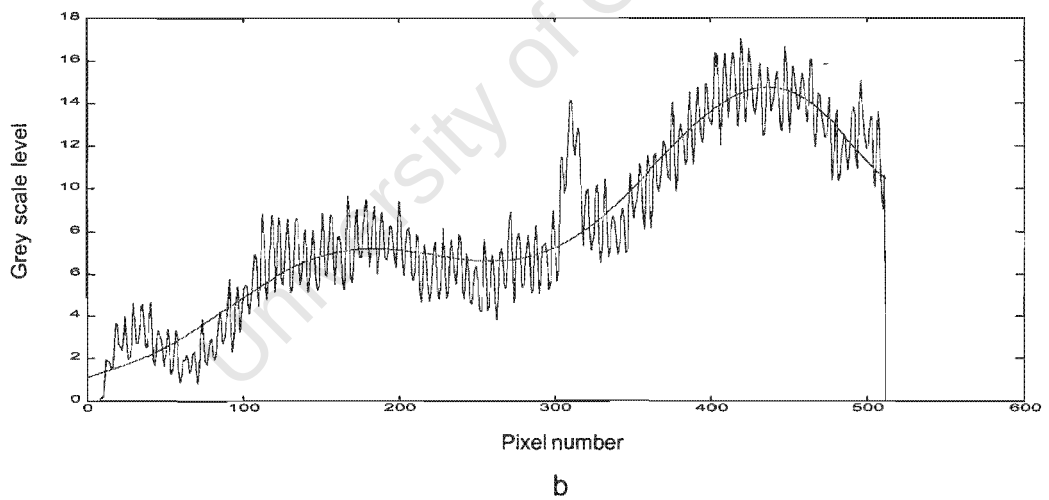


Figure 4.11 Various steps in the process of finding the peak from a passive marker obscured by strong light reflection from the background: (a) taking the average from multiple lines; (b) applying polynomial curve approximation to the signal to remove the background noise.

This interpolation was followed by the subtraction of the approximated noise from the signal and subsequent filtering. The last three steps could be repeated if necessary to obtain the final result such as the one shown in Figure 4.12, from where the exact position of the target could be easily determined.

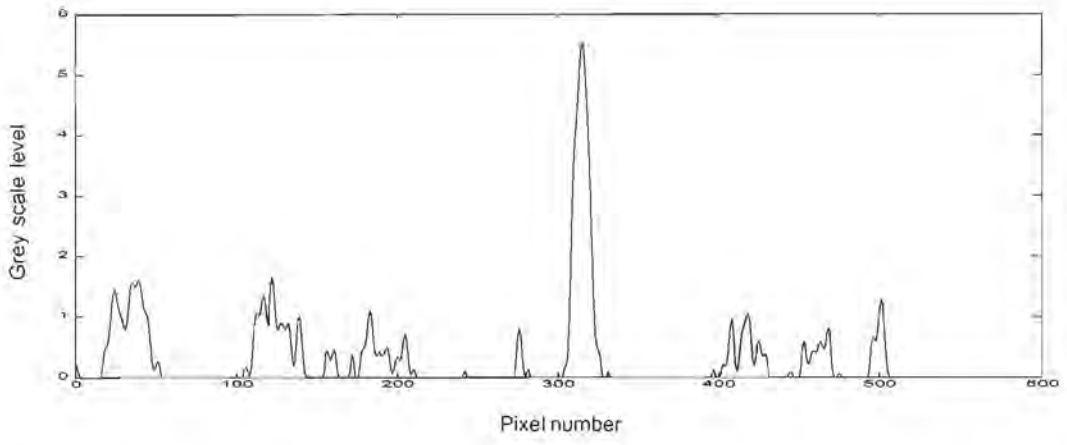


Figure 4.12 Peak from the passive marker on Figure 4.11after two consecutive steps of filtering and background removal using curve approximation.

The second experiment, referred to as Test 4.1 in the following discussion, was done with multiple passive and active targets in which a different approach to the background removal was used. The first objective of the following experiment was to show that multiple passive retro-reflective targets can be detected successfully by means of a rectangular CCD camera equipped with a cylindrical lens. Another objective was to show the difference in the intensity between both types of targets in the detection process. For this purpose a strong IR source was designed consisting of 100 IR LEDs (LD275, GaAs - Siemens), which illuminated a set of 22 flat targets, 2 spherical targets and one active (IR LED of the same type) marker, placed 2 metres from the camera. Images of these point targets were transformed by the cylindrical lens into parallel straight lines.

In order to compare the input signal, when using active targets, to those from passive targets an image of both was captured and analyzed. The above figures illustrate the markers' patterns on the image (Figure 4.13) and the steps taken for extracting the peaks from the signal (Figure 4.14).

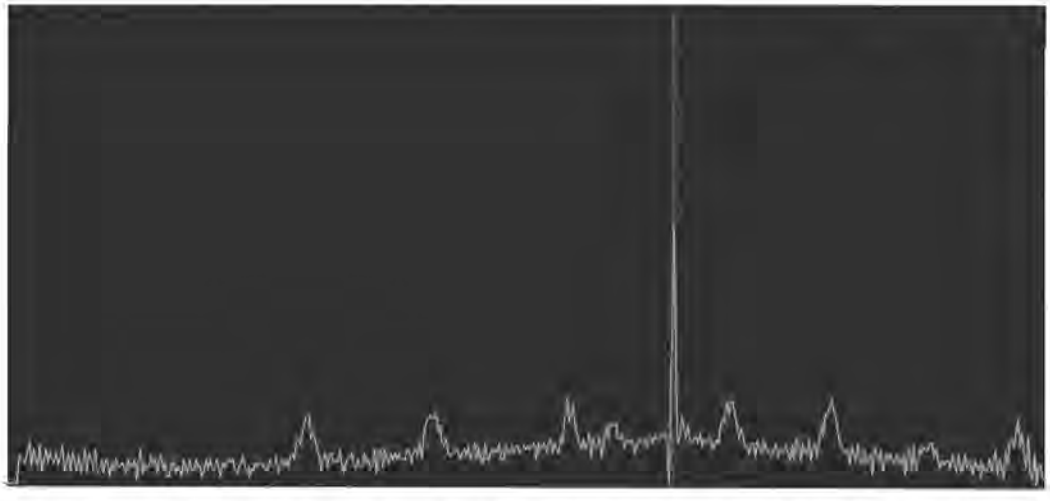


Figure 4.13 Image from one active and eight passive markers and their line profile showing the differences in intensity.

(1) The raw signal (Figure 4.14a, top) was being filtered to reject the noise from the digitization and a threshold was applied (Figure 4.14a, middle, bottom).

(2) The signal obtained must be separated from the background. This was done by isolating the peaks and by linearizing the rest of the signal (Fig.4.14b, top). The profile obtained was superimposed over the initial signal to fill the gaps between peaks and filtered again (Figure 4.14b, second and third). Figure 4.14b, (bottom) shows the signal after the subtraction of the background has been done.

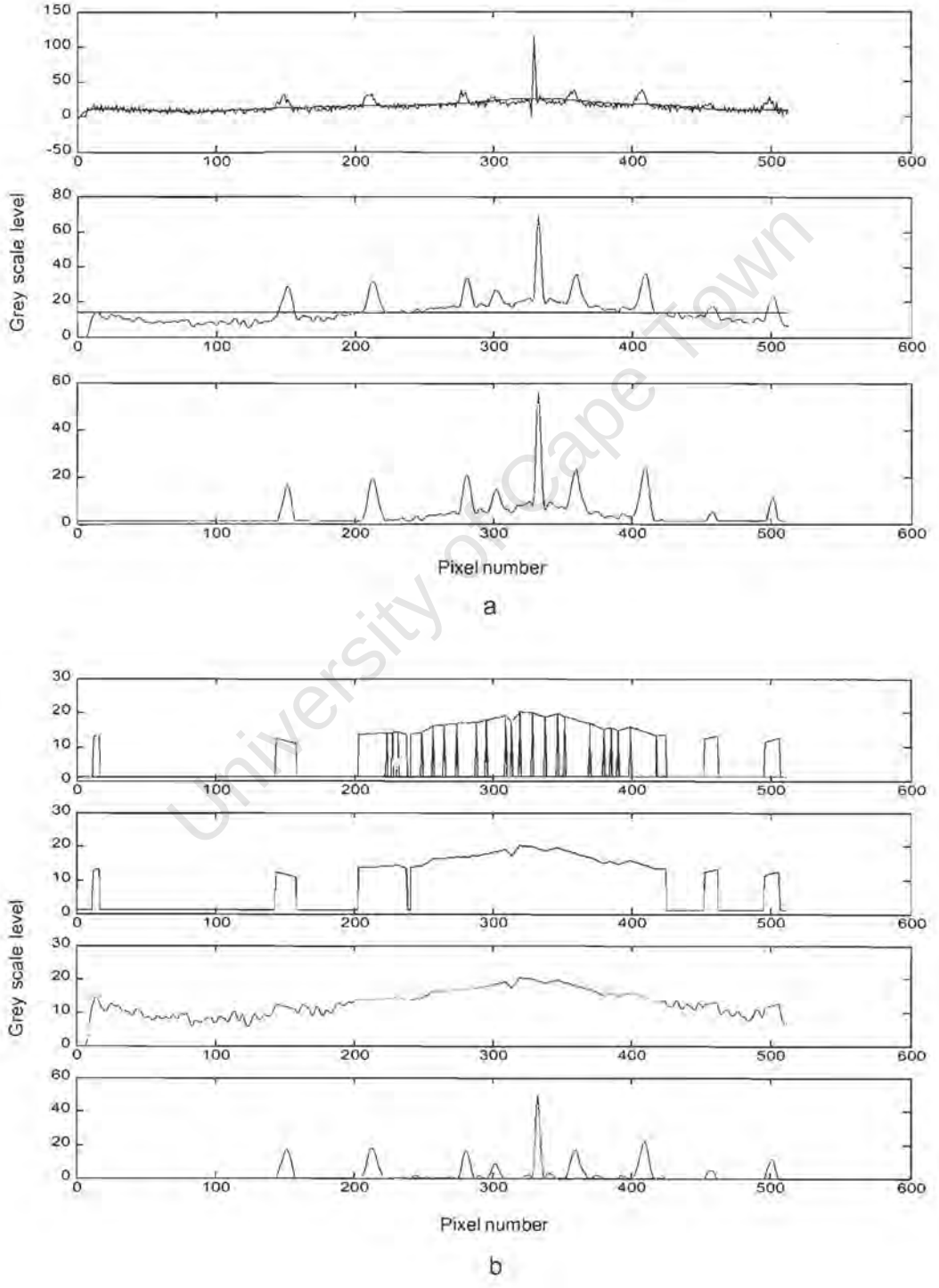


Figure 4.14 Images of multiple passive targets and one active marker, showing: (a) the raw signal, the filtered signal and approximated background removal; (b) noise reduction by reconstructed background removal and peak separation.

(3) This output was amplified and a threshold is again applied to remove the low intensity noise, leaving only the peaks from the reflected targets.
(4) The algorithm calculated the positions of the peaks and their centroid values (Georgiev *et al.*, 1998).

Another experiment, referred to as Test 4.2 in the following discussion, was performed only with the set of 22 flat targets. The contrast of the line patterns representing the markers was satisfactory, although there was a strong reflection from the surrounding environment appearing as a high level background on the picture (Figure 4.15a,b) and on the line profile (Figure 4.16a,b).

a b

Figure 4.15 Images of 22 passive targets taken with a rectangular camera with a spherical (a) and cylindrical lens (b).

In a similar fashion to the previous experiment, the following steps were completed:

(a) Firstly, the raw signal (Figure 4.16a,top) was filtered to smoothen the noise from the digitization and a threshold was applied (Figure 4.16a, middle, bottom). The first graph (Figure 4.16a, top) shows how the interpolation of the background would look like if it was used.

(b) Secondly, the signal obtained which contained a reflection from the ambient light must be separated from the background. This was done in the same way as described in the previous test with the steps of the process shown in Figure 4.16b (first, second and third graph);

(c) The next step was to amplify this output (Figure 4.16b,bottom) by multiplication with an integer value and then a threshold was applied once again to remove the low intensity noise resulting from the background subtraction, leaving only the peaks from the reflected targets.

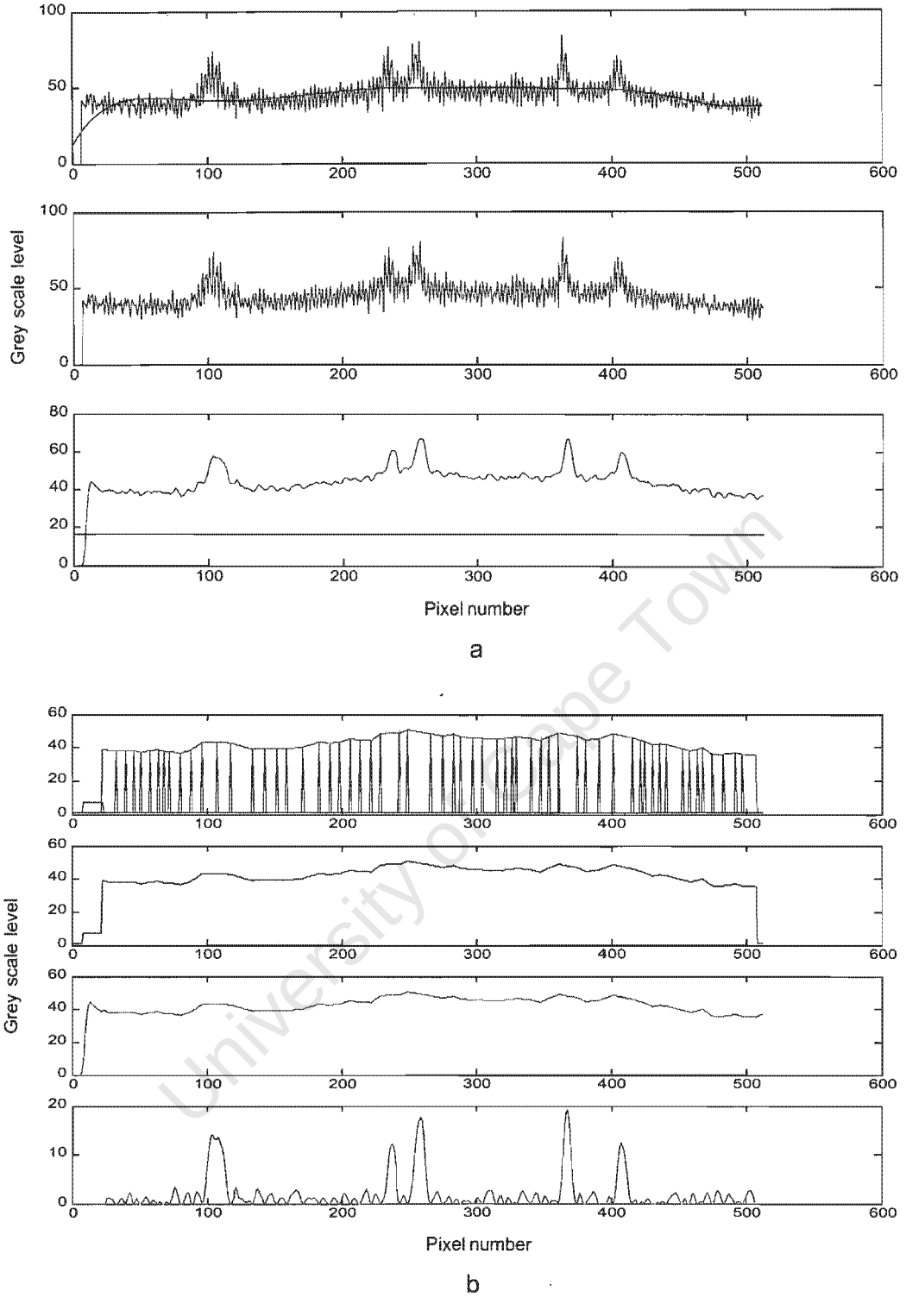


Figure 4.16 Images of multiple passive targets, (a) showing the raw signal and the filtering process and (b) the process of background removal and peak separation.

These experiments with passive and active targets show that both types are detectable from the linear sensor. The graphs of the signal intensities point out clearly the advantages of the active targets when the direction of the incident light coincides with the optical axis. However, active targets could also be affected by reflections occurring between the source and the sensor. In most surgical applications for tracking targets attached to instruments the space

around the patient is equipped with various tools and the lights directed towards the area of operating activity could produce severe parasite reflections that could affect the signal from the targets. In such situation the detection of active targets may become as difficult as that of passive targets. One effective solution for this, besides all signal processing techniques mentioned previously in this section, is to use wider range of infrared wavelengths and to combine the sensor with appropriate filters. Switching the source in a pulse mode with higher electric current supplied to the LEDs would additionally improve the signal-to-noise ratio.

Passive targets in turn are more convenient when imaged at wide angles, but their use necessitates more signal processing and mathematical manipulation. These requirements, together with the target activation control, are among the main requirements in the process of building a system for 3D measurements based on linear sensors. In this regard the specified limitations determine some of the important functional parameters, such as the target detection rate and the maximum number of targets to be used. As long as algorithms, such as filtering and centroid calculation, are applied for both types of targets, specific noise reduction and background removal will only be necessary in the case of passive targets, whereas active targets might only require a simple thresholding as a pre-processing step. These mathematical methods and advanced signal processing techniques could be implemented either by the host computer or special hardware running as a peripheral device. In both cases this will add more complexity to the system and should be considered in terms of cost-efficiency.

4.5 The problem with targets located close to one another

When tracking multiple targets simultaneously a problem may arise with two or more closely located targets, whose peaks are not clearly distinguishable. This can be seen in Figure 4.16b, (bottom) where the first pattern has a shape containing two peak values, that suggests closely located targets (Figure 4.17). An algorithm for recognition and separation of these peaks is proposed to solve this problem using a line profile of a picture where multiple peaks are captured. Initially the algorithm was tested on images from a CCD with a spherical lens, where the targets were presented in 2 dimensions. A combined line profile was constructed by projecting all the peaks onto one line. The true centroid values were calculated beforehand for comparison with those from the superimposed peaks to check the effectiveness of the algorithm. A block diagram of the algorithm is given in Appendix D. The method consists of the following steps, organized as functions:

(I) Set a threshold (zero line) to remove the small peak shapes that could be considered as noise. The threshold value is calculated as 1/5 of the greatest value on the curve.

(II) Find the number and locations of the maxima [$dP(n)/dx_n > 0$, $dP(n+1)/dx_{n+1} < 0$] on the curve.

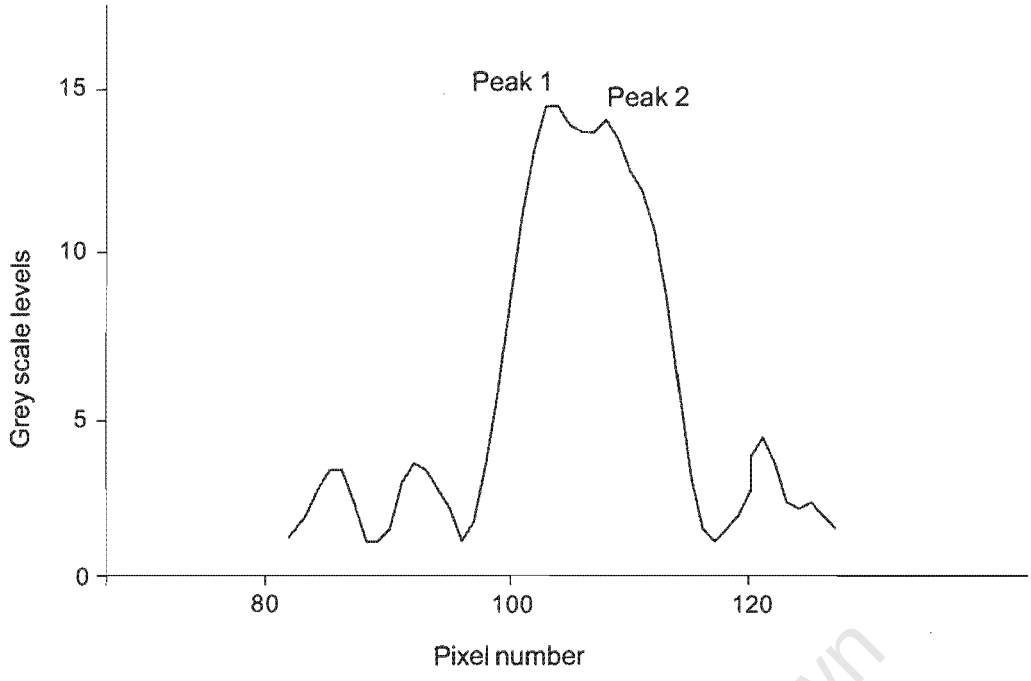


Figure 4.17 An example of two closely located targets on the sensor from Figure 4.16b.

(III) Find the number and locations of the valleys between the maxima which are greater than the zero level [$dP(n)/dx_n < 0$, $dP(n+1)/dx_{n+1} > 0$] and the number and locations of the zero points separating the peaks whenever they appear on the curve.

(IV) Extract the peaks from the profile, which are determined by two neighbouring valleys or zero points.

(V) Reshape the peaks by linear extension of these slopes which do not reach the zero line.

(VI) Place the reshaped peaks at their original locations along the horizontal axis.

The function mentioned in step 5 performs the linear slope extension. There cannot be more than two intersection points for each peak on both its sides. First the peak height is divided into three intervals and the algorithm checks for the point of intersection with the neighbouring peaks (valley point). The intervals are defined as:

(1) $0 < P(n_s) < 1/2\, P_{max}$

(2) $1/2\, P_{max} < P(n_s) < 3/4\, P_{max}$

(3) $3/4\, P_{max} < P(n_s) < P_{max}$

where

P = value of the peak;

n = current pixel number;

$P(n_s)$, $P(n_e)$ = value of the start point (n_s) and end point (n_e);

P_{max} = value of the peak maximum.

Next, depending on which interval the intersection point falls into, the extension segment delta is determined for use in completing the peak shape. According to 1), 2) or 3) delta is:

(a) delta = $|P(n_s+1) - P(n_s)|$

(b) delta = max[D'] where $[D'] = [d_1', d_2' \ldots d_{n-1}']$
$d_1', d_2' \ldots d_{n-1}'$ are the differences between each two points on the same side of the peak, denoted by prime

(c) delta = max[D"] where $[D''] = [d_1'', d_2'' \ldots d_{n-1}'']$
$d_1'', d_2'' \ldots d_{n-1}''$ are the differences between each two points on the opposite side of the peak, denoted by double prime

Subsequently, the extension segment is applied until its length becomes greater than the value of the current point, otherwise the length of the segment is set to 1/2 of the last calculated delta. The same procedure is repeated for the other side of the peak if there is a non-zero valley point. An example of the use of this method on the targets from Figure 4.18 is shown in Figure 4.19 and the centroid values of the peaks in pixels are given in Table 4.4 where:

c	=	true centroids of the markers
c1	=	centroids of each peak calculated after a peak separation by crossing the valley points was applied.
c2	=	centroids of each peak calculated after a peak separation by linear slope extension was applied.

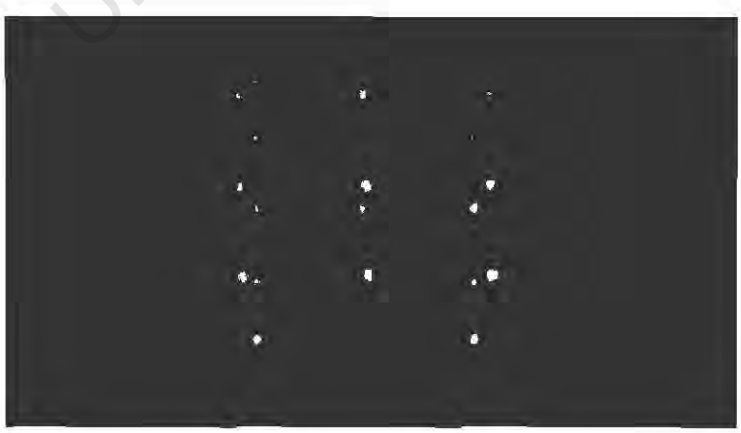


Figure 4.18 The actual positioning of the targets on the sensor shown in Figure 4.19.

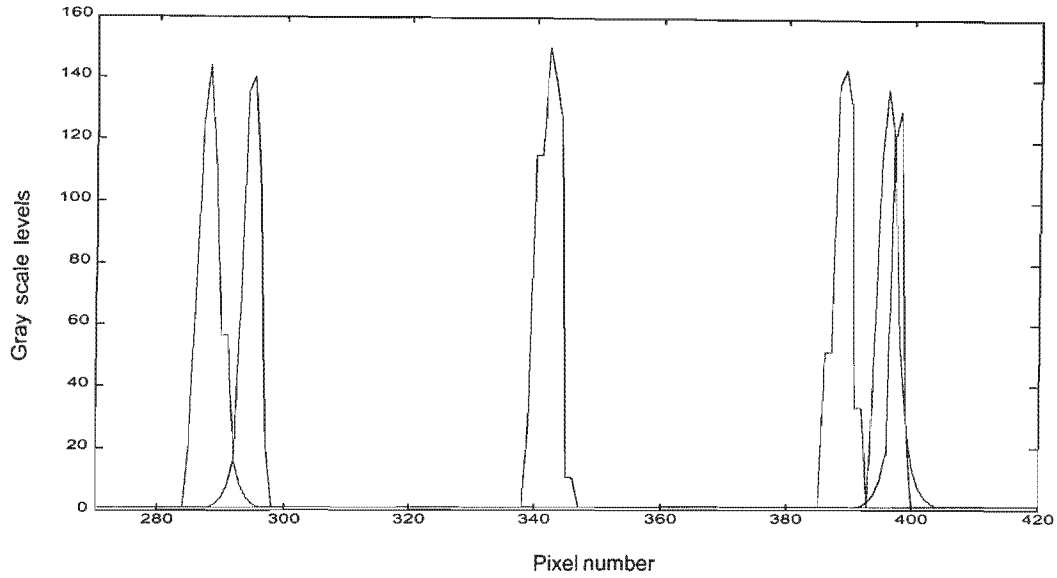


Figure 4.19 The superimposed patterns of the 22 markers from Figure 4.18 on the sensor showing the peak separation after the algorithm for linear slope extension was applied.

Peak position	c	c1	c - c1	c2	c - c2
288	287.84	288.28	-0.44	288.35	-0.50
295	294.88	294.67	0.21	294.39	0.49
342	342.15	342.16	-0.01	342.16	-0.01
389	389.03	388.97	0.06	389.08	-0.05
396	396.05	395.63	0.42	396.05	0.00
398	396.93	397.75	-0.83	397.20	-0.27
Mean			-0.10		-0.06
Stand. dev.			0.46		0.33

Table 4.4 Pixel positions of the targets from Figure 4.18 and 4.19, their true centroid values and the centroid values calculated after two different methods for peak separation were applied to the raw signal.

This algorithm was applied to the peaks from Test 4.1 and Test 4.2. There were no closely positioned targets in Test 4.1. In Test 4.2, however, peaks 1 and 2 in Figure 4.16b (bottom) were closely positioned, covering one another. Using the earlier described algorithm, the two peaks were separated and their centroid values calculated. This process is shown in Figure 4.21. In Table 4.5 are given the values from Test 4.1 and Test 4.2 corresponding to the targets found in Figures 4.20 and 4.21 respectively.

The precision of this separation algorithm depends on the condition whether the two maximum values of the peaks could be registered when separated by at least one pixel. It can be seen from Table 4.4 that the last two peaks (rightmost in Figure 4.19), have a difference of about

1.2 pixels between their centroid values when separated. This compared to their true centroid values, separated by about one pixel, gives the range of the positioning error of the algorithm, which is a fraction of a pixel in this case. This error is important for the 3D reconstruction of the position of the target in a space and which is ultimately involved in the target identification process.

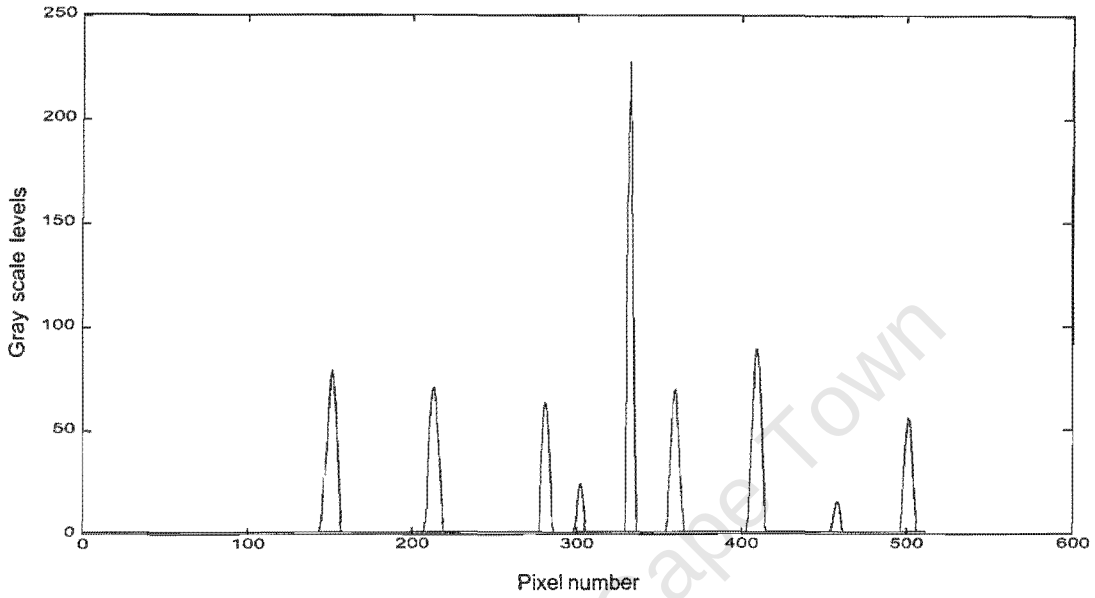


Figure 4.20 Peaks of the targets from Figure 4.14b (bottom) - multiple passive and one active target (Test 4.1) after the algorithm for separation of close targets was applied.

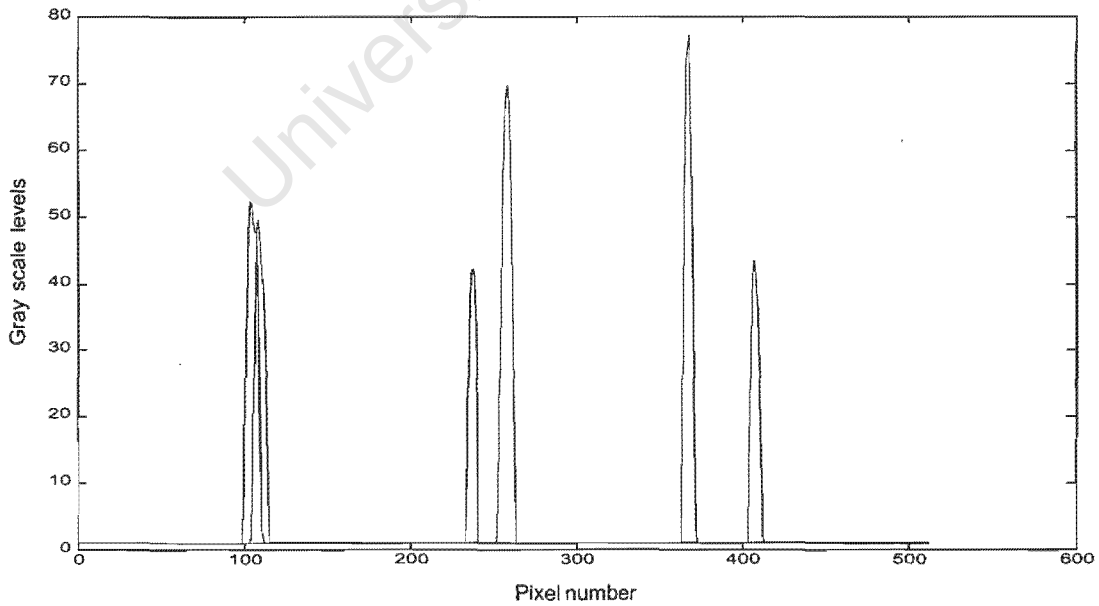


Figure 4.21 Peaks of the targets from Figure 4.16b (bottom) - multiple passive targets (Test 4.2) after the algorithm for separation of close targets was applied.

	Test 4.1		Test 4.2	
	Peak maximum (pixels)	Peak centroid (pixels)	Peak maximum (pixels)	Peak centroid (pixels)
Peak 1	152	151.32	103	104.45
Peak 2	213	212.94	108	109.02
Peak 3	281	281.11	237	236.81
Peak 4	302	301.90	259	258.28
Peak 5	332	332.20	368	367.67
Peak 6	360	359.48	407	407.52
Peak 7	410	409.09		
Peak 8	458	457.31		
Peak 9	501	501.13		

Table 4.5 Peak positions determined by the maximum value and the centroid values of the targets from Test 4.1 (Figure 4.14; Figure 4.20) and Test 4.2 (Figure 4.16; Figure 4.21).

Passive targets are more problematic to distinguish because of a loss of intensity after the light conversion from the cylindrical lens and their peaks will not be as sharp as the ones from active targets. The example in Figure 4.16, showing multiple passive targets, was performed with a rectangular CCD sensor. The intensity of these peaks is relatively high reaching about 60 grey scale levels. However, if a linear sensor is used instead of rectangular sensor, the acquired signal will contain less amount of light, which lowers the intensity, the top area of the peaks will be smoother and they will stretch over a larger area on the sensor. This effect could eventually cause the two adjacent peaks to overlap and to make the separation process impossible.

This problem of close targets will often arise when multiple targets are imaged simultaneously. When the specified targets are in motion they can be registered close to one another on the sensor, especially when their tracks cross or when they lie in a plane orthogonal to the cylindrical lens surface. The separation of their peaks will then be important to identify their positions on the sensor as well as to provide more information for the tracking process.

As it was mentioned in the beginning of section 4.3 active targets have the advantage to be activated sequentially. This helps the process of identification and in this way the problem of close positioned targets is avoided. Each LED target is usually controlled by a circuit receiving signals from an infrared or a radio link or by direct wire connections. The target itself releases pulses of light (visible or infrared), which could also be modulated for easier identification. This process could use Frequency Modulation (FM), Pulse-Width Modulation (PWM) or Pulse-Code Modulation (PCM), where every target operates at a different frequency, pulse width or pulse code sequence. The method using FM signal for identification is usually applied when the output signal from the sensor is further processed by analogue electronics and

consequently converted into a digital signal for computation. The method using PWM or PCM is preferred when the signal after the sensor is converted into a digital form and directed to a digital hardware. In the cases of FM after the low-pass filter that removes the noise, additional band-pass filters should be applied for each frequency to identify the targets. For identification using PWM this is done by means of comparators driving a switching circuit and for PCM the low-pass filter is followed by multiplexing and decoding modules.

The experiments in the second half of this chapter were conducted to evaluate the effect of the cylindrical lens in the process of target detection on the sensor. Following the fact that large variety of applications use rectangular sensors with spherical lenses, the latest was replaced by a cylindrical one for the measurements. The images and the graphs from these measurements were compared to the images and graphs taken with a spherical lens. Further, the effect of the linear sensor was achieved by taking a single line of the image from the rectangular sensor and this approach proved to be efficient in this research. From the experiments in this chapter it could be concluded that the combination of a linear sensor and a cylindrical lens is efficient to detect both passive and active targets. The sensor also provides the necessary linear resolution. For close range applications, such as neurosurgery, the stress would fall as well on how precisely the targets are identified. In these cases close targets separation could be a major contribution at the stage when processing algorithms prepare the data for further computation.

Chapter Five

Reconstruction of Three Dimensional Coordinates from N Linear Sensors

The final step of the data processing in a target tracking system is the algorithm for 3D reconstruction of the space coordinates, which entails the main part of the computation. On this topic two major methods have been explored during the current research: the Direct Linear Transformation (DLT), (Abdel-Aziz and Karara, 1971) and the Bundle Adjustment (BA) method, based on the laws of perspective projection (Gruen, 1996). This chapter will focus on the theory of both methods as well as on a number of different tests based on software simulated data and measured real data.

5.1 The Direct Linear Transformation (DLT) Method

As has been described in Chapter 2, the DLT provides the relationship between XYZ object space coordinates and x,y image coordinates, given by equations [2.9]. One of the main objectives in this research is to test different algorithms for coordinate reconstruction based on data obtained from 1D sensors. For this purpose the equations from [2.9] were transformed by reducing the one equation and rearranging the other in a system of equations for all the sensors. Thus, the image coordinates will be described by just one component p, the target position on the sensor, and registered by its pixel value. The transformation equation that forms the basis for calibrating the one-dimensional camera will then be the following:

$$p_{ik} + \Delta p_{ik} = \frac{L_{1k}X_i + L_{2k}Y_i + L_{3k}Z_i + L_{4k}}{L_{5k}X_i + L_{6k}Y_i + L_{7k}Z_i + 1} \qquad [5.1]$$

where

p	=	image coordinate in one-dimensional camera
i	=	control point, where $i = 1 \ldots N$
k	=	camera number, where $k = 1 \ldots N$ (minimum 3)
Δp	=	error introduced by lens distortion
L_{jk}	=	transformation parameters, where $j = 1 \ldots 7$
X	=	object coordinate in 3D cartesian space
Y	=	object coordinate in 3D cartesian space
Z	=	object coordinate in 3D cartesian space

The term characterizing the distortion, according to the theoretical considerations in Chapter 3, has a non-linear form, that could be approximated by a high order polynomial and one recommended expression to describe it is

$$\Delta p_{ik} = D_k(p_{ik} - p_{ok})^3 \qquad [5.2]$$

where

D_k = distortion parameter for each camera

p_{ok} = principal point for each camera

Similar to the DLT algorithm for two-dimensional cameras (Marzan and Karara, 1976) the principal point can be calculated using the following relationship:

$$p_{ok} = \frac{L_{1k}L_{5k} + L_{2k}L_{6k} + L_{3k}L_{7k}}{L_{5k}^2 + L_{6k}^2 + L_{7k}^2} \qquad [5.3]$$

where the subscript k indicates the camera number used in the calibration and reconstruction process. Here, k = 3 is the minimum according to Macellari (1983). The calibration can be further determined by calculating the transformation parameters L_{jk} and the distortion parameters D_k. This is accomplished using the following assumption:

Firsly, the lens distortion is considered negligible:

$$\Delta p_{ik} = 0 \qquad [5.4]$$

Equation [5.1] then becomes

$$p_{ik} = L_{1k}X_i + L_{2k}Y_i + L_{3k}Z_i + L_{4k} - L_{5k}X_ip_{ik} - L_{6k}Y_ip_{ik} - L_{7k}Z_ip_{ik} \qquad [5.5]$$

or for each camera

$$\mathbf{p}_i = \mathbf{L}_j\mathbf{B}_{ij} \qquad [5.6]$$

where **B**, **L** and **p** are in the form:

$$\mathbf{B} = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & -X_1p_1 & -Y_1p_1 & -Z_1p_1 \\ X_2 & Y_2 & Z_2 & 1 & -X_2p_2 & -Y_2p_2 & -Z_2p_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_i & Y_i & Z_i & 1 & -X_ip_i & -Y_ip_i & -Z_ip_i \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_N & Y_N & Z_N & 1 & -X_Np_N & -Y_Np_N & -Z_Np_N \end{bmatrix} \quad \mathbf{L} = \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \end{bmatrix} \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ p_i \\ \cdot \\ \cdot \\ p_N \end{bmatrix}$$

This matrix equation may now be solved for $\mathbf{L}_j$ (j = 1 . . 7) using at least 7 pixel values p_i to obtain a fully determined system of equations. Therefore, a minimum of 7 control points need to be measured so that each camera can provide enough data to solve the set of calibration parameters. However, if more than 7 control points are used, the system will become over determined and will have multiple solutions. In this event the least squares approach in a matrix form could be used to calculate the solution:

$$\mathbf{L} = (\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{p}) \qquad [5.7]$$

This is the first estimate for the calibration parameters assuming zero distortion in the system. The next step is to calculate the first estimate of the distortion parameters. Using the L_{jk} values obtained from equation [5.7], the principal point p_{ok} may be obtained from equation [5.3], and the substitution in equations [5.1] and [5.2] yields:

$$D_k(p_{ik} - p_{ok})^3 = \frac{L_{1k}X_i + L_{2k}Y_i + L_{3k}Z_i + L_{4k}}{L_{5k}X_i + L_{6k}Y_i + L_{7k}Z_i + 1} - p_{ik} \qquad [5.8]$$

Equation [5.8] may be simplified further for one camera as:

$$C_i = DE_i \qquad [5.9]$$

where C_i is the right side of equation [5.8] and where the index k is dropped and

$$E_i = (p_i - p_o)^3 \qquad [5.10]$$

Then again using the least squares approach

$$D = (\mathbf{E}^T\mathbf{E})^{-1}(\mathbf{E}^T\mathbf{C}) \qquad [5.11]$$

Combining equation [5.2] and the left side of equation [5.1] yields

$$q_{ik} = p_{ik} + D_k(p_{ik} - p_{ok})^3 \qquad [5.12]$$

This iterative process of computing the transformation parameters L_{jk} and the distortion parameters D_k is continued until convergence is achieved for the corrected camera value q_k. Now equation [5.1] is ready to be used to reconstruct the unknown object space coordinates XYZ from the measured image coordinates q_k. Substituting equation [5.12] in equation [5.1] and rearranging gives

$$q_{ik} - L_{4k} = X_i(L_{1k} - L_{5k}q_k) + Y_i(L_{2k} - L_{6k}q_k) + Z_i(L_{3k} - L_{7k}q_k) \quad [5.13]$$

And it can be further set

$$F_k = q_k - L_{4k}$$

$$Q_1 = X$$
$$Q_2 = Y$$
$$Q_3 = Z$$

$$A_{k1} = L_{1k} - L_{5k}q_k$$
$$A_{k2} = L_{2k} - L_{6k}q_k$$
$$A_{k3} = L_{3k} - L_{7k}q_k$$

Then equation [5.13] can be written in matrix form as

$$F_k = Q_m A_{km} \quad \text{where } k = 1 \ldots 3, \text{ and } m = 1 \ldots 3 \quad [5.14]$$

And finally it can be solved for Q_m (*i.e.* X, Y and Z) by

$$\mathbf{Q} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{F}) \quad [5.15]$$

A good check on the accuracy of the calibration would be to compare the known object space coordinates $X_iY_iZ_i$ with the estimated values Q_{mi} from the solution in equation [5.15].

5.2 A test of the DLT Method with simulated data points

To test the algorithm described above a variety of tests were performed with simulated and measured data. A simulation software package called Optical Playground (OPG) was developed, which constructed a linear sensor-cylindrical lens system and a source point emitting light towards the sensor according to the drawing in Figure 3.3. The detected image on the sensor shown in Figure 5.1 was similar to the target profiles presented in the previous chapter and in the same way centroid values were calculated from the raw pixel data. The theory and structure of the OPG simulation softare are given in Appendix E.

Several of these sensor-lens elements could be positioned in 3D space and then they were used to capture images of the target point. The output from each sensor would corresponded to the position of the target. The peak values from the sensors were measured from the beginning of the sensor and given in pixels. The position of the source, on the other hand,was

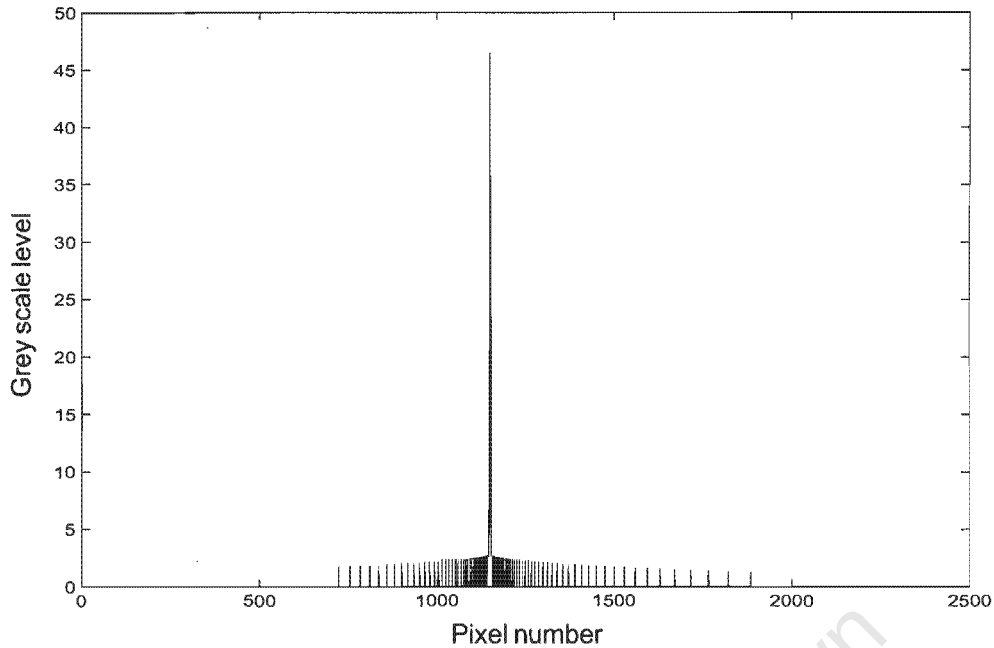


Figure 5.1 The intensity of the pixel values as a line profile along the CCD sensor created with OPG. Peak in the middle corresponds to the target position after the light was converted by the cylindrical lens in front of the sensor.

defined by its XYZ coordinates. Following the condition for a determined system from equation [5.1], multiple sensors were positioned in space to create a detecting system which provided multiple pixel values for each target point. To create a determined system the minimum number of equations is three, the same number applying to the minimum number of sensors.

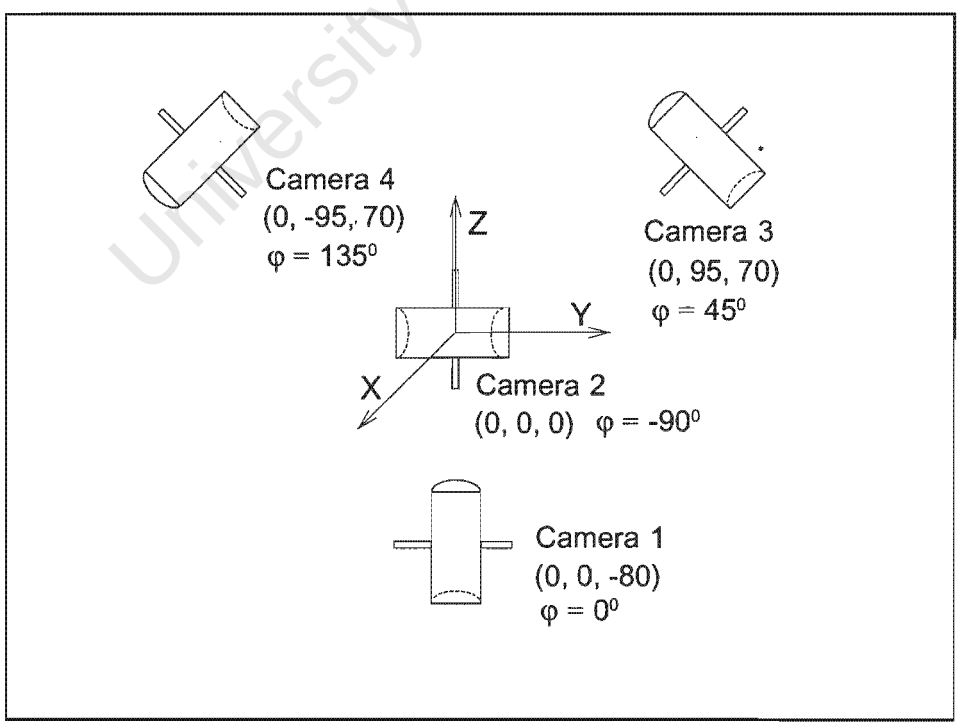


Figure 5.2 Space positions and orientation of four linear CCDs with cylindrical lenses created with OPG for detecting targets in three dimensions. Each CCD sensor has 2048 pixels and a total length of 24.58 mm. The detector is positioned at the origin of the XYZ coordinate system and φ represents the orientation of the sensors.

However, if more sensors are used more pixel data can be obtained, thereby creating redundancy and making the system overdetermined. For the initial experiment, 4 sensors were chosen, the number of which could be extended further to test different spatial configurations.

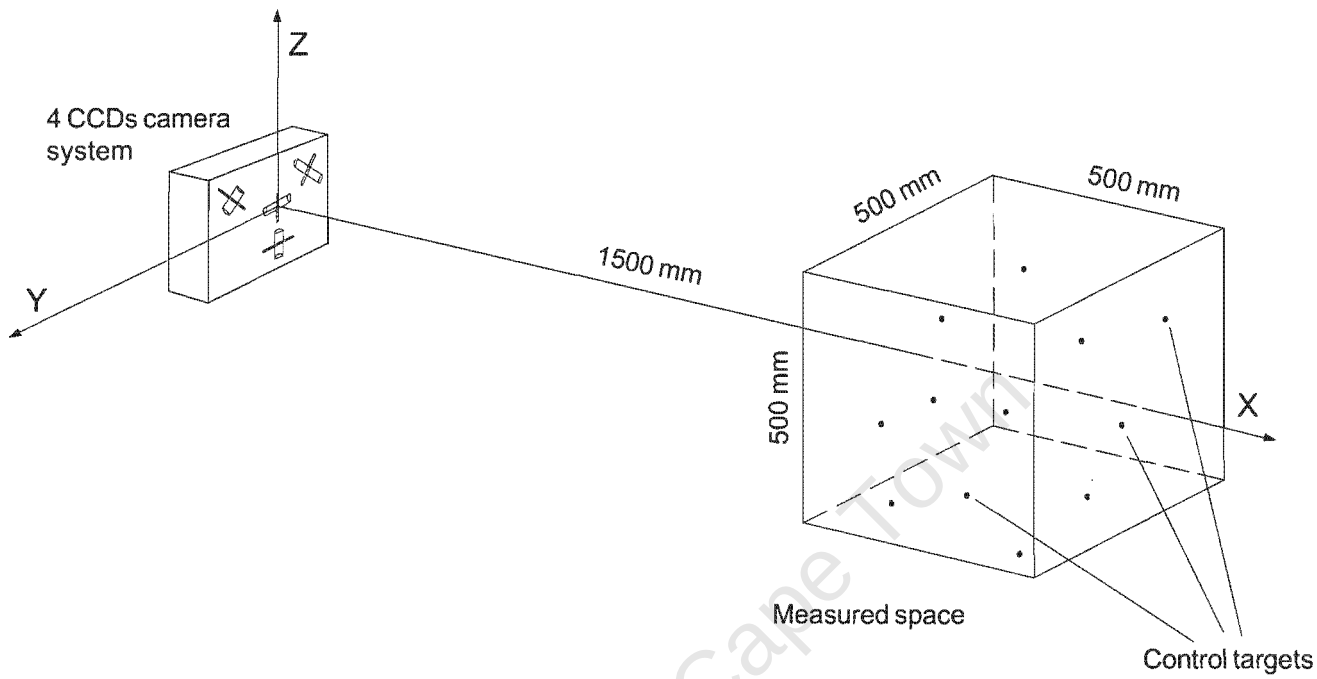


Figure 5.3 Positioning and orientation of a 4-linear CCD cameras detecting system in the origin of an XYZ cordinate system and a volume containing multiple control targets sitting in the corners and inside the cubic space.

The OPG software was used to simulate a detecting system with 4 linear CCD cameras equipped with cylindrical lenses, as is shown in Figure 5.2. Twelve targets simulating active Light Emitting Diodes (LEDs), also created with OPG, were positioned in space according to Figure 5.3 at a distance of about 1500 mm from the system. The pixel values for each CCD were then calculated by the software, taking into account light refraction from the glass material of the lens and integrating the amount of light that falls onto the sensor. The obtained pixel values, together with the XYZ coordinates, were passed to the DLT algorithm, first to calculate the calibration parameters according to equations [5.1] - [5.12]. Subsequently the space coordinates XYZ were reconstructed, using the pixel data and the calibration parameters according to equations [5.13] - [5.15]. In the process of compensating the pixel values for distortion, 30 iterations were completed to update the distortion term from equation [5.12]. This process was repeated until full convergence of the principal point for each linear sensor was achieved, as shown in Figure 5.4.

In Table 5.1 the differences (expressed by their mean, standard deviation and absolute maximum error values) are given between the original and calculated XYZ coordinates of the 12 points, which are randomly positioned in a cubic shaped volume with sides measuring

500 mm. The X direction corresponds to the distance away from the camera, while Y-Z is a plane parallel to the lens surface. The greatest error of 3.86 mm appeared in the X direction which is the depth of the measured space. In both other directions the inaccuracy was less than 0.5 mm. Another simulation was performed with 40 points located along straight lines within a smaller volume sized 10x75x70 mm from the optical axis and separated by small steps, but still located at about 1500 mm in the X direction from the camera. The maximum error in this case dropped to less than 2 mm, and was observed in all directions (Georgiev *et al.*, 1998).

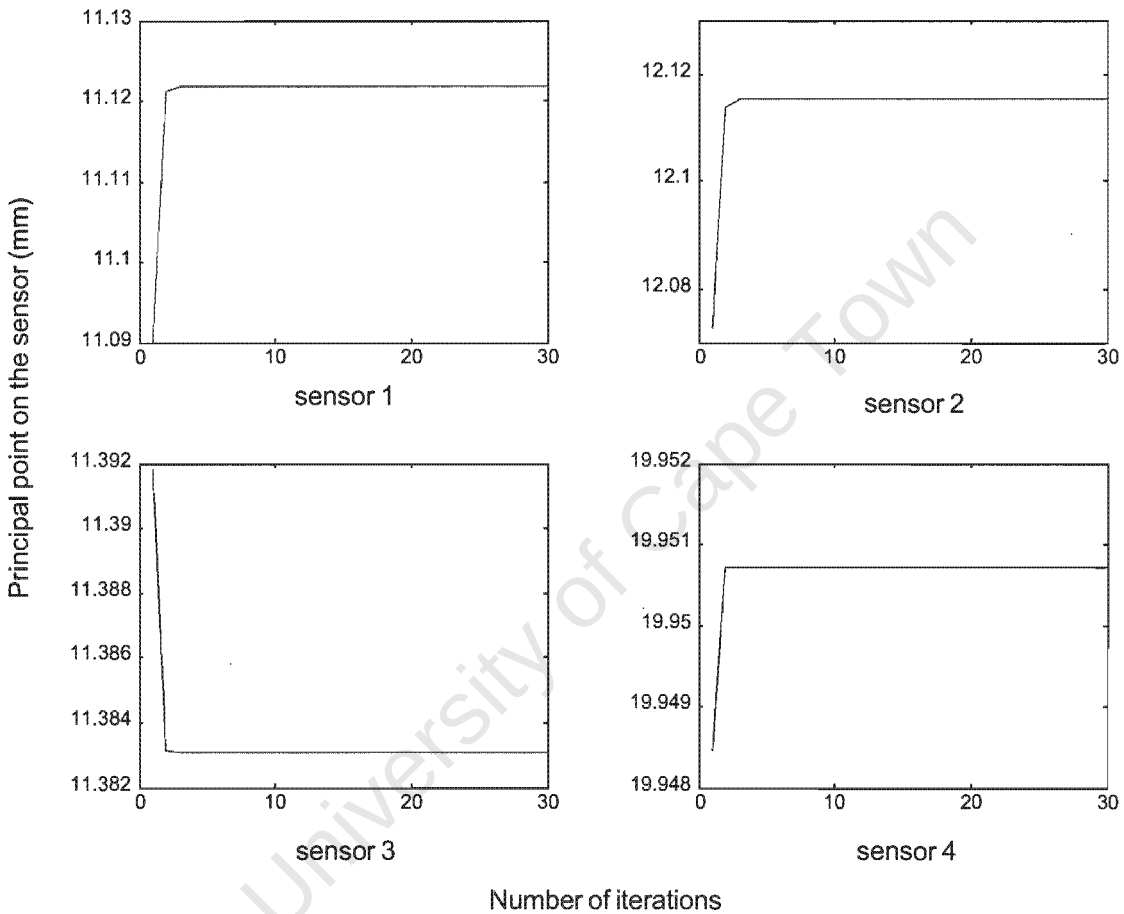


Figure 5.4 Convergence curves of the principal point of 4 linear sensors after a DLT calibration of 30 iterations with principal distances given in mm. Twelve randomly positioned targets in 3D space were used in this process.

	12 point set			40 point set			8 points from the corners		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	0.57	0.08	0.02	0.16	0.07	0.10	-0.10	0.01	-0.01
stand.dev.	1.28	0.15	0.08	0.75	0.37	0.57	1.04	0.19	0.09
abs. mean	0.97	0.13	0.06	0.56	0.28	0.47	0.64	0.12	0.06
abs. stand. dev.	1.01	0.11	0.05	0.52	0.26	0.34	0.83	0.15	0.06
abs. max. error	3.86	0.34	0.19	1.95	1.07	1.40	2.54	0.39	0.19

Table 5.1 Mean, standard deviation and absolute maximum of the difference between the calculated and actual X,Y,Z coordinates of three separate data sets of targets reconstructed with the DLT algorithm.

A reconstruction procedure for the data set of 12 points was performed again, taking into account the distortion produced by the cylindrical lens. The distortion for this type of lens was determined, as explained in Chapter 3, by measuring targets along the axis parallel to the lens curvature. From these curves the differences between the ideal and distorted image were found and correction coefficients were calculated to compensate the measured data. In this test the calibration was done with the raw data where the distortion was compensated using equation [5.2]. Then the pixel values were compensated and used for the reconstruction. The results in Table 5.2 show slight improvement in the X direction, but otherwise no significant changes appear in the other two directions.

	12 point set without distortion compensation			12 point set with distortion compensation		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	0.57	0.08	0.02	-0.47	-0.03	-0.09
stand.dev.	1.28	0.15	0.08	1.25	0.21	0.12
abs. mean	0.97	0.13	0.06	1.09	0.18	0.10
abs. stand.dev	1.01	0.11	0.05	0.78	0.12	0.11
abs. max. error	3.86	0.34	0.19	2.25	0.41	0.37

Table 5.2 Differences in the X,Y and Z directions from the reconstruction of the coordinates of 12 point data set after a correction for lens distortion was applied to the raw pixel data prior to reconstruction with the DLT algorithm.

The values of the L parameters can vary for the different calibration procedures. These values are dimensionless and do not have any physical meaning in the DLT algorithm. The parameter L_4 is a free term in the formula, while the others are coefficients in front of X,Y and Z respectively. The absolute value of the free term could be used as a general estimate of the distance to the principal point from the beginning of the sensor, which can be expressed either in pixels or in millimetres.

	sensor k			
	k=1	k=2	k=3	k=4
L_1	1.78E-01	1.80E-01	2.32E-01	1.06E-01
L_2	4.54E-01	-1.50E-03	4.30E-01	1.57E-01
L_3	3.00E-04	4.64E-01	-4.09E-01	2.68E-01
L_4	1.05E+01	1.12E+01	6.87E+01	-9.65E+00
L_5	1.46E-02	1.47E-02	1.88E-02	8.50E-03
L_6	-5.00E-04	-1.00E-04	5.00E-04	-1.90E-03
L_7	1.00E-04	-1.00E-04	1.30E-03	6.20E-03

Table 5.3 The values of the DLT parameters for a system of 4 linear sensors after the iterative calibration process was completed.

Table 5.3 shows the range of the numbers of the DLT calibration parameters for the test with the 12 simulated points. In this example the calculated principal points from Figure 5.4 are p_{o1} = 11.12 mm, p_{o2} = 12.12 mm, p_{o3} = 11.38 mm and p_{o4} = 19.95 mm. The calculation of the principal point does not, however, directly involve L_4 in formula [5.3]. The other parameters are adjusted in a way to balance the weight of the coefficients, so that the obtained pixel value reflects target position projected onto the direction of the sensor axis. The parameters L_5, L_6 and L_7 are small numbers and since they appear in the terms of the denominator in equation [5.1] and [5.3], influence the accuracy of the calibration and reconstruction results. Theoretically the distance to the principal point for these linear sensors should be expected to be about midway from the beginning or 12.3 mm. From the numbers above it is seen that only sensor 2 is close to this value. Sensors 1 and 3 have produced small deviation from the estimated point, which could be explained with instability of the calibration parameters. The value of the principal point for sensor 4, however, has significant shift towards the end point. One possible reason for this could be the insufficient number of points (only 12) used in the calibration and their distribution in the Y-Z plane parallel to the plane of the sensors. For this particular data set it could be seen that for sensors 1 and 2 in the middle of the system most of the points are in the observed area. For sensors 3 and 4 positioned at 95 mm away from the centre the points are distributed mainly in the right half of the field providing better calibration data for sensor 3. Sensor 4 in this case has fewer points in the central area and most of the points fall in the left half of its field of view, moving the position of the principal point away from the midpoint. A complete list of results from this example using simulated data is given in Appendix F (Test 5.1).

Better calibration could be achieved if the points are regularly distributed in each direction. The second example from Table 5.1 used 200 points for calibration positioned along each diagonal of the Y-Z coordinate axes and separated by small steps. These points also followed the diagonals of the X-Y and X-Z coordinate axes. The differences from the data set of 40 points reconstructed after this calibration showed that it was better than the previos one that used 12 random points, however, the principal points were also shifted as a result of the arrangement of the calibration data. Results with similar accuracy were achieved when the targets used for calibration and reconstruction were positioned in the corners of the measured volume in Figure 5.3. This example confirms the hypothesis that symetrical distribution of the points used for calibration would improve the values of the calibration parameters.

It was recommended by Vaughan *et al.* (1998) that significant improvement of the accuracy could be achieved if more than 4 linear sensors were used. According to this concept such a detecting system, which consists of several detectors equipped with 4 linear sensors each and looking at different angles at the object space, can decrease the error to less than 1 mm. To test this concept, two detectors were arranged to measure the initial set of 12 points. Additional pixel data were generated for the second detector, which was positioned at an

angle of 60° related to the first detector. The data from both detectors were used for calibration and reconstruction of the space coordinates and the results are summarized in Table 5.4.

	12 points from two detectors consisting of 4 linear CCDs		
	X (mm)	Y (mm)	Z (mm)
mean	0.07	0.03	0.00
stand.dev.	0.65	0.18	0.12
abs. mean	0.45	0.15	0.09
abs. stand.dev.	0.47	0.10	0.08
abs. max. error	1.65	0.36	0.30

Table 5.4 Differences in the X,Y and Z direction from the reconstruction of the coordinates of 12 point data set, where two detectors with 4 linear sensors each were used to capture the raw pixel data for the reconstruction with the DLT algorithm.

The general observation, which ensues from applying the DLT Method for calibration and reconstruction of 3D coordinates, is that the algorithm in this form is not stable enough to produce consistent accuracy. This conclusion followed the fact that for some measurements the error after the reconstruction was minimal, while for other measurements, such as the example with real data shown later in this chapter, the differences were considerable. Although the calibration process achieved convergence of the principal point for all the sensors, the reconstruction of the coordinates could still bring about large errors at certain space positions. Lens distortion is an important factor in both the calibration and the reconstruction stage. The expression in equation [5.2] is a good estimate for the pixel displacement obtained due to this optical aberration and utilizes the algorithm to achieve convergence at a certain value on the sensor. This convergence point, however, does not always correspond to the middle of the sensor as expected and in some instances falls outside the central area.

The reconstruction part of the DLT algorithm is based on a Least Squares approach for an overdetermined system of equations. These equations are independent because none of the sensors had the same spatial orientation for the described experiment, unlike the DLT method for rectangular sensors where x and y pixel values are linked by an orthogonal relationship. In this regard, to test the performance of the DLT algorithm further, additional constrains could be incorporated into the system of equations for 1D sensors. A similar approach was suggested in the formulation of DLT using 2D data (Bopp and Krauss, 1978). In this example the convergence point becomes more stable after an iterative approach was used to solve the system (Karara, 1980). As a result the error in the final calculations of the reconstructed points was also minimized.

5.3 Bundle Adjustment (BA) Method

This method was described by Gruen (1996), for coordinate reconstruction of targets, imaged from multiple two-dimensional cameras. The idea to apply the same technique for multiple one-dimensional sensors was presented by Vaughan *et al.* (1998). It uses the concept of the perspective projection of a point lying in the object space, connected by a straight line through the perspective centre of the camera lens and projected onto the sensor. This straight line condition is also known as the "collinearity condition" and is expressed by the following relationship:

$$\begin{bmatrix} x - x_p \\ y - y_p \\ -c \end{bmatrix} = \lambda \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X - X_o \\ Y - Y_o \\ Z - Z_o \end{bmatrix} \qquad [5.16]$$

where

(X,Y,Z)	=	object space coordinates of object point P
(X_o,Y_o,Z_o)	=	object space coordinates of perspective centre 0
$(x,y,0)$	=	measured image space coordinates of point P'
$(x_o,y_o,0)$	=	image space coordinates of principal point H'
c	=	distance from perspective centre to image plane
λ	=	scaling factor
$\mathbf{R}$	=	rotation matrix between image and object space coordinate systems shown in Figure 5.5.

The interior orientation of the sensor is defined by the coordinates of the principal point x_o,y_o and the distance c, which in this case is equivalent to the focal distance of the camera. X_o,Y_o,Z_o are the exterior coordinates defining the sensor position and $\mathbf{R}$ is the rotation matrix defined by the three angles to determine the rotation between the interior and exterior coordinate system. As seen in equation [5.16], dividing the first and second equations by the third and cancelling the scaling factor λ yields

$$x = x_o - c\left[\frac{r_{11}(X - X_o) + r_{21}(Y - Y_o) + r_{31}(Z - Z_o)}{r_{13}(X - X_o) + r_{23}(Y - Y_o) + r_{33}(Z - Z_o)}\right]$$

$$y = y_o - c\left[\frac{r_{12}(X - X_o) + r_{22}(Y - Y_o) + r_{32}(Z - Z_o)}{r_{13}(X - X_o) + r_{23}(Y - Y_o) + r_{33}(Z - Z_o)}\right] \qquad [5.17]$$

Here,

r_{11} = cosine of the angle between X and x axes

r_{21} = cosine of the angle between Y and x axes

r_{31} = cosine of the angle between Z and x axes

r_{12} = cosine of the angle between X and y axes

r_{22} = cosine of the angle between Y and y axes

r_{32} = cosine of the angle between Z and y axes

r_{13} = cosine of the angle between X and z axes

r_{23} = cosine of the angle between Y and z axes

r_{33} = cosine of the angle between Z and z axes

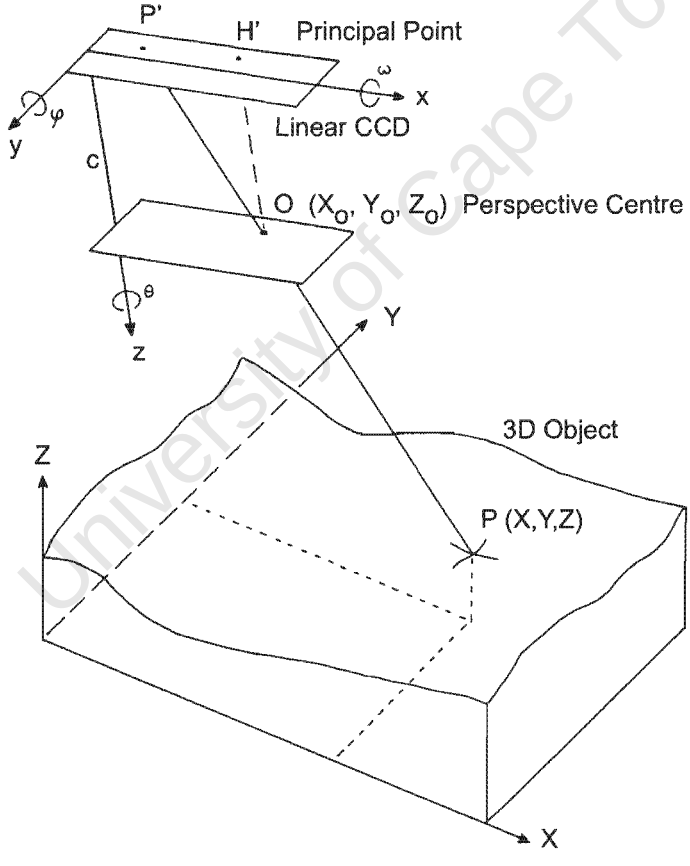


Figure 5.5 A point P in 3D space can be imaged with a linear CCD and its coordinates reconstructed using the laws of perspective projection.

If a linear sensor is used instead of the rectangular, the image coordinate y will drop off and the sensor will be measured only along x. Then

$$p = x$$

$$p_o = x_o$$
$$L_1 = r_{11} \qquad L_2 = r_{21} \qquad L_3 = r_{31} \qquad L_4 = r_{13} \qquad L_5 = r_{23} \qquad L_6 = r_{33}$$

and

$$q = - \frac{(p - p_o)}{c} \qquad [5.18]$$

Substituting the above parameters into equation [5.17] yields

$$q = \frac{L_1(X - X_o) + L_2(Y - Y_o) + L_3(Z - Z_o)}{L_4(X - X_o) + L_5(Y - Y_o) + L_6(Z - Z_o)} \qquad [5.19]$$

which could be rearranged to

$$(L_1 - qL_4)X + (L_2 - qL_5)Y + (L_3 - qL_6)Z = (L_1 - qL_4)X_o + (L_2 - qL_5)Y_o + (L_3 - qL_6)Z_o \qquad [5.20]$$

Equations [5.18] and [5.20] determine the relationship between the measured image coordinate p and the object space coordinates X,Y,Z of the point P. There are 11 coefficients specific to the linear sensor: c, p_o, X_o, Y_o, Z_o and L_1 through L_6. To calculate these calibration parameters, 12 or more data points p_i are needed to capture with known X_i, Y_i, Z_i space coordinates. The problem could then be stated as a non-linear optimization problem, where the object function to be minimized is

$$J = \sum_{i=1}^{N} [(L_1 - q_i L_4)(X_i - X_o) + (L_2 - q_i L_5)(Y_i - Y_o) + (L_3 - q_i L_6)(Z_i - Z_o)]^2 \qquad [5.21]$$

and the unknown variables are the 11 calibration coefficients.

Once the calibration coefficients are determined for each sensor, then the reconstruction of unknown X, Y, Z positions should be possible by the measured p values from 3 or more linear sensors. From equation [5.20] it can be written

$$A_{j1} = L_{1j} - q_j L_{4j}$$
$$A_{j2} = L_{2j} - q_j L_{5j}$$
$$A_{j3} = L_{3j} - q_j L_{6j}$$

Assuming that index j represents the number of the sensors and

$$F_j = A_{j1}X_{oj} + A_{j2}Y_{oj} + A_{j3}Z_{oj}$$

so

$$[A_{j1}\ A_{j2}\ A_{j3}] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [F_j] \qquad j=1..n\_sensors \qquad [5.22]$$

and representing the matrix [A] as

$$\mathbf{A} = [A_{j1}\ A_{j2}\ A_{j3}]$$

and the vectors [Q] and [F] as

$$\mathbf{Q} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\mathbf{F} = [F_j]$$

Equation [5.8] may be written as

$$F_j = Q_m A_{jm} \quad j = 1..n\_sensors,\ m = 1..3 \qquad [5.23]$$

which may be solved using the standard form:

$$\mathbf{Q} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{F}) \qquad [5.24]$$

providing n _sensors ≥ 3.

5.4 A test of the BA Method using multiple sensors and simulated points

This BA Method was tested with 16 simulated points (Vaughan *et al.*, 1998), whereby applying the inverse problem to equation [5.20], the sensor values of the points were generated from known (X,Y,Z) coordinates and a known set of 11 calibration coefficients which were determined by a certain geometry (position and orientation) of 4 linear sensors arranged in one detector with size 200 x 100 mm. These sensor values were considered to be ideal data, obtained using the perspective projection theory, which is the basis for the BA Method. Random noise was then added to all pixel values and calibration and reconstruction was performed to calculate X,Y,Z coordinates using the initially defined calibration parameters. In this test multiple positions for the detector were chosen to measure a volume of 1000 mm^3 at a distance of 2000 mm from the detector (Figure 5.6). The examples using the BA Method with one and two detectors, similar to those with the data generated by a simulated camera-lens system,

were repeated with a high level of success and confirmed that the largest error appeared in the direction corresponding to the depth of the measured space Table 5.5.

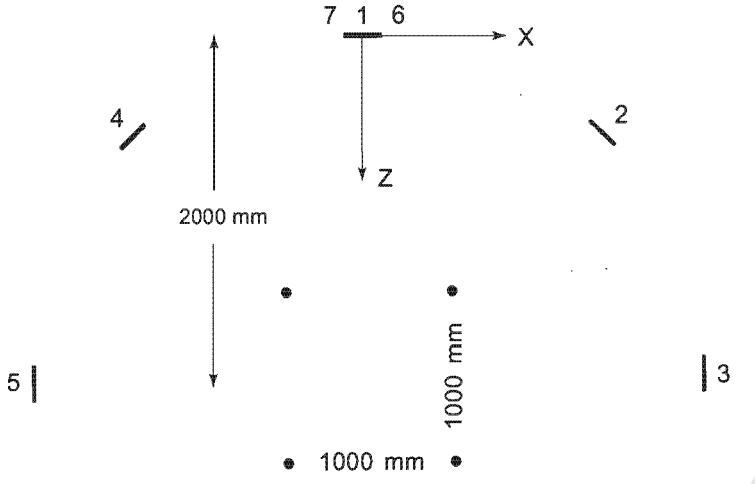


Figure 5.6 A view of the X-Z plane of the test volume, illustrating the location of the data points (solid dots) and the location of the seven detector positions. For clarity, detectors 6 and 7 have not been drawn.

# Cameras	1st	2nd	3rd	Noise (mm)	X (mm)	Y (mm)	Z (mm)
1	1			0.005	0.68	0.71	3.54
1	2			0.005	2.25	0.66	2.28
1	3			0.005	3.42	0.69	0.83
1	4			0.005	2.33	0.66	2.36
1	5			0.005	3.51	0.72	0.85
1	6			0.005	0.67	0.69	3.45
1	7			0.005	0.68	0.68	3.44
2	1	2		0.005	0.16	0.14	0.28
2	1	3		0.005	0.17	0.12	0.17
2	2	4		0.005	0.14	0.12	0.17
2	3	5		0.005	0.35	0.12	0.13
2	6	7		0.005	0.35	0.35	1.76
2	3	4		0.005	0.28	0.13	0.18
3	1	2	3	0.005	0.14	0.10	0.15
3	1	2	4	0.005	0.10	0.10	0.17
3	1	3	5	0.005	0.16	0.10	0.12
3	1	6	7	0.005	0.31	0.32	1.58
7				0.005	0.07	0.06	0.09

Table 5.5 Number of detectors used in the reconstruction process and the differences in the X,Y and Z directions after coordinate reconstruction with BA Method of 16 points within a volume of 1000 mm^3 from their pixel values generated using the Perspective Projection Theory.

Further, the number of the detectors used in the reconstruction process was also increased to as many as 7 (*i.e.* 28 sensors) for targeting high spatial resolution improvement.

This Bundle Adjustment Method was also applied to the simulated 12 points from Test 5.1 (Appendix F), solving the inverse problem to obtain the sensor values from the X,Y, Z coordinates with a set of calibration parameters corresponding to the geometry of the detector consisted of four cylindrical lens-linear sensor elements, shown in Figure 5.2. These so-called perfect camera values were further used by the DLT algorithm and the BA Method for reconstruction of the initial X,Y,Z coordinates without adding a noise. In Tables 5.6 and 5.7 the following are given: the mean; standard deviation; absolute mean; absolute standard deviation and the absolute maximum error of the reconstructed data. The full list of the results can be found in Appendix F (Test 5.2).

	12 points reconstructed with DLT from their projected coordinates, generated using the Perspective Projection Theory		
	X (mm)	Y (mm)	Z (mm)
mean	-0.01	-4.00E-04	4.00E-04
stand.dev.	0.22	0.02	0.02
abs. mean	0.15	0.01	0.01
abs. stand.dev	0.15	0.01	0.02
abs. max. error	0.48	0.05	0.06

Table 5.6 Differences in the X,Y and Z directions after coordinate reconstruction with DLT algorithm of 12 points, where their pixel values were generated using the Perspective Projection Theory.

	12 points reconstructed with Bundle Adjustment Method from their projected coordinates generated using the Perspective Projection Theory		
	X (mm)	Y (mm)	Z (mm)
mean	-2.27E-11	9.00E-14	-7.20E-13
stand.dev.	3.39E-11	4.21E-12	3.40E-12
abs. mean	2.92E-11	2.62E-12	1.89E-12
abs. stand.dev	2.79E-11	3.21E-12	2.87E-12
abs. max. error	9.46E-11	1.09E-11	1.05E-11

Table 5.7 Differences in the X,Y and Z directions after coordinate reconstruction with Bundle Adjustment Method of 12 points, where their pixel values were generated using the Perspective Projection Theory.

The results from the BA Method show significantly better accuracy compared to those from the DLT algorithm and in both cases the error in the X direction (the depth of the object space) is an order of magnitude greater than the error in the other directions. Figure 5.7

shows the process of convergence of the principal point for all four sensors after completion of 30 iteration steps.

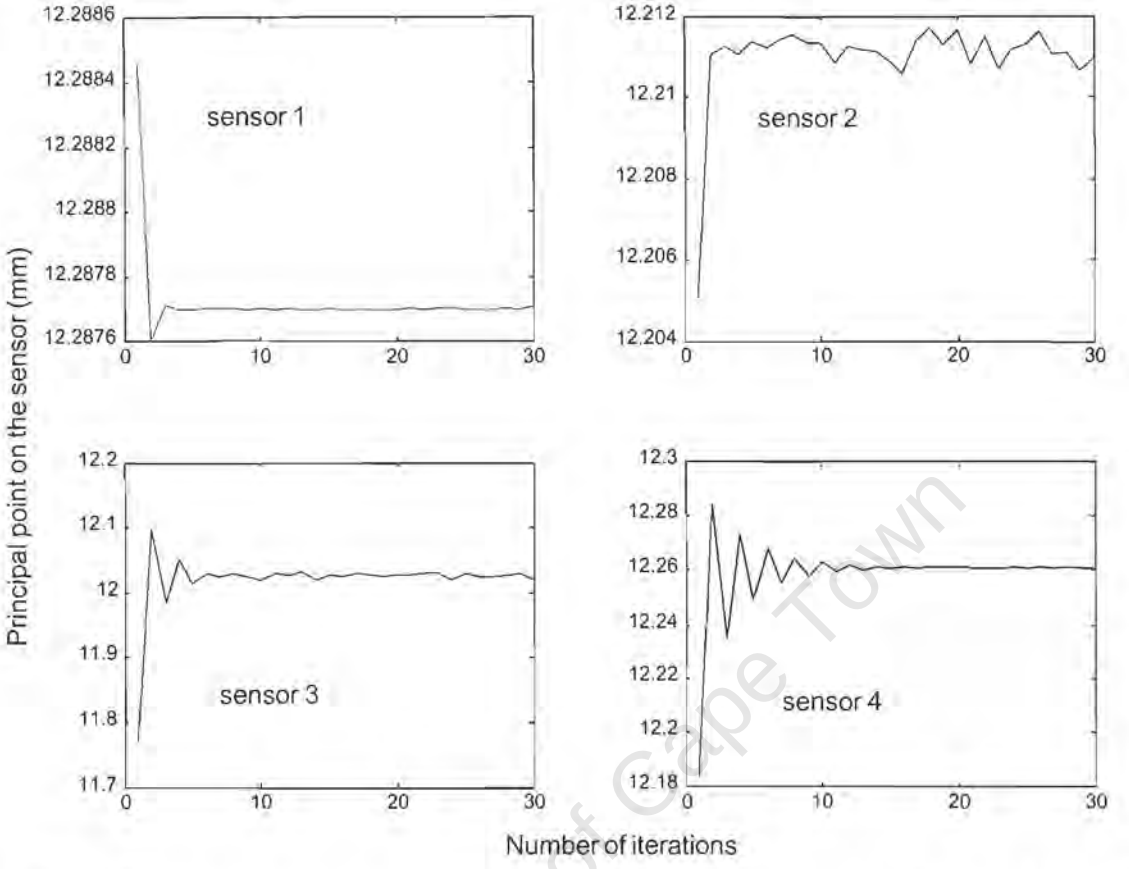


Figure 5.7 The convergence process of the principal point for sensors 1 through 4 using DLT algorithm for calibration and reconstruction of the X,Y,Z coordinates for the 12 points from their pixel values generated according to the Perspective Projection Theory (Bundle Adjustment Method).

5.5 A test with measured data points using the BA and DLT Methods

In addition to Tests 5.1 and 5.2, a separate experiment was performed with a real calibration frame constructed for this purpose, which was made of aluminium bars arranged in a pyramidal shape, consisting of 19 active (IR LED) targets as shown in Figure 5.8.

Figure 5.8 Calibration frame consisting of 19 active targets with 3 of them switched on (left) and their image on the camera where the light is converted by the cylindrical lens (right).

The calibration was done with two rectangular cameras Sony DFW- V500 with a CCD crystal, measuring 640x480 pixels with a diagonal of 0.3 inches, equipped with cylindrical lenses, so that only one dimension along the camera was used, similar to the output of a linear sensor. Infrared filters were used to improve the signal-to-noise ratio during the data acquisition stage. The cameras were positioned at two levels in such a way as to provide four different camera positions and orientations as is shown in Figure 5.9.

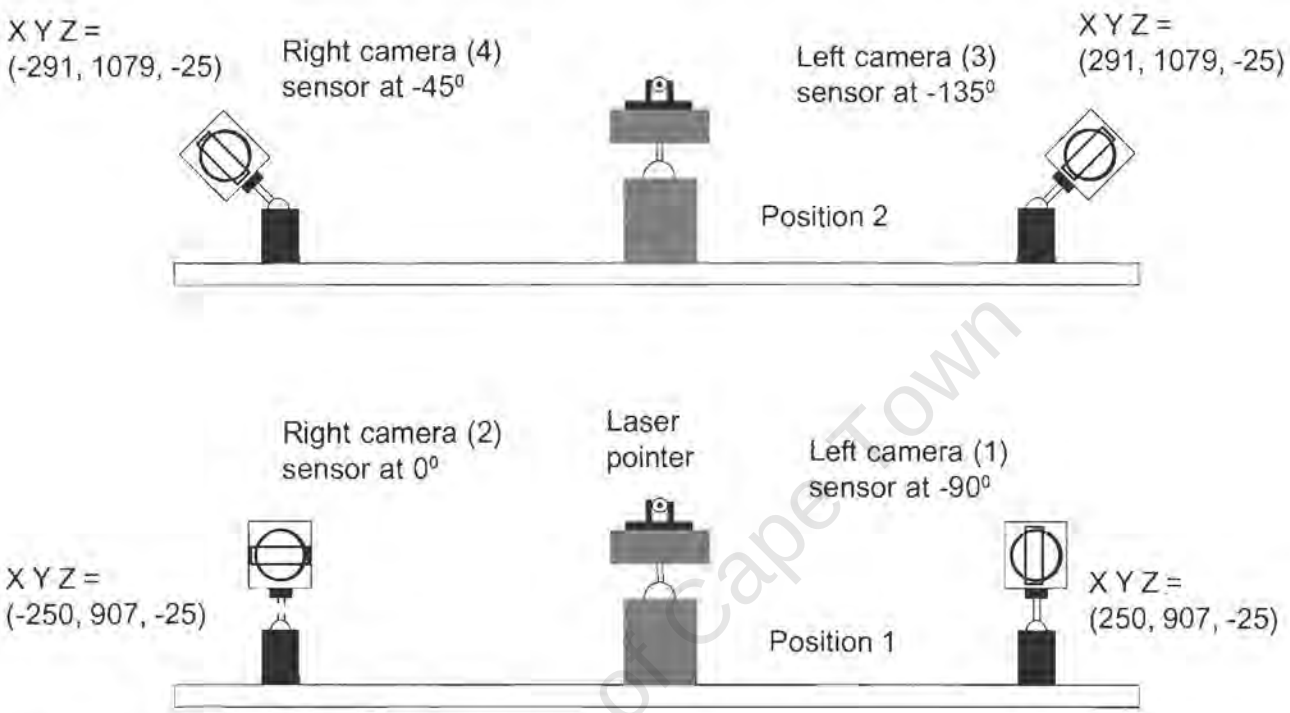


Figure 5.9 Camera arrangement used to image the calibration frame from Figure 5.5. Laser in the middle is to ensure that the Z direction is orthogonal to the X-Y (horizontal-vertical) plane and to fix X=0.

The frame was positioned at two distances: at 3 m from the cameras for the first test, and at 1 m for the second test. The cameras were aligned using a laser pointer directed to the measured space. The optical axes of the lenses and the cameras were then set parallel to the laser beam. The camera angles around the vertical axes were corrected by taking images of the spot formed from the laser and next the distances to the spot from the sides of each image were equalised. After the images of the targets were taken, a calibration, followed by reconstruction of their 3D coordinates was done, using DLT algorithm in the same way as in the previous tests. Results from these two tests, presented in Appendix G (Test 5.3 and Test 5.4), were not satisfactory as large errors appeared in all directions proving the instability of this algorithm.

Similar to the procedure done with the data from Test 5.1 which generated the data in Test 5.2 by solving the inverse problem from the BA Method, the same problem was solved for Tests 5.3 and 5.4. The initial data was acquired with real cameras from the targets on the calibration frame in Figure 5.8. The generated data was plotted against the measured data

to discover their relationship. In Figure 5.10 the points from Test 5.3 are plotted, where the distance to the targets was 3 m. The data acquired from Test 5.4 is shown in Figure 5.11,

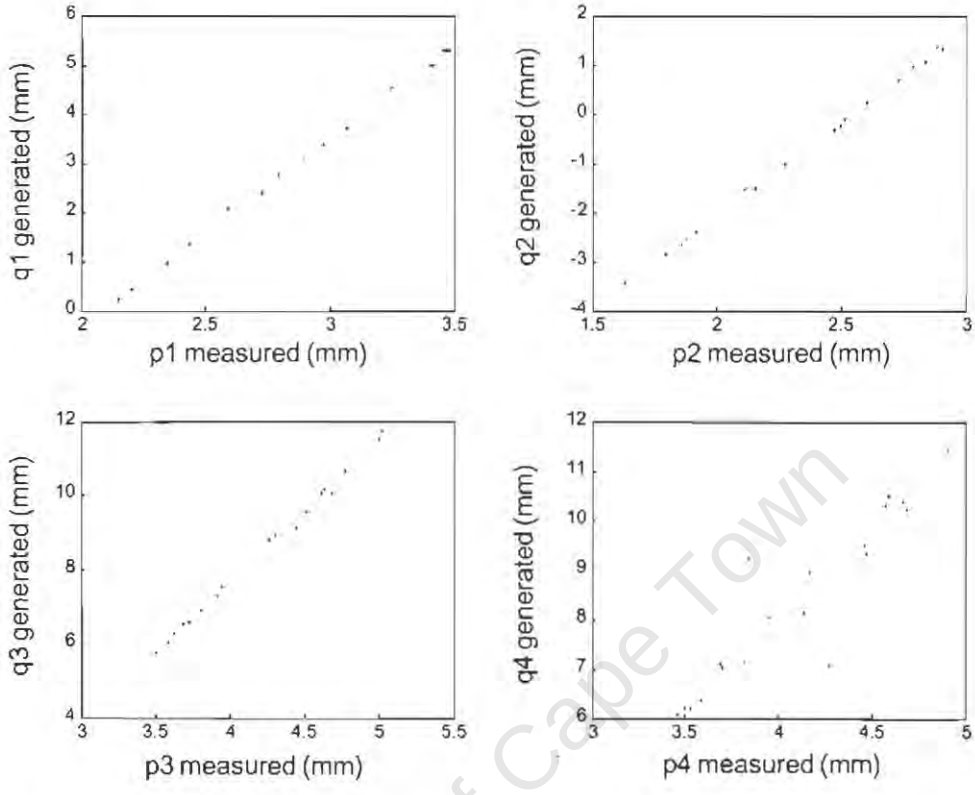


Figure 5.10 Generated vs measured pixel values from Test 5.3.

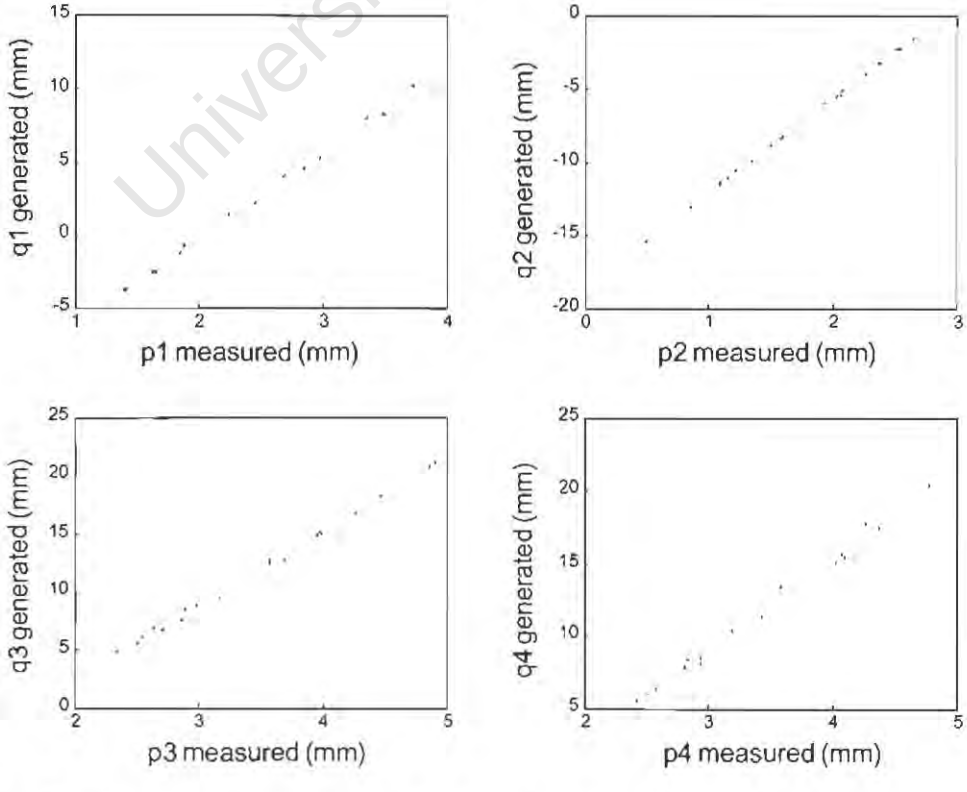


Figure 5.11 Generated vs measured pixel values from Test 5.4.

where the distance to the targets was 1m. The measured pixel values in Figures 5.10 and 5.11 are given by p (1,2,3,4) for each sensor respectively, while the generated values, using the Perspective Projection Theory, are given by q (1,2,3,4). Theoretically the data points in both figures should follow straight lines. The slight dispersion that occurred is a result of the hand positioning of the frame and the cameras, which could best be seen in the graph p4-q4 in Figure 5.10. The range of the axes given in millimetres should be a measure of the position on the sensor where any particular point is projected. The total sensor range is from 0 to 5 mm for the measured points for sensors 1, 3 and 4, while for sensor 2 the total range is from 0 to 6.7 mm. Some of the generated points, however, do not fall within those limits as is seen on the graphs. The scale of the q-values is offset in Figure 5.10 (q2, q3, q4), but the range of the numbers falls within the mentioned physical size of the sensor. In Figure 5.11 the q-values are offset, as well as their range exceeds by far the size of the sensor. These results followed the solution of the inverse problem for the BA Method, which used the initial estimates for the camera parameters. These estimates do not account for hand positioning errors, poor focusing and lens distortion displacement of the principal point which could be a reason for this data not being completely consistent.

In the next step, the generated data was used by both the DLT and Bundle Adjustment algorithms to reconstruct XYZ coordinates of the targets. In the case of DLT, the calibration parameters were calculated prior to the reconstruction, while the Bundle Method used the set of initial parameters that had been used to generate the pixel values. Table 5.8 shows the

	19 point reconstructed with DLT from their projected coordinates, generated using the Perspective Projection Theory			19 point reconstructed with Bundle Adjustment Method from their projected coordinates generated using the Perspective Projection Theory		
Test 5.5	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	1.20E-05	-6.10E-05	-2.10E-04	1.80E-14	1.32E-13	2.39E-13
stand.dev.	7.79E-04	1.16E-03	4.49E-03	1.00E-13	1.94E-13	1.35E-12
abs. mean	5.71E-04	8.61E-04	3.53E-03	6.13E-14	1.32E-13	8.14E-13
abs. stand.dev	5.30E-04	7.83E-04	2.78E-03	8.00E-14	1.94E-13	1.09E-12
abs. max. error	2.06E-03	3.01E-03	1.02E-02	2.27E-13	4.55E-13	3.64E-12
Test 5.6	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	5.23E-05	-2.00E-07	4.89E-04	-1.65E-14	2.51E-13	1.92E-13
stand.dev.	2.59E-04	3.20E-04	8.28E-04	1.55E-13	3.76E-13	1.25E-12
abs. mean	1.99E-04	2.50E-04	6.83E-04	1.03E-13	2.75E-13	8.14E-13
abs. stand.dev	1.75E-04	1.99E-04	6.77E-04	1.15E-13	3.58E-13	9.54E-13
abs. max. error	5.11E-04	6.69E-04	3.07E-03	4.55E-13	1.36E-12	3.64E-12

Table 5.8 Differences in the X,Y and Z direction after coordinate reconstruction with DLT algorithm and Bundle Adjustment Method on 19 points from the calibration frame on Figure 5.8, where their pixel values were generated using the Perspective Projection Theory and camera parameters of the system from Figure 5.9.

results from the reconstruction for each test with both methods. The complete list of these results can be found in Appendix G (Test 5.5 and Test 5.6).

It was also interesting to see how the system would perform when one set of data is used for calibration and the other to reconstruct the target coordinates without having any XYZ information beforehand. In this example the generated data from Test 5.6 was used to calculate the 7 calibration L parameters for the DLT algorithm. In other words the calibration frame was placed 1 m away from the cameras.

	19 point from Test 5.5 (3m) reconstructed after calibration was done with another 19 points from Test 5.6 (1m)			19 point from Test 5.6 (1m) reconstructed after calibration was done with another 19 points from Test 5.5 (3m)		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	0.0002	-0.0002	0.0044	0.0073	-0.0139	-0.0189
stand.dev.	0.0019	0.0008	0.0182	0.0176	0.0227	0.0575
abs. mean	0.0016	0.0007	0.0158	0.0151	0.0171	0.0494
abs. stand.dev	0.0011	0.0004	0.0100	0.0112	0.0203	0.0351
abs. max. error	0.0044	0.0015	0.0373	0.0409	0.0610	0.1027

Table 5.9 Differences in the X,Y and Z directions after coordinate reconstruction with DLT algorithm on 19 points from the calibration frame in Test 5.5, where generated data points from Test 5.6 were used for calibration and vice versa. For both data sets pixel values were generated using the Perspective Projection Theory and camera parameters of the system from Figure 5.9.

At the next step, space coordinates of the points from Test 5.5 (calibration frame positioned 3 m away from the cameras) were calculated using these parameters and the pixel values of the targets. Table 5.9 (left half) shows that this calibration, performed with generated data, can considerably decrease the error, provided there is no change in the cameras' positioning and orientation. When the data sets switch their sides in the process of calibration and reconstruction (right half of Table 5.9) the accuracy decreases slightly, but is still within the submillimetre range. Complete list of results from these calculations is given in Appendix H (Test 5.7 and Test 5.8).

As it was expected, the BA Method produced very accurate results when the forward problem was solved with the data created from the inverse problem of the Perspective Projection Theory. From these various tests it can be concluded that the BA Method is more stable than the DLT method. Comparing the results from the reconstruction, it can also be pointed out that for the data generated, according to the Perspective Projection Theory, the BA Method produces an error several orders of magnitude lower than the DLT. In Tests 5.5 and 5.6 the trend for both methods to generate a higher error in the direction of the depth of the measured space, reported earlier by Georgiev *et al.* (1998) and Vaughan *et al.* (1998), was not retained. There was a regular distribution of the error in all directions.

The major differences between the two methods are the number and physical meaning of the camera parameters. In the case of DLT there are 7 parameters for each linear sensor represented as dimensionless numbers, while in the case of the BA Method there are 11 sensor parameters and their description is given in section 5.3. The first five parameters are measured in millimetres, while the remaining six are dimensionless.

The term $\Delta p = D_k(p_{ik} - p_{ok})^3$ included in the DLT equations, used to model the lens distortion requires an iterative solution. Following this condition to achieve convergence for all the sensors, however, may set the values of the calibration parameters, so that the accuracy is affected. Using perfect data for the sensor calibration shows that the error at the reconstruction stage could be minimized. This could be a useful approach to work out corrections for eventual calibration procedure with real measured data. Increasing the power in the term Δp from 3rd to 4th, as was suggested in equation [3.22] does not bring any significant change in the principal point convergence and the final accuracy of the system.

The BA Method, unlike the DLT method, does not contain a term describing the lens distortion and instead of initial calibration it requires the camera parameters to be given certain values. They could be adjusted further in the process, using data from a set of calibration targets similar to ones shown in Figure 5.8, with known XYZ coordinates and pixel values. An example of such sensor calibration is given in the following section where an optimization procedure of the function from equation [5.21] is described.

5.6 Optimization of the Camera Transformation Function

As was pointed out in equations [5.18] through [5.20] the relationship between the measured image coordinates and the object space coordinates could be expressed as a nonlinear problem. For any number of points greater then 12 it can be assumed that their X, Y, Z values are constants and the 11 camera parameters ($c, p_o, X_o, Y_o, Z_o, L_1, L_2, L_3, L_4, L_5, L_6$) are variables of the object function. The meanings of these camera parameters were defined in section 5.3. Now, if a set of targets with known X, Y, Z coordinates and a certain configuration of the system consisting of four linear sensors is used, the measured pixel values will be represented by the output function,

$$p = f(\xi_1 \ldots \xi_{11}) \qquad [5.25]$$

where the calibration parameters are written as $(\xi_1 \ldots \xi_{11})$. It will also be assumed that due to distortion and other aberrations these pixel values are not perfect. The camera parameters defined by the position and orientation of the sensors could serve as a first estimate in the vicinity of which a minimum of the error function will be sought after.

$$\frac{\partial E(\xi_i)}{\partial \xi_{si}} = 0 \qquad [5.26]$$

The camera parameters used for Tests 5.5 and 5.6 are shown in Table 5.10. The first five parameters are measured in millimetres, while the next six are dimensionless. Using this initial estimate new pixel data was calculated according to the Perspective Projection Theory with no distortion taken into account.

parameter	sensor 1	sensor 2	sensor 3	sensor 4
c	50	50	50	50
p_o	2.54	3.39	2.54	2.54
X_o	250	-250	291	-291
Y_o	907	907	1079	1079
Z_o	-25	-25	-25	-25
L_1	0	1	-0.707	0.707
L_2	-1	0	-0.707	-0.707
L_3	0	0	0	0
L_4	0	0	0	0
L_5	0	0	0	0
L_6	1	1	1	1

Table 5.10 Sensor parameters for the system from Figure 5.9 used to generate the so called "perfect pixel values" according to the Perspective Projection principle.

This pixel data was different from the measured data. The sum of the squares, from the differences between the generated and the measured pixel values for each sensor, will be a measure of the system error. As a next step the initial values from where to start the iterative process of minimizing the error function were chosen to be $[\xi_{1s} \ldots \xi_{11s}]$ and one example of this optimization was done using as initial values the data from Table 5.11. Subsequently the process followed a stepwise calculation of the new pixel values by sequential change of all parameters.

parameter	sensor 1	sensor 2	sensor 3	sensor 4
c	3	3	3	3
p_o	0.01	0.01	0.01	0.01
X_o	225	-275	266	-316
Y_o	882	882	1054	1054
Z_o	-200	-200	-200	-200
L_1	-0.25	0.75	-0.957	0.457
L_2	-1.25	-0.25	-0.957	-0.957
L_3	-0.25	-0.25	-0.25	-0.25
L_4	-0.25	-0.25	-0.25	-0.25
L_5	-0.25	-0.25	-0.25	-0.25
L_6	0.75	0.75	0.75	0.75

Table 5.11 The initial values of the sensor parameters in the optimisation process.

The current value in the process was given by:

$$p_{j+1} = f(\xi_{1j} \ldots \xi_{11j}) + (\Delta\xi_1 \ldots \Delta\xi_{11}) \qquad [5.27]$$

where $\Delta\xi_1 \ldots \Delta\xi_{11}$ were the iteration steps.

As a result such values, for the camera parameters for which the system error would be negligible, were expected to be found. Although it was not absolutely necessary that the new calculated parameters were close in value to the initial estimates, some of them reached almost the same values, while others differed considerably.

This process of optimisation has reached a local minimum for each of the 11 variables. A certain range was found for each variable where the error function was minimal. This value was used further to find out at which iteration step it has occurred. The calculations to achieve the point of the minimum were implemented iteratively using the criterion for a minimal error given by the sum of squared differences between the measured and generated pixel values where the pixel values on the sensor were converted into millimetres

$$E_k(\xi_k) = \sum_{i=1}^{19} [\, p(\xi_k)_i - p_{ik} \,]^2 \qquad [5.28]$$

where

$E(p_o)$ = error function
$\xi = p_o$ = principal point
$p(p_o)_i$ = generated pixel value
p_i = measured pixel value
k = sensor number

If all minima values were found at the same iteration step, then this would have been the case of a global minimum. However, this was not achieved and all minima values were found at different iteration steps. The borders of the error function were not clear enough to apply any stop criterion. Another interesting observation from this process was found when the optimisation was repeated again within the same range for all parameters, and produced lower error values. The minima values were also found to float around a certain point, which could be determined after sufficient number of repetitions of the optimisation procedure.

Table 5.12 contains the result of such an optimization process with the pixel data from Test 5.6, completing 13 error minimization steps where each of these steps contained 500 iterations.

parameter	sensor 1	sensor 2	sensor 3	sensor 4
c	9.60	9.00	4.80	4.60
p_o	2.68	3.78	1.89	1.90
X_o	250.00	-263.50	266.10	-291.00
Y_o	889.20	907.00	1057.30	1054.10
Z_o	133.00	-1.00	-153.00	-86.00
L_1	0.01	-1.00	-0.78	0.81
L_2	-1.01	0.03	-0.72	-0.86
L_3	0.01	0.02	0.07	0.09
L_4	0.06	-0.23	-0.13	0.05
L_5	-0.25	0.13	-0.18	-0.14
L_6	1.00	1.02	0.91	0.85

Table 5.12 The values of the sensor parameters where the error function, defined as a sum of the squared differences between the generated and the measured pixel values for all 19 targets, reached a minimum

One possible reason for reaching a local rather than a global minimum could be the fact that the equations in the Bundle Adjustment Method are independent. This fact is due to the selected arrangement of the sensors as shown in Figure 5.9. A different arrangement of the sensors or applying constraints by adding more equations may solve the problem of finding a global minimum and it could be a potential field for further investigations.

Some other examples show that when only one parameter is optimized the improvement of the accuracy is at the same level as the above and the values of this camera parameter appear to be much closer to the initial estimate.

The following case presents the optimization according to the second parameter, which is the position of the principal point p_o along the sensor. Following the condition for a minimum of the function $E(\xi)$

$$\frac{\partial E_k(\xi_k)}{\partial \xi_k} = 0 \qquad [5.29]$$

it is found at different locations along each sensor and the character of the error function $E_k(\xi_k)$ is shown in Figure 5.12 and in Table 5.13.

	Sensor 1	Sensor 2	Sensor 3	Sensor 4
Distance to principal point (mm)	2.48	3.42	2.42	2.42
Min. Error value (mm)	0.0221	0.0025	0.8901	0.9644

Table 5.13 The values of the minimum for each sensor and and the values of the error function when optimisation was performed according to one parameter - distance to principal point p_o.

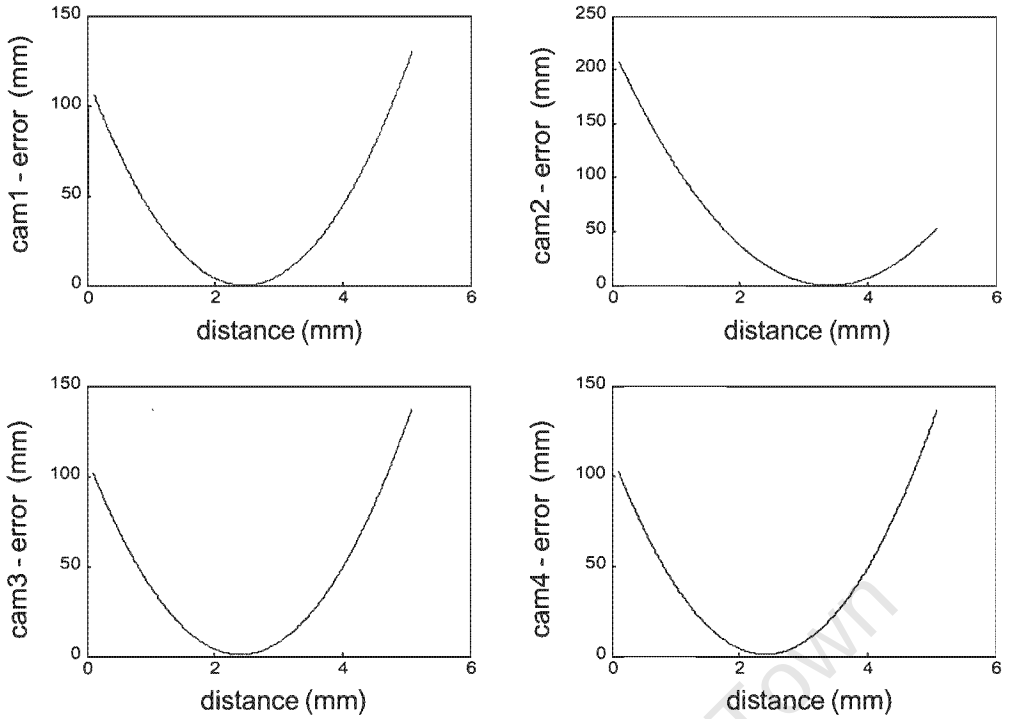


Figure 5.12 Distance to the principal point from the beginning of the sensor assuming that the numbers for p_o from Table 5.10 show the middle of the sensor.

The new sensor values generated using the optimized parameters differed from the measured data by some 10 to 20 pixels in the worst case and the results from the calibration and reconstruction are shown in Table 5.14.

	19 points calibrated and reconstructed with DLT from their optimised pixel data			19 points calibrated and reconstructed with BA from their optimised pixel data		
	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
mean	2.90E-04	7.00E-05	2.00E-05	3.44E-14	-7.78E-14	-2.15E-13
stand.dev.	1.30E-03	2.40E-03	4.10E-03	1.13E-13	3.71E-13	8.76E-13
abs. mean	9.00E-04	1.80E-03	3.40E-03	8.53E-14	2.45E-13	6.94E-13
abs. stand.dev	1.00E-03	1.60E-03	2.30E-03	7.91E-14	2.84E-13	5.54E-13
abs. max. error	3.10E-03	5.80E-03	9.80E-03	2.27E-13	9.09E-13	1.82E-12

Table 5.14 Differences in the X,Y and Z directions after coordinate reconstruction with DLT algorithm and Bundle Adjustment Method was completed on 19 points from the calibration frame, where their pixel values were generated using the Perspective Projection Theory and the optimised sensor parameters.

This parameter optimization procedure improved the results significantly from the reconstruction after the difference between the generated and the measured pixel values had been minimized. The definition of the sensor parameters, however, could be considerably

different from the initial set up. This difference could come from improper focusing and positioning error and is ultimately compensated for after the parameters are optimized. The results from the reconstruction in Table 5.14 show that the error remains within the same range shown in Table 5.8 where the tests of the 19 points were done, using DLT for calibration and DLT and the BA Method for reconstruction of the generated pixel data.

The process of calculating the sensor parameters in order to optimize the camera transformation function included several hundred iteration steps and the updating of each parameter after every step. This proved to be a time consuming operation, which depended on the hardware performance. A Pentium II computer operating at 466 MHz, under Windows 98, running the Matlab 6.1 routine to implement the optimization, needed 2 to 4 hours to complete the calculations.

The study of the actual effectiveness of the DLT and BA Method, when calibrating and reconstructing coordinates in 3D, showed that the BA Method is more stable during the calibration stage and more accurate at the reconstruction stage. In fact both methods could be used with a potential 3D coordinate measuring system, provided that an accurate calibration was done to determine the correct sensor parameters, thereby compensating for the lens distortion as well. At the reconstruction stage, however, the fact should be considered that for each point in the measured 3D space a set of pixel values exists, from which the 3D coordinates could be reconstructed uniquely. These sensor values could be calculated using the Perspective Projection Theory. On the other hand, the measured sensor values could be significantly different from the projected or "perfect" values. For this purpose it is recommended that the target space be accurately mapped and measured. On this basis, correction values should be derived from the measured and the projected differences for every part of the volume where eventually targets could be imaged. The central part of this measured space and the sections close to the optical axes are expected to require smaller correction values, whereas those sections on the sides of the measured volume are expected to require much larger values. These corrections could also be dependent on the position and orientation of each sensor and may differ in their magnitude and sign from one sensor to another.

Some commercial systems use different sets of calibration targets. In many cases a static calibration is done first. This provides the rough values for the calibration parameters, which is followed by a dynamic calibration that provides a fine adjustment to the system. There is also a factory calibration, which uses a large number of targets describing the entire measured space. The use of such net of targets could help for the precise determination of the sensor parameters and could utilise the optimisation when searching for a global minimum of the error function. In this process some of the parameters may need to be changed significantly and thus, to rearrange the geometry of the detector and the position and orientation of the sensors. This field is recommended for future study with a prospective to analyse in detail the character of the camera transformation function based on multiple variables.

Chapter Six

Target Identification and Tracking

In the last two chapters the emphasis fell on the different algorithms for the 1D and 3D determination of target positions in space. Attention was drawn to the significant problem of improving the accuracy of these calculations. In this new chapter the objective is to show how all the techniques for signal processing and coordinate reconstruction could help track multiple targets in 3D space. In addition, the identification of the targets during the process has proved to be of crucial importance.

As pointed out earlier, the detection of the targets from the linear sensor could be either a sequential or a simultaneous process. In the first example, targets are activated by an external device which controls, by means of infrared (IR) light pulses or radio frequency (RF) pulses, an electronic circuit that switches each target on and off over a period of a few milliseconds. This sequential technique is also known as Time Division Identification (TDI) where the targets are activated one at a time. The advantage of this mode is faster signal processing and the unique identification of every target. The disadvantages stem from the need for additional circuitry for the targets, the larger space occupied by the target and the complex process of target activation. In this application linear sensors, used as detecting elements instead of rectangular sensors, could provide a better solution. This is due to fact that they operate with less data and at much higher sampling rates where only one target needs to be active during every reading cycle.

The second example, the simultaneous process, usually occurs when passive targets are used and reflect the signal back to the sensor. This signal could be an infrared light or other types of signals such as radio or ultrasonic signals which are widely used in radar applications. The identification of these targets is considered to be more difficult because of the simultaneous detection and often the possible ambiguous confusion of their position on the sensor. Two-dimensional sensors used for this purpose are more reliable but they operate with more data and require more computational time. On the other hand linear sensors, equipped with cylindrical lenses, can improve the performance of the tracking system with less data to be processed but the potential number of targets used will decrease because these sensors have only one dimension.

In addition to the above-mentioned factors, one more aspect should be taken into account when solving the problem of target tracking. During the detection phase it may happen that the sensor registers either few or no targets. The situation with missing data complicates both identification and tracking. The reasons for this are numerous but usually occur when active targets, such as infrared light emitting diodes (IR LEDs) are moved in the space, the

problem arising from their limited angle of light radiation (about 120°). When passive retro-reflective targets are used the problem usually lies in the low intensity of the reflected light which causes a poor signal-to-noise ratio. Various methods to solve this problem have been developed in the past, some of them using a statistical approach, others based on time-space determination of the signals. In this study the problem will be formulated as a 3D coordinate reconstruction at close range. The following experiment s include simultaneously detected, moving, active and passive targets. The noise, produced from the reflection of the background, was removed from the signal at the pre-processing stage to avoid additional ambiguity at the tracking stage.

6.1 Tracking a pair of active moving targets

6.1.1 An algorithm for one-dimensional tracking

The problem of tracking two active targets with four linear sensors, will be considered as one-dimensional in the first instance and could be referred to as one-dimensional tracking. In this example, as it has been noted, tracking and identification stages will go together to achieve a unique correspondence between tracks and targets.

One model as demonstrated in the literature (Bar-Shalom *et al.*, 1990) is called the dynamic model and represents the target state with the following equation:

$$\mathbf{X}_t(k+1) = \mathbf{F}\,\mathbf{X}_t(k) + \Delta T \mathbf{w}_t(k)\mathbf{q}_t \qquad [6.1]$$

where

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_t \\ \dot{y}_t \end{bmatrix}$$

is the current state vector and k represents the discrete time index

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is the transition matrix for the sampling period ΔT, $\mathbf{w}_t(k)$ represents white Gaussian noise

with zero mean and known variance and $\mathbf{q}_t$ represents the variance of the velocity. This is a model of 2 coordinates and 4 variables applied to 2D cameras. If the second term on the right hand side of equation [6.1] is absent, the model will represent a constant velocity motion.

When considering the use of linear sensors, [6.1] can be reduced to just one dimension with 2 variables:

$$\begin{aligned} X_t(k+1) &= X_t(k) + \dot{X}_t(k)\Delta T \\ \dot{X}_t(k+1) &= \dot{X}_t(k) \end{aligned} \qquad [6.2]$$

6.1.2 Data acquisition from moving targets

In order to implement targets moving in 3D space, a system consisting of a DC motor driving a horizontal bar, as shown in Figure 6.1, was constructed. At both ends of the bar two active targets (infrared LEDs) were attached, separated by a distance of 292 mm and activated to produce continuous IR light. The motor was rotated at a constant velocity, so that the bar with the targets could complete a certain number of turns at a given time interval. To control this speed a signal from the motor was passed to a pulse encoder operating at 1000 pulses/ revolution and connected further to an oscilloscope in order to display the pulse rate.

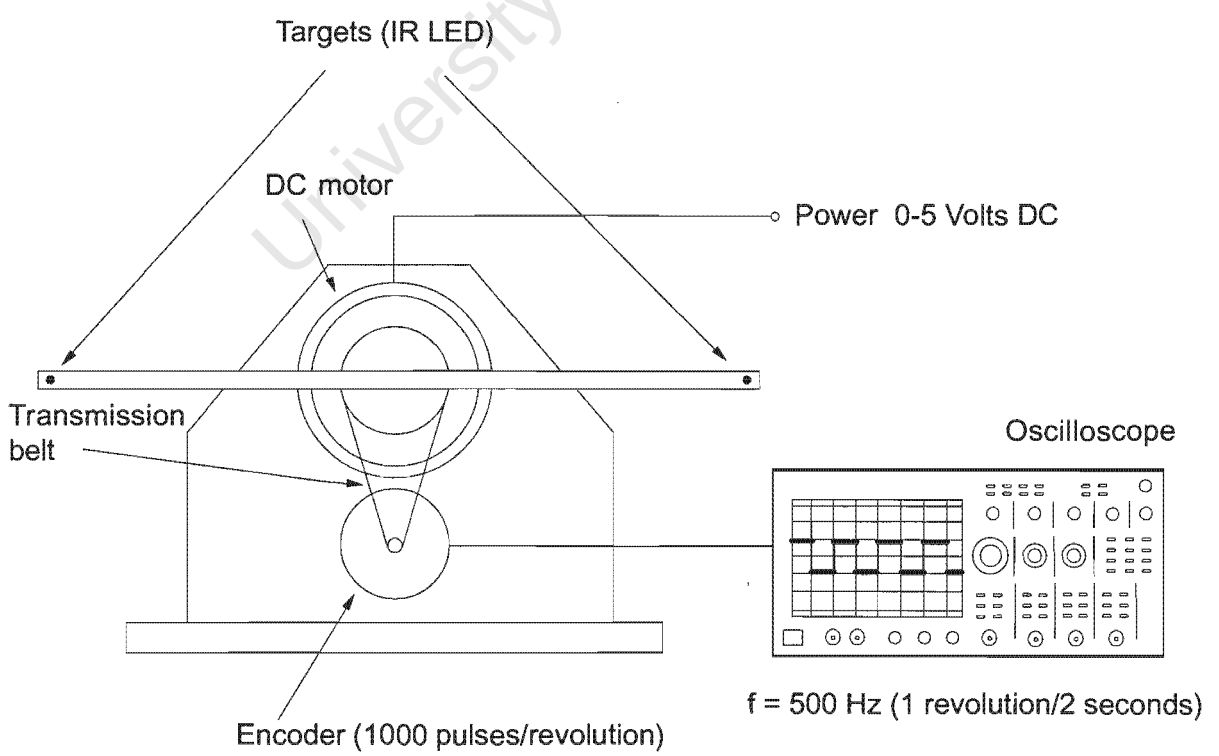
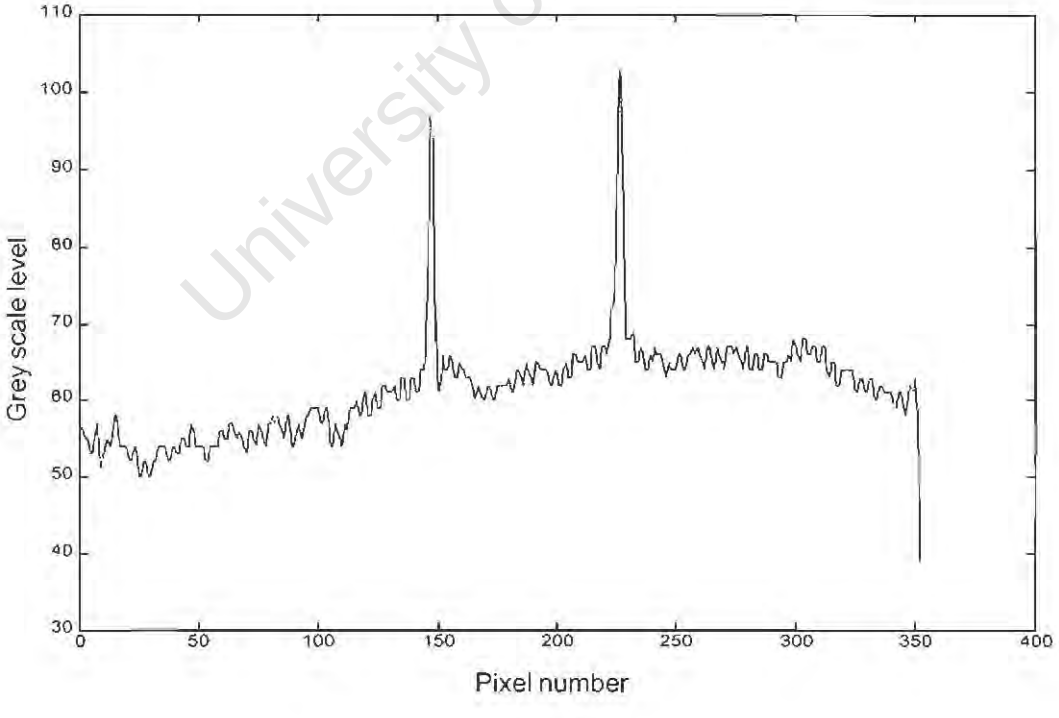


Figure 6.1 A system of a DC motor controlled by frequency, driving two active targets in a circle which were tracked by four cameras equipped with cylindrical lenses.

The power supply voltage needed to produce a motion of the bar equivalent to one full turn completed in 2 seconds was used to control the motor. As a detecting system the set up displayed in Figure 5.6 in Chapter 5 was used, where four cameras with cylindrical lenses captured images of the targets at about 30 frames/second. The camera system was calibrated according to the procedure which uses 19 markers from the calibration frame shown in Figure 5.5.

a



b

Figure 6.2 Two active targets attached to the horizontal bar captured from a camera with cylindrical lens (a) and their line profile taken across the image (b) during the tracking process. During rotation these two peaks move laterally along the sensor crossing over every half cycle.

The rotating targets were recorded on videotape from each of the four cameras. In the beginning there was an interval of 10 seconds when targets were stationary, followed by another 10 second interval during which the bar with the targets completed a specified number of turns. After this cycle it was recorded again during a 10 second interval of the stationary targets. The intervals with stationary targets were necessary for the synchronization of time during the tracking. The entire motion sequence for each camera contained up to 350 frames including a short interval when targets were stationary at the start and at the end. All images resembled the one in Figure 6.2a and the line profile taken across the image corresponded to the linear sensor reading (Figure 6.2b). In order to obtain only the signal from the target, signal processing, including filtering and background removal followed by threshold subtraction, was applied to each line profile. Thereafter their centroid values, representing the positions on the sensor, were calculated. Two subsets of points, with the positions of the targets on each frame, were formed from this data. Where there was a target detected in every frame, the total number of points in a subset had to amount to 350.

The points from the two subsets were plotted on a graph (Figure 6.3) where the pixel positions along the sensor were shown on the ordinate and the frame number on the abscissa.

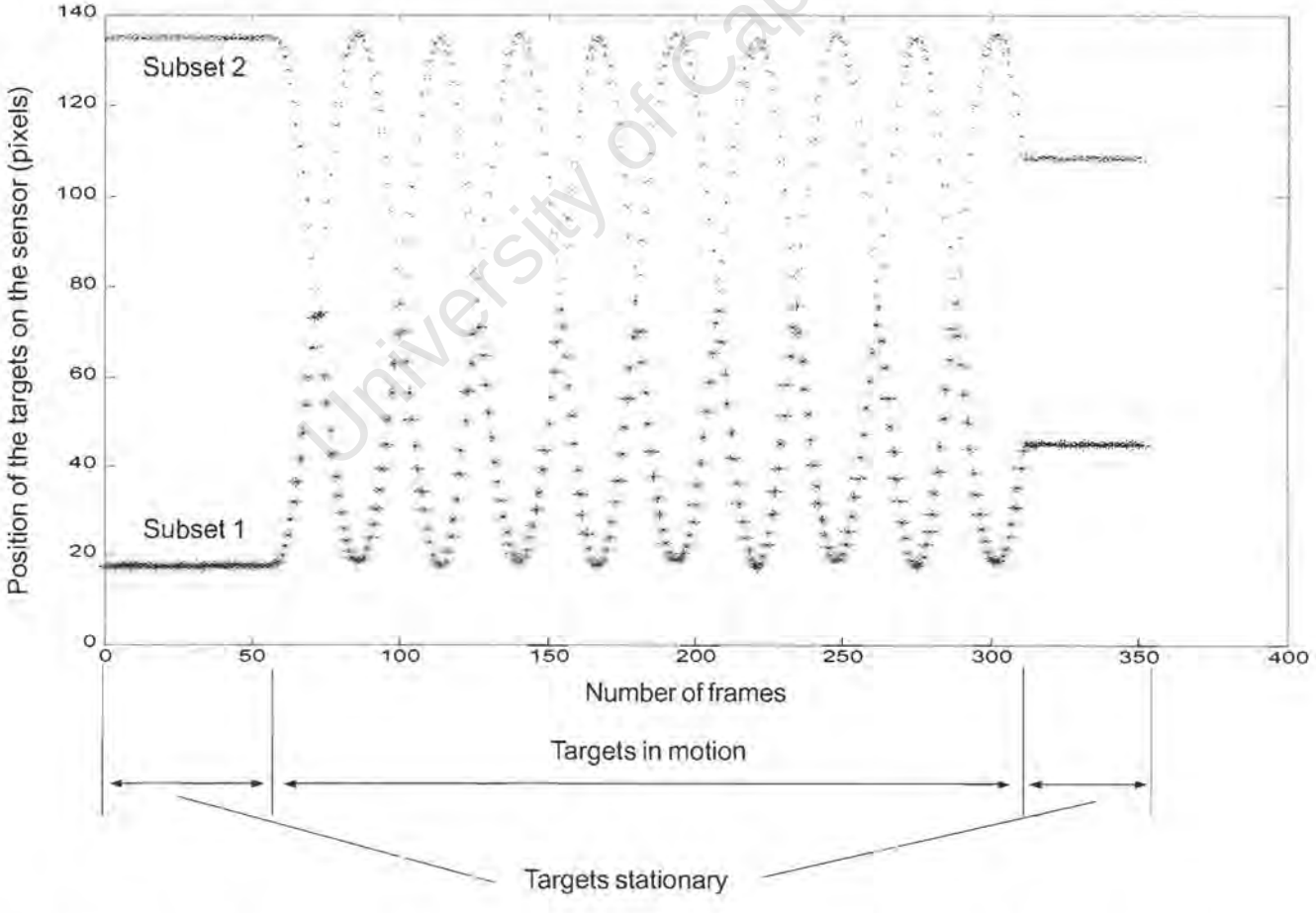


Figure 6.3 Raw data points stored in Subset 1 (plotted in green '*') and Subset 2 (plotted in red 'x') represent the positions of Target 1 and Target 2 within each section of the sensor along all the 350 frames.

These points, shown in two different colours, corresponded to the projections from the left and the right half of the circle along which targets were moving (where the sensor was positioned at 0°). The actual trajectory for each target could be constructed from the points in Subset 1 and Subset 2, forming two intersecting sine curves out of phase by 180°. As pointed out earlier, missing data was reported in some of the frames where either only one target or no target was captured.

The algorithm placed the data points into Subset 1 and Subset 2 in the same order it detected the centroid values of Target 1 and Target 2. As the targets had changed their positions alternatively on the sensor, the pixel values in Subset 2 did not refer only to Target 1, but to Target 2 as well. In other words, the pixel positions on the curve Subset 2 corresponded to the physical position of Target 1 for only half of the sine curve periods (every even period). The positions from all odd periods therefore corresponded to Target 2. This data order had to be rearranged by identifying each target individually and the tracking process could only continue after the correct positions had been determined.

6.1.3 The process of target tracking

In the next section a simple tracking procedure is described for the two targets in the test. This is the minimum number of targets in the case of multiple targets over a simple trajectory in the shape of a circle. The method is based on recorded data and performed off-line, but it could be used in on-line procedures as well. The model from [6.2] was applied to calculate the predicted position of the target and its velocity over the entire data set, point by point, following the reception from each sensor scan.

At this first stage, a quadratic curve fit over a certain number of points (n = 7), was calculated to evaluate the trajectory of the target. This window function was applied from the beginning of both data sets for Target 1 and Target 2 and the fitted data was denoted as Track 1 and Track 2. During tracking the point lying next to the points within the window was calculated as a predicted position lying on the fitted curve. The window with the approximated data was then moved one point forward, until the calculation from the last frame was completed. These calculations were also used to determine the initial state of the targets, so that Target 1 was associated with Track 1 and Target 2 with Track 2. The accuracy was acceptable, not exceeding a fraction of a pixel. The continued process was described for Target 1 and was repeated in the same way for Target 2. The next objective of this process was to associate only the data from the initial Subset 1, corresponding to the physical positions of Target 1, by using the approximated data. In addition, the results were stored in a vector called Track 1, transferring data samples between Subset 1 and Subset 2. The same was done for Target 2 and its data was stored in a vector called Track 2. The associated data identified the two targets from the pixel data captured at each scan.

At the second stage, when the algorithm came to a point of intersection between the current tracks of Target 1 and Target 2, the object was to find out which subset contained those pixel values related to the physical position of Target 1. At this point the target had already passed to the other section of the linear space. Technically, this means that in order to follow the sinusoidal curve correctly for Target 1 after the intersection point, the algorithm should stop tracking data from Subset 1 and start tracking data from Subset 2. The decision was taken by checking the magnitude and sign of the linear velocity vector, calculated between the points from the last two scans in accordance with the tracking model.

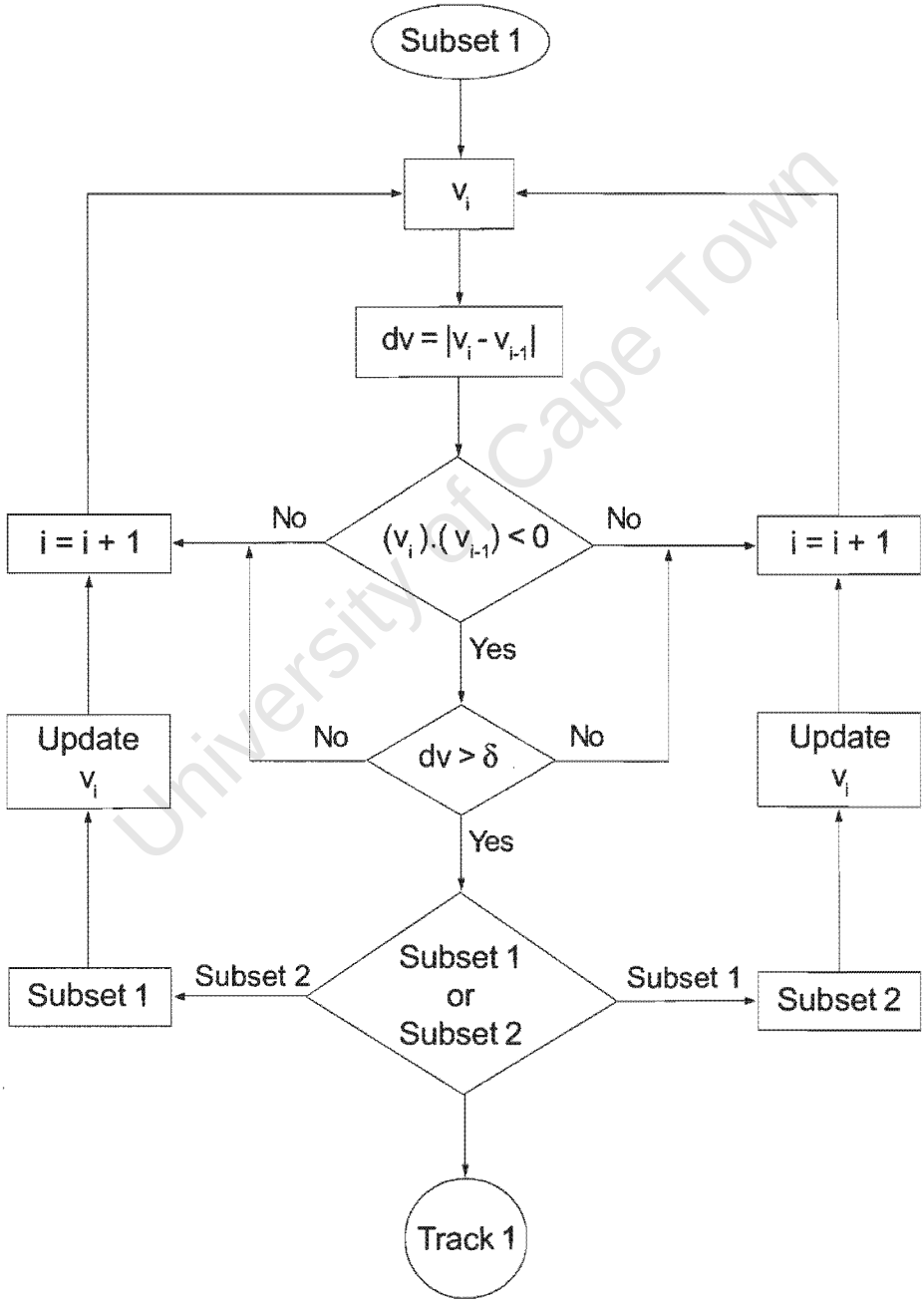


Figure 6.4 Block diagram for the algorithm to track and identify a moving target based on the dynamic model given by [6.2] and modified to take into account the velocity change.

Theoretically the sign of this vector should only change at the end points of the cycle, *e.g.* the points lying on the diameter of the circle, perpendicular to the longitudinal axis of the lens. These end points were also projected on to the sensor as end points of the lateral motion and they lay on the arcs of each curve. At these positions the magnitude of the velocity vector should have also reached a value close to zero as it was calculated as a fist derivative of the travelled distance.

One the other hand, according to the data from Subset 1 and Subset 2 there was also a change of the sign around each intersection point because the targets changed their position from one section of the linear space to the other. This point was essential for the decision taken by the algorithm whether to change the current track or to stay on the same track. The distance travelled by the target along these sections was almost a linear function of time and rapid changes in the magnitude of the vector were not likely to occur, based on the assumption that the velocity had fulfilled the criteria of a monotonic function (nearly constant velocity model). At the end points of the lateral motion there was a change of the sign of the velocity vector. The magnitude of the vector underwent a slight absolute change around the extremal points on the curve.

For this reason, whenever the change was below a certain value (velocity limit δ), the algorithm kept the same track. At the point of intersection the change of the sign (in the signal) occurred together with a significant change in magnitude. If it was over the velocity limit, this meant that the current track had been taken by the other target, and that the algorithm had started taking data from the other subset, in keeping with the character of the motion.

In this example the character of the motion was not a perfectly constant velocity type and this fact was accounted for in the suggested model. Velocity was calculated using a differential approach and the velocity limit δ was provided with a larger tolerance for error compensation. Figure 6.4 presents a block diagram illustrating the steps which the algorithm had performed to track Target 1 from its initial data set, Subset 1, and Target 2 from Subset 2.

Prior to the step of tracking, another procedure was applied to fill the gaps of missing data. These gaps occurred when either the target remained undetected or two targets covered each other, so that the sensor saw only one target. The missing points in the subset could be recreated later by using a target's motion behaviour from the previous points and inserting a point based on calculation of this data. Such an approach is not very accurate especially for the nonlinear sections and for the sections with successive missing points, which are preceded by nonlinear sections. As a result of this inaccuracy the newly created points may differ from the real pixel values. Generally, however, a considerable improvement was observed in the tracking process.

An example of accomplished tracking of Target 2, which initially started from Subset 2, is given in Figure 6.5 and in Figure 6.6 the linear velocity and acceleration along the sensor of the same target are shown.

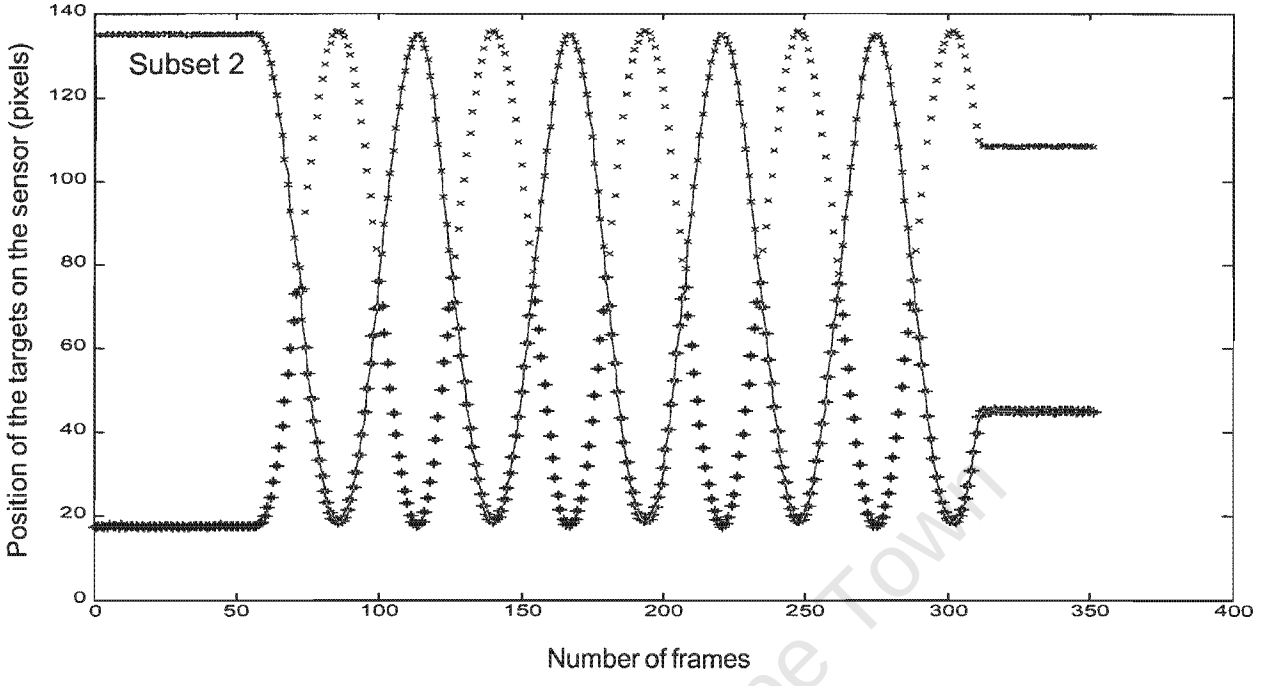


Figure 6.5 Data points from Figure 6.3, Subset 2 (plotted in red 'x') used to track the motion of Target 2 by the algorithm from Figure 6.4. The solid line shows the positions of Target 2 along its track for each of the 350 frames.

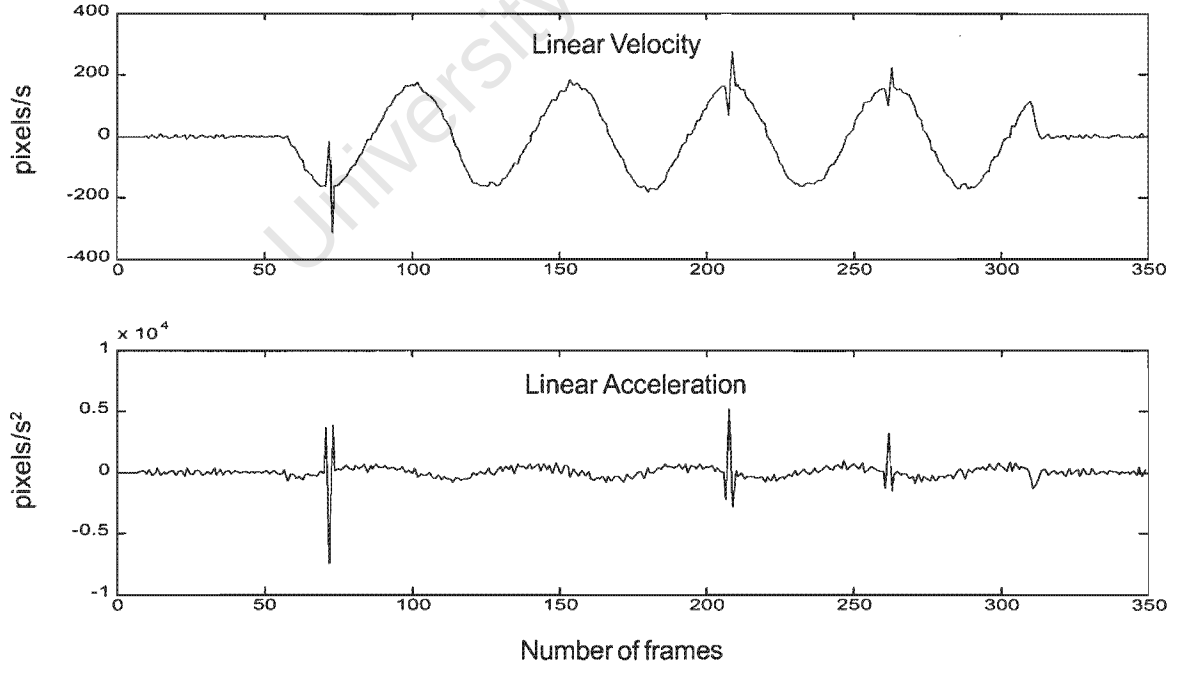


Figure 6.6 Linear velocity and acceleration of Target 2. The spikes on the velocity curve are due to incorrect identification at the intersection points, where the algorithm changes the direction of the current track temporarily. The reason for this is the low and unstable sampling rate of the sensor and the digitising hardware, and the velocity tolerance introduced in the algorithm.

6.2 Tracking multiple passive targets

6.2.1 An algorithm to improve signal-to-noise ratio

The tracking in the last example was utilised after the positions of all targets had been calculated and stored in vectors. If those targets were activated sequentially, the identification would have been unique and the deviation of the kinematic parameters would have been minimal. However, this effect could not be avoided in the case of passive targets where the detection by the sensor is simultaneous and the received signal is much noisier. Subsequently, if more than 2 targets are tracked the ambiguity will increase and the reports of "false" targets and "wrong" tracks would be more frequent. The algorithm for identification and tracking described earlier, in section 6.1.3, which is based on an entirely deterministic approach, would not be efficient enough to achieve a satisfactory level of accuracy.

An improved method based on probabilistic data association between targets and tracks has been suggested to estimate the transition probabilities of each target state (Bar-Shalom *et al.*, 1990). Another method described by Kurien (1990), called the hybrid state model, uses discrete variables when solving the target association problem and uses continuous variables for the track estimation problem.

When considering the character of the initial data from a linear sensor detecting passive targets, it can subsequently be assumed that after the signal processing stage some noise may still appear on the scan, and that this noise should be removed prior to the start of the tracking. Following the idea of the above probabilistic methods, the tracks will be initiated immediately from the first scan, and then the objective is to reduce the number of "false" targets by decreasing the noise to a minimum. This can be achieved by using an integral evaluation of the signal containing targets, assuming that they will acquire the larger portion from the activated pixels on the sensor. This can be expressed as

$$S = A_{t1} + A_{t2} + \ldots + A_{tn} + N_1 + N_2 + \ldots + N_m \qquad [6.3]$$

where $A_{t1}, A_{t2}, \ldots A_{tn}$ are the parts of the integrated signal from the targets and $N_1, N_2, \ldots N_m$ are the areas of the integrated noise. These may be represented as sums of the pixels contained in each area

$$A_{tn} = \sum_{i=p_r}^{p_s} p_i \qquad N_m = \sum_{i=p_u}^{p_v} p_i \qquad [6.4]$$

The sums calculated in this way may happen to have similar values if the noise has occupied

a large number of pixels. For this reason a weighted sum is introduced to eliminate the noise from the targets by multiplying the integrated sum by the maximum pixel value P_m found within the area.

$$w_t = A_{tn} = P_m \sum_{i=p_r}^{p_s} p_i \qquad w_n = N_m = P_m \sum_{i=p_u}^{p_v} p_i \qquad [6.5]$$

The sums formed by noise, which contain small maximum values, will be eliminated from further evaluation, while the sums from the targets are expected to retain or increase their values. The effectiveness of this filtering is shown in Figure 6.7 and Figure 6.8 where 3 passive targets were captured from one of the four linear sensors for over 250 scans. These 3 targets were placed on the same horizontal rotation bar, one at the centre of rotation and the other two on the sides at different distances to form two individual circular trajectories.

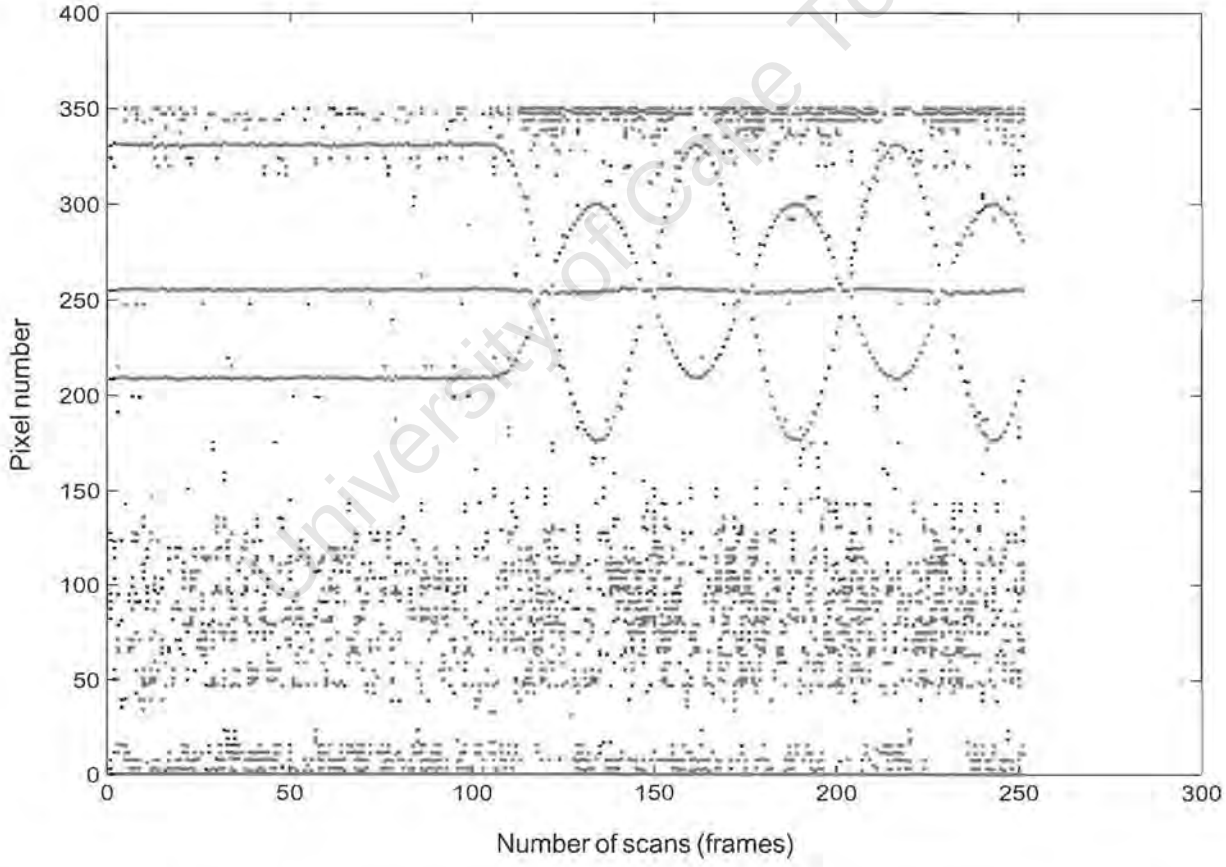


Figure 6.7 Scans of 3 rotating passive targets before weighted data elimination was applied.

In Figure 6.7 it is seen that this low threshold type of detection leaves large abundance of points that could be recognised as "false" data. The approach used in Chapter 4, consisting of filtering and background reduction and high level thresholding, which completely removes the noise has the disadvantage to discard from the signal target positions characterised by

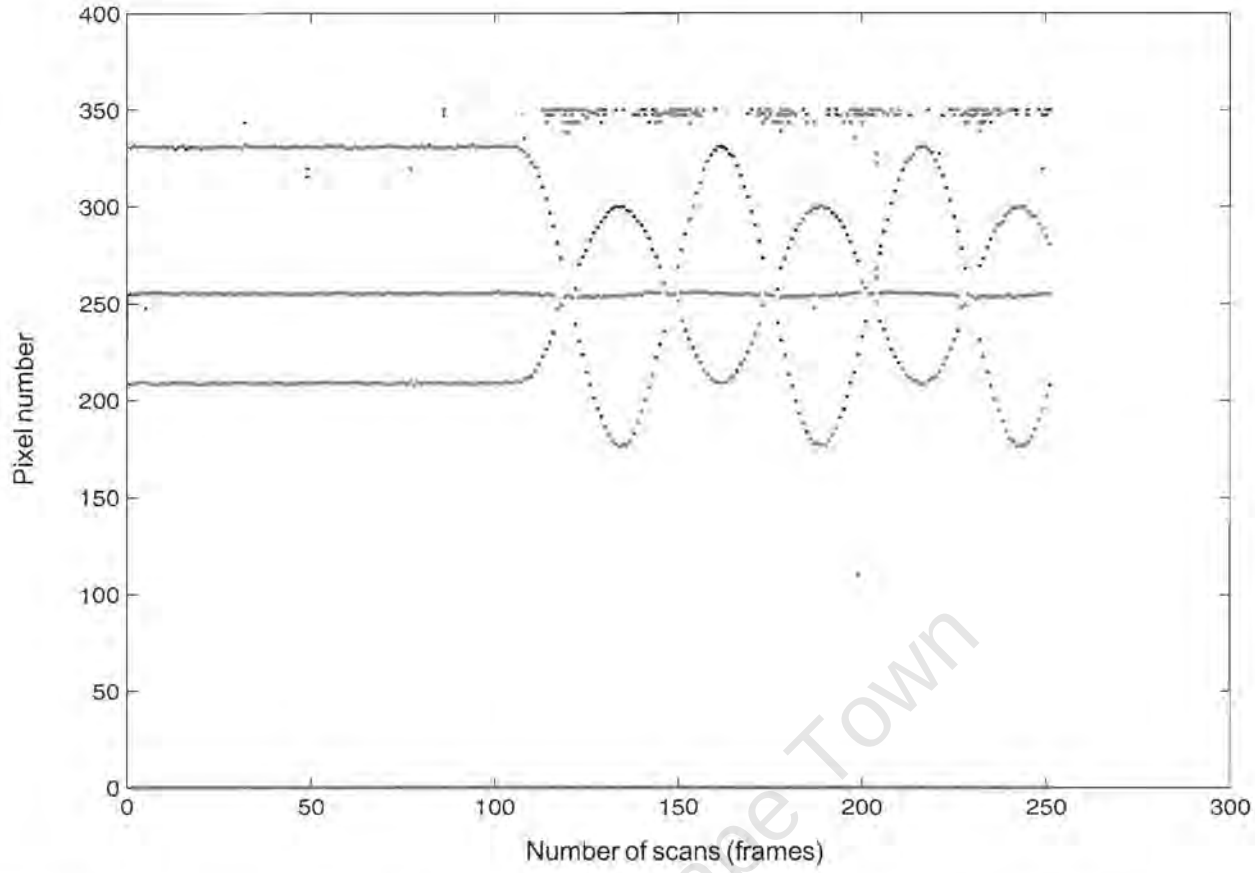


Figure 6.8 Scans of 3 rotating passive targets after weighted data elimination was applied.

low intensity. This is another reason for detection of missing points at the signal processing stage, although the primary cause was pointed out to occur during the data acquisition, because of limited angular irradiance of the light emitting diodes and low reflection level of the retroreflective markers.

Low level thresholding followed by intergral evaluation of the targets compensates this disadvantage providing more complete data sets describing the motion curves, but compromises by leaving small quantities of remaining noise. This noisy data will be included further in the 1D tracking process and it will be the proirity for the tracking algorithm to make a decision for their elimination.

6.2.2 A probabilistic model to determine the target state

As an essential part of the tracking algorithm, the dynamic model from [6.2] is used for trajectory (positions) and velocity estimation. The situation at which the target occupies certain track (or its state as it was defined in section 2.5) could be expressed through the conditional probability

$$P(M(k)) = P(M(k)_{j+1} = N | M(k)_j = K) \qquad [6.6]$$

where the following notations are used: P(M(k)) is equal to the probability of target M(k) at scan j+1 to be associated with track N, provided that at scan j it was associated with track K. This could also be referred to as a condition where the targets will exchange their tracks, to which the appropriate criteria should be introduced.

In the following set of circumstances the focus fell again on several passive targets which were simultaneously detected by the sensor and appeared on each scan together with noise, also recognised as "false" targets. Tracks from every target were initiated from the first scan and were constructed according to the data evaluation from the dynamic model. Then the state of each target was estimated and the probable association region for the next scan was determined as follows.

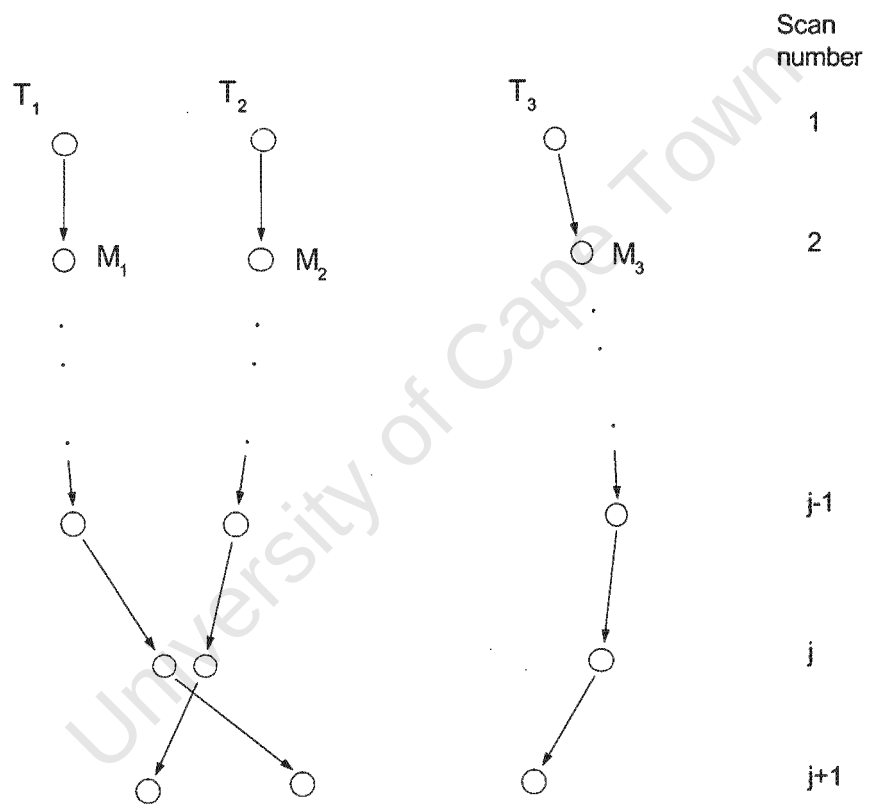


Figure 6.9 Track initiation with further state and track evaluation for multiple targets.

It was assumed that the velocity of the target along the 1D sensor was within the interval $[-v_{max}; v_{max}]$, where the mean value was taken between each two scans, and 2 standard deviations were then applied on each side of the mean. This was the initial estimation of the probable position on the sensor. On the following scan some of the kinematic parameters of the target might have changed and this would have affected its estimated position as well. If at the start an initiation process for targets M_1, M_2 and M_3 to tracks T_1, T_2 and T_3 has begun as shown on Figure 6.9 then for the scan j we can estimate the velocities of each target as

$$M_1 - v_{1j} \qquad M_2 - v_{2j} \qquad M_3 - v_{3j}$$

and for the further target position we can obtain

$$X_{j+1}(k) = X_j(k) + v_j(k)\Delta T + 2\sigma_{v_j}(k)\Delta T \qquad [6.7]$$

where for the k^{th} target

v_j = the target's mean velocity
σ_{v_j} = the standard deviation of the velocity
ΔT = the sampling period

From equation [6.7] the position interval occupied by the k^{th} target will be

$$X_{j+1} \in [X_j - \Delta X_j; X_j + \Delta X] \qquad [6.8]$$

where $\Delta X_j = v_j(k)\Delta T + 2\sigma_{v_j}(k)\Delta T$.

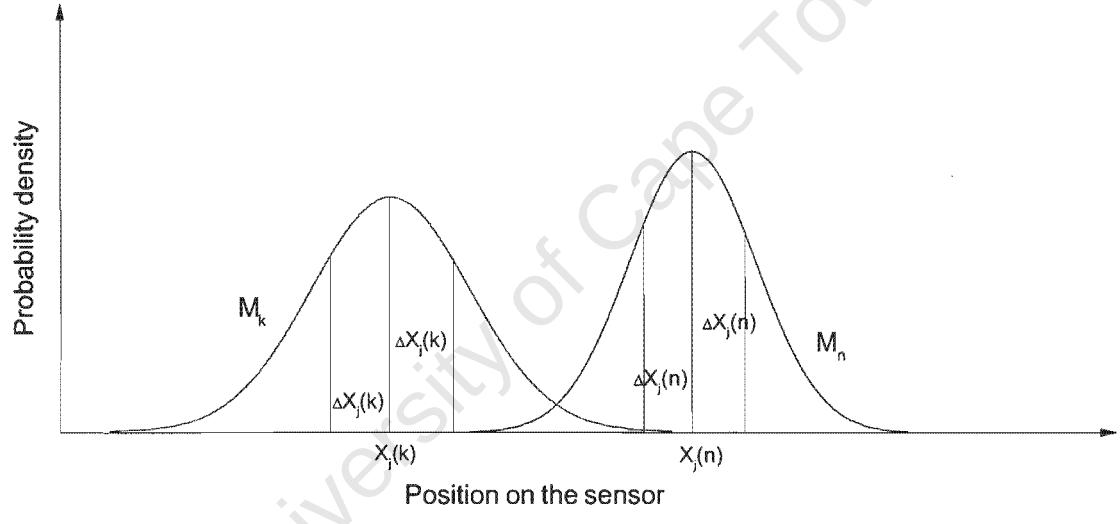


Figure 6.10 Probability density function of neighbouring targets showing transition of the state from separate tracks.

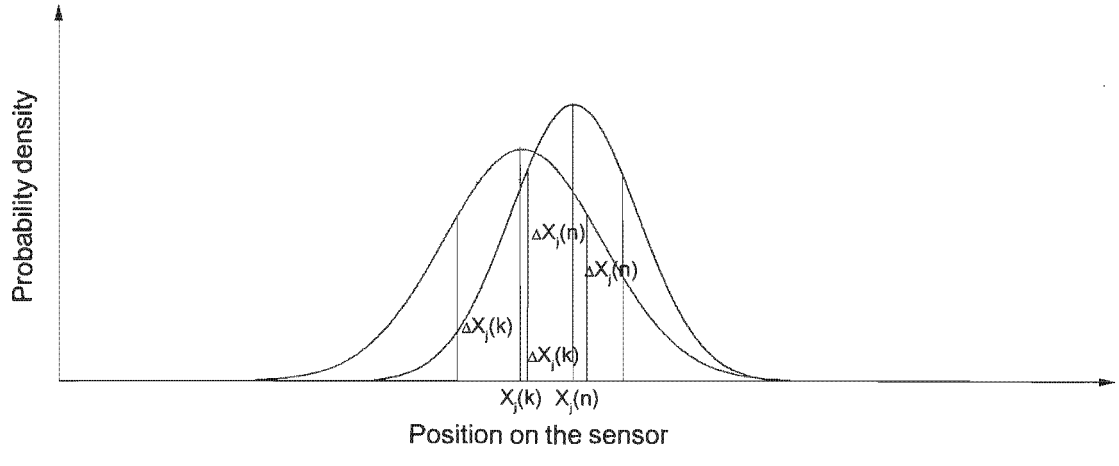


Figure 6.11 Probability density function of neighbouring targets showing transition of the state to the same track.

The condition determined from equations [6.7] and [6.8] is expressed graphically in Figures 6.10 and 6.11, where a pair of neighbouring targets are represented by their probability density functions (PDF) and occupy certain positions on the sensor. The area of the intersection corresponds to the probability of the two targets exchanging their tracks according to three possible conditions:

- the two targets from different tracks occupy the same track;
- the two targets from the same track split into separate tracks;
- the two targets from separate tracks have their tracks cross over

To take a decision in this situation it is necessary to apply a probability threshold, for example 0.5, where the first situation will occur at a higher level of probability than the threshold at scan j+1 (Figure 6.11), provided that it has been lower than the threshold at scan j (Figure 6.10). The second situation will occur at a probability level lower than the threshold at scan j+1, provided that it has been higher than the threshold at scan j. The third situation will occur if the probability level changes from a level below the threshold at scan j to a level above the threshold at scan j+1, and then drops again to a level below the threshold at scan j+2. If the probability value for a certain target does not change according to the threshold value for three siccessive scans, the target should not change its state to the current track which it occupies. This probability is therefore expressed as

$$P(M_k) = P(M_{j+1}(k)= N|M_j(k)= K) = \frac{P(X_n) \cap P(X_k)}{P(X_n) \cup P(X_k)}$$

$$= \frac{P(X_n) \cap P(X_k)}{P(X_n) + P(X_k) - P(X_n) \cap P(X_k)} \qquad [6.9]$$

where $P(X_n)$ and $P(X_k)$ are the probability density functions of targets N and K with certain positions along the sensor on the j^{th} scan.

Finally substituting [6.7] and [6.8] into equation [6.9] gives the estimated probability for the target intersection area through their position intervals:

$$P(M_k) = X_j(n) - \Delta X_j(k) - \frac{\Delta X_j(n) - \Delta X_j(k)}{2} \qquad [6.10]$$

Equations [6.7] and [6.10] serve to evaluate at each scan the association of a given target to a certain track.

The restriction of using the constant (or nearly constant) velocity model will prevent "false" targets from being associated with existing tracks. The reason for this is that their probability of exchanging tracks will be much higher and their probability to retain their current tracks will be much lower, which would ultimately result in their termination (Bar-Shalom *et al.*, 1990). In the process of track initiation such targets will also have a "short life" after the start and such tracks will end with the target's disappearance.

6.3 Reconstruction of coordinates for tracked targets in 3D

To complete the final stage of the process, tracking in 3D has to be performed, which means that multiple sensors are needed, and the two models explained earlier will be applied after the 3D reconstruction. These steps of 1D tracking and identification, followed by 3D reconstruction and target-to-track verification, should be performed for each scan and could be referred to as 1D → 3D transformation.

Coordinate reconstruction of moving targets can be done using Direct Linear Transformation and the Bundle Adjustment Method from Chapter 5. After the tracking of each target from the previous example had been completed and the pixel values for each of the four sensors had been rearranged, in such a way that Target 1 and Target 2 always stayed on Track 1 and Track 2 respectively, their space coordinates could be reconstructed from their pixel values, using either of the two methods above. In order to calibrate the system prior to reconstruction,

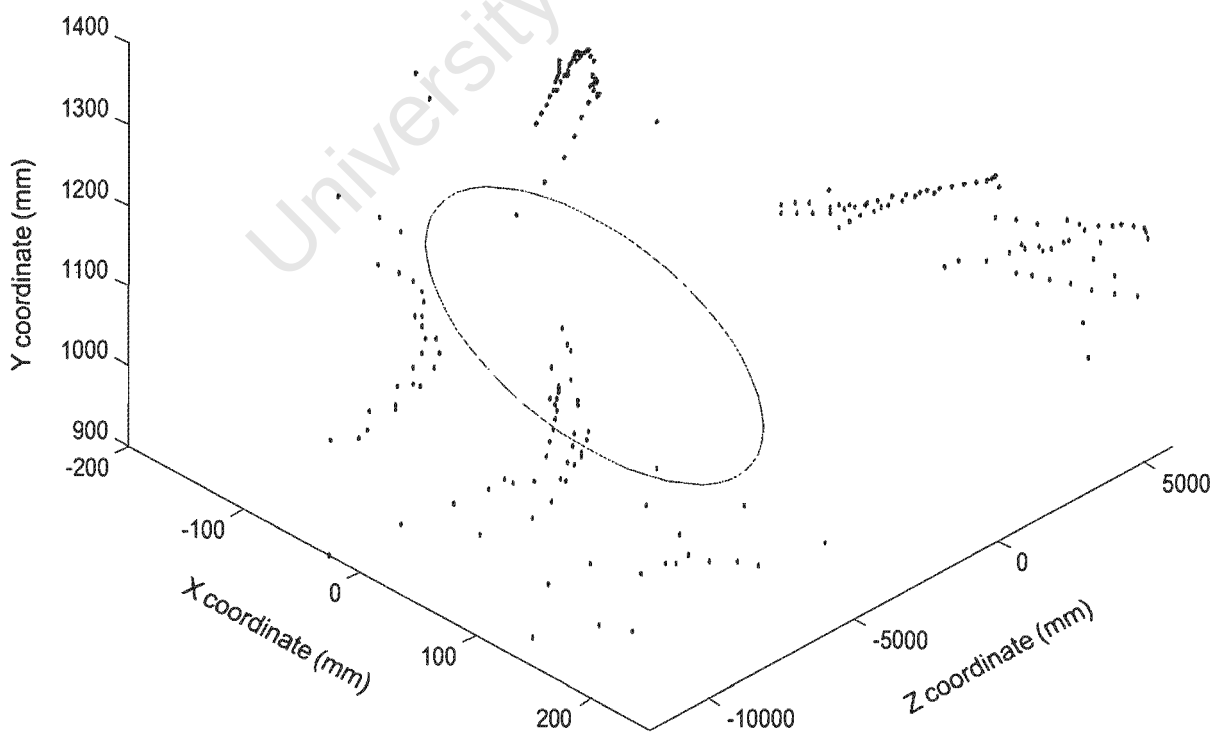


Figure 6.12 Three dimensional presentation of the theoretical and reconstructed trajectory of Target 1 from the measured data.

a procedure was applied to calculated the camera parameters using the 19 control points from the calibration frame as already described in Chapter 5. Similar to the results from Test 3 and Test 4 in Chapter 5, the reconstruction from the measured data of the targets was not accurate enough according to the calculated 3D coordinate errors, and only fragments of the circular trajectory were reconstructed properly. Majority of the points were chaotically distributed in the 3Dspace (Figure 6.12).

	Multiple points tracked in space and reconstructed with DLT, using their generated pixel values, after calibration with 19 points from Test 5 (3m)			Multiple points tracked in space and reconstructed with Bundle Adjustment Method, using their generated pixel values, after calibration with 19 points from Test 5 (3m)		
	diff. X (mm)	diff. Y (mm)	diff. Z (mm)	diff. X (mm)	diff. Y (mm)	diff. Z (mm)
mean	3.00E-04	2.32E-02	1.51E-01	2.85E-13	-1.31E-12	-1.70E-12
stand.dev.	2.71E-02	6.87E-02	1.46E-01	6.10E-13	4.91E-12	1.44E-11
abs. mean	2.38E-02	6.32E-02	1.94E-01	5.00E-13	4.16E-12	1.21E-11
abs. stand.dev	1.31E-02	3.56E-02	8.09E-02	4.49E-13	2.91E-12	7.96E-12
abs. max. error	4.72E-02	1.47E-01	3.96E-01	2.36E-12	1.17E-11	2.91E-11

Table 6.1 Differences in the X,Y and Z directions after coordinate reconstruction with the DLT and Bundle Adjustment algorithms was completed for the two moving targets from Figure 6.1. Their pixel values were generated using the Perspective Projection Theory and sensor arrangement from Figure 5.6 using camera parameters given in Table 5.8.

New pixel values were generated from the so-called "ideal camera system," using Perspective Projection Theory, and reconstruction was repeated with this data, once again using the DLT and Bundle Adjustment algorithms from which the results were summarized in Table 6.1. Following the examples of Test 5 and Test 6 (Chapter 5), it can be pointed out that the Bundle Adjustment method achieved much higher accuracy than the DLT method and again the greatest error is observed in the Z direction in both examples.

The results from the reconstruction and tracking of Target 1 and 2 with the DLT method are shown in three dimensions in Figure 6.13. The initial position of stationary targets before the motion starts is recognisable along the horizontal axis. After the start the rotation proceeds in a counter clockwise direction and 3D tracking is completed for the first turn of 360°, where the trajectories of the targets are plotted in two different colours. The theoretical trajectory is also plotted on the picture to give an idea of the error in 3D, after the calculations were done. The scale along the Z axis is chosen to show the difference between the original and the calculated data, which is less than 1 mm. The errors in the other directions are an order of magnitude smaller.

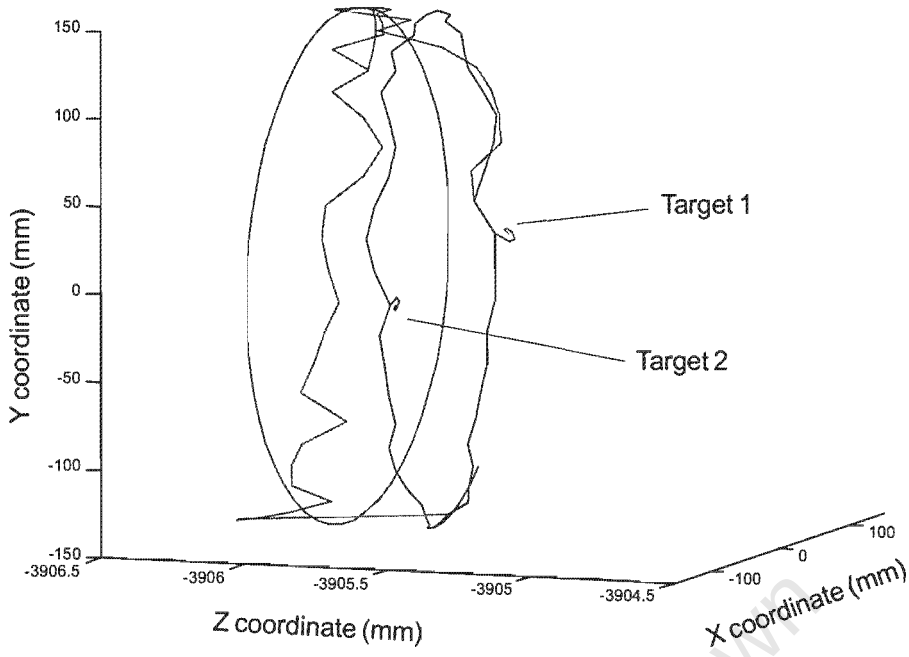


Figure 6.13 Three dimensional presentation of the theoretical and reconstructed trajectory of Target 1 and Target 2 showing the initial stationary positions of both targets. Note the scale in the Z direction is different to emphasize the difference between the curves.

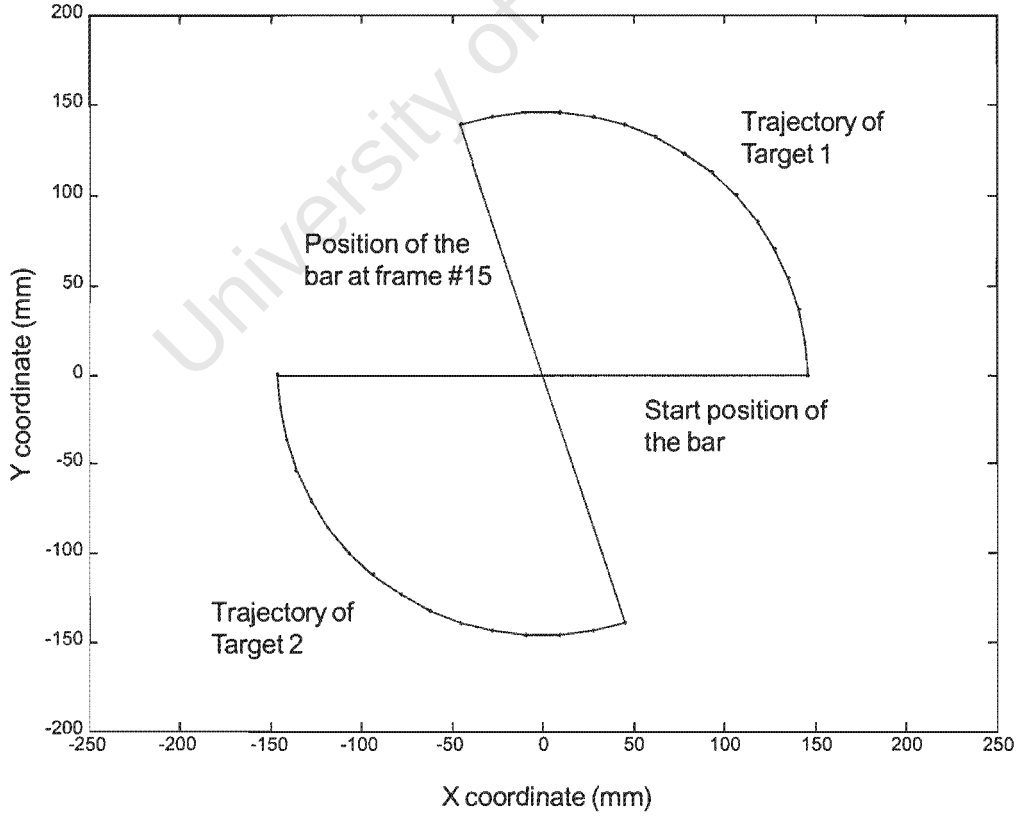


Figure 6.14 Partial reconstruction using the generated data and completed with the DLT method showing the initial stage of the tracking process (dots) of the two active targets over the theoretical trajectory (solid line).

Figure 6.14 shows from the XY prospective the initial stationary position and sections of the trajectories after the motion had started for both targets. From this reconstruction the distance between the targets was calculated for each pair of reconstructed points and compared to the original distance between the targets on the bar. These results were satisfactory and varied in the same range with those obtained from the individual coordinate reconstruction, showing an average error of -0.024mm. The angular position was also examined and the average error was 0.0379°.

	Reconstructed with DLT, using the optimised pixel values, according to p_o only			Reconstruction with Bundle Adjustment Method, using the optimised pixel values according to the all camera parameters		
	diff. X (mm)	diff. Y (mm)	diff. Z (mm)	diff. X (mm)	diff. Y (mm)	diff. Z (mm)
mean	-4.81E-02	1.11E-01	8.02E+00	1.50E-13	-1.42E-12	-2.06E-12
stand.dev.	1.29E-02	3.29E-02	1.13E-01	5.60E-13	4.32E-12	1.25E-11
abs. mean	4.81E-02	1.12E-01	8.02E+00	4.30E-13	3.72E-12	1.03E-11
abs. stand.dev	1.29E-02	3.29E-02	1.13E-01	3.80E-13	2.62E-12	7.38E-12
abs. max. error	6.77E-02	1.81E-01	8.25E+00	1.28E-12	1.02E-11	2.91E-11

Table 6.2 Magnitude of the error in the X,Y and Z directions after coordinate reconstruction with DLT algorithm and Bundle Adjustment Method was completed for the two moving targets. Their pixel values were generated using the Perspective Projection Theory and the optimised sensor parameters.

An optimisation of the camera parameters was suggested at an earlier stage and it was found that this strategy could improve the accuracy considerably. This was done using the same initial conditions given in the last section of Chapter 5. The process was performed step by step and pixel data was generated after each step, using the optimised camera parameters. The termination of the optimisation was achieved after the error, equivalent to the square difference between the generated and measured data, had been set to a minimum. This error for some of the points reached up to a hundred pixels. Nevertheless, the results from the reconstruction with optimised data, using the Bundle Method (Table 6.2), are comparable to the results obtained from the generated data in Table 6.1. When DLT was used for reconstruction of the optimised data the results were found to be much less accurate.

Another optimisation was attempted with only one parameter (p_o- distance to the principal point) included in the process. These data were again reconstructed with the DLT and Bundle Adjustment Methods and the accuracy achieved was in the same range for both methods (Table 6.2). Although the greatest error appeared in the Z direction, cforresponding to the depth of the object space, which could have been due to improper focusing during the optimisation process, the reconstruction and the tracking performed well in the X and Y directions, showing good correlation with the original XY data as seen in Figure 6.15.

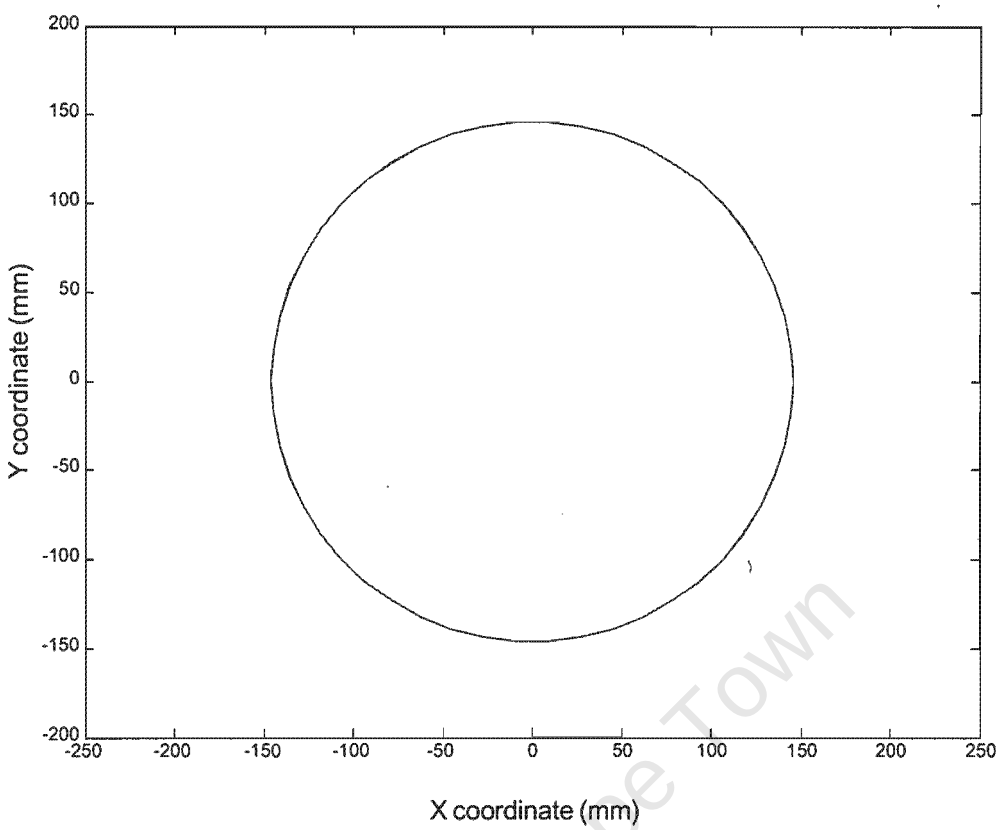


Figure 6.15 Position in the X-Y plane of the circular trajectory of Target 2 after reconstruction with DLT using optimised pixel data by changing one camera parameter p_o.

As a result of the optimisation the relationship between the generated and the optimised pixel data improved to form almost perfect straight lines (Figure 6.16). It was also observed that the number of the calibration parameters involved in the process may cause a variation of the points distribution. As it was established in Chapter 5, when solving the optimisation problem a local rather than a global minimum of the camera parameter function was found. This leads to the conclussion that using only one parameter, for example, the focal distance p_o, would generate better results for the pixel values, compared to the results obtained with all parameters used in the optimisation. This linear relationship is an evidence to this hypothesis, although the range of the generated and the optimised data does not agree with the range of the measured data. The physical size of the cameras was 640 pixels in the horizontal direction and the numbers in Figure 6.16 show that most of the values fall outside this range. Among the reasons for obtaining such an offset, hand positioning, improper estimation of the focal distance and lens distortion, were mentioned in section 5.5. In this particular experiment with moving targets it could be added that the variation of the angular velocity of the motor at every start and stop phase, as well as instability of the sampling rate of the digitiser, would benefit to the data inconsistancy.

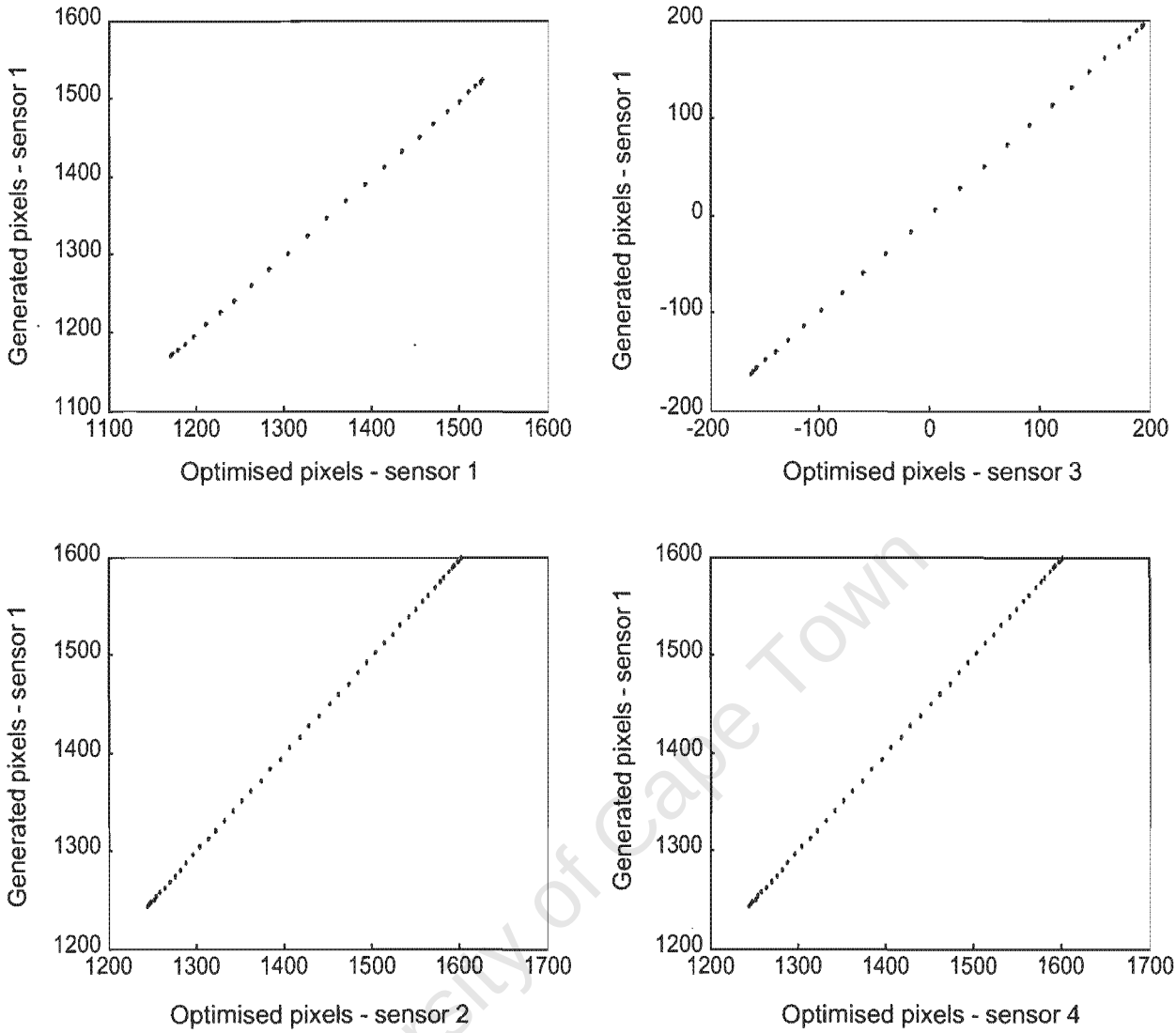


Figure 6.16 Optimised vs generated pixel values for Target 2 using one camera parameter p_o in the optimisation process.

The process of 3D coordinate reconstruction considerably improves the enire target tracking as it helps for a unique target identification and the estimation of the kinematic parameters. Using the results from Table 6.1 when the DLT method was used to calculate the target coordinates and the reconstructed trajectories shown in Figure 6.13, the linear velocity and acceleration along the coordinate axes are obtained. Figure 6.17 shows these two parameters for Target 1, taken along the X axis

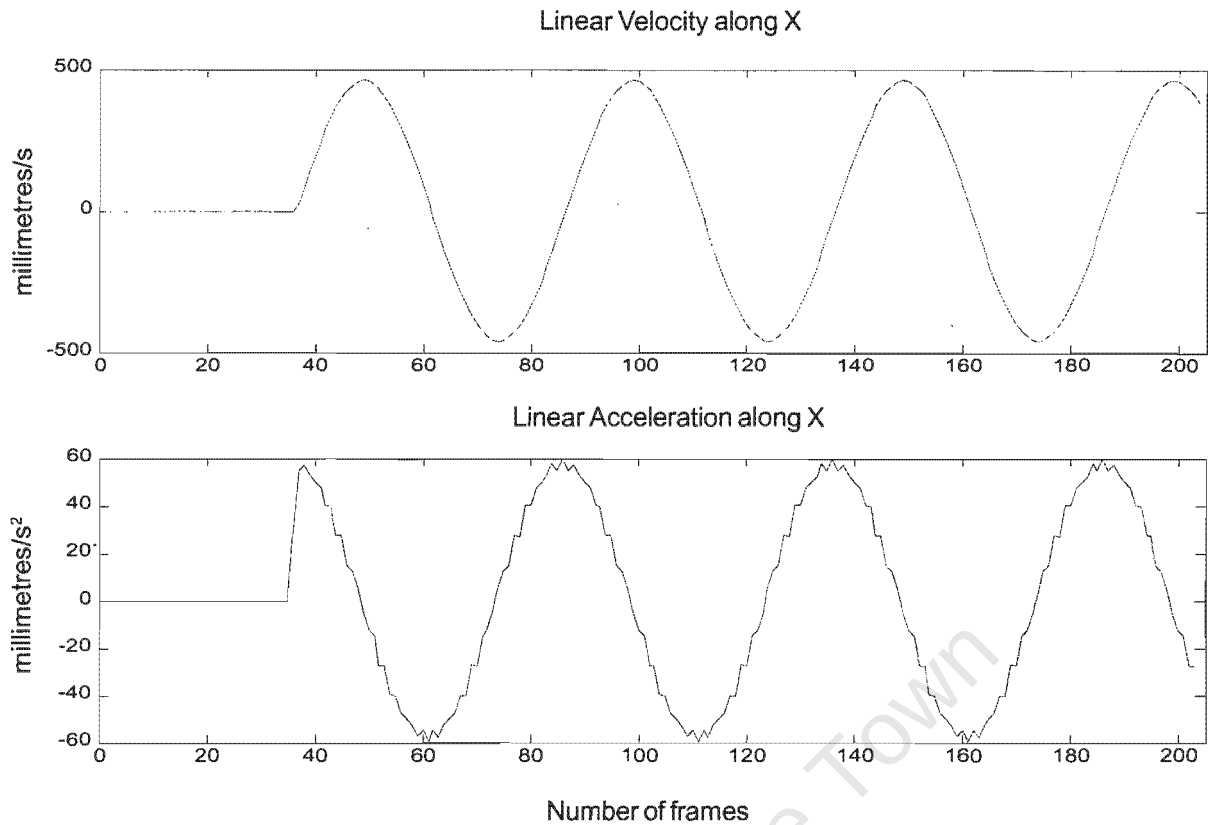


Figure 6.17 Linear velocity and acceleration along the X axis for Target 1. The spikes observed in Figure 6.6 have disapeared as a result of precise position determination after the 3D reconstruction. The acceleration still contains slight variations from small position errors introduced by the tolerance of the calibration parameters of the DLT method.

Completing the tracking and identification in 3D could solve many of the problems that could not be solved in the 1D sensor space due to great ambiguity and uncertainty during target association and the definition of tracks. For this purpose the number of parameters in the dynamic model from [6.1] will increase to 6 if their description in the equations is limited to positions and velocities. Some tasks may require accelerations to be included in the system as well in order to track targets capable of executing manoeuvres (Dufour and Mariton, 1990). The number of parameters in the state vector will then increase to at least 8 in such a system if two accelerations are added to obtain accurate results and the state vector and the transition matrix will be:

$$\mathbf{X}_t = \begin{bmatrix} X_t \\ Y_t \\ Z_t \\ \dot{X}_t \\ \dot{Y}_t \\ \dot{Z}_t \\ \ddot{X}_t \\ \ddot{Y}_t \end{bmatrix} \qquad \mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & \Delta T & 0 & 0 & \tfrac{1}{2}\Delta T^2 & 0 \\ 0 & 1 & 0 & 0 & \Delta T & 0 & 0 & \tfrac{1}{2}\Delta T^2 \\ 0 & 0 & 1 & 0 & 0 & \Delta T & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \Delta T & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta T \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad [6.11]$$

Following the discussion on the trade-off between the complexity and the effectiveness of such a model, it should be pointed out that adding accelerations and transforming equation [6.2] in 3D, would allow a single target model to track an accelerated target and would remove the need for $\mathbf{q}_t$ required in [6.1], where a four state model is considered. The drawback of such a 3D model is that it makes the overall task much more complicated and would increase the requirements for computational power and the time needed to reach a reasonable solution. In comparison to the four state model, this 3D model would therefore show a poorer performance on constant velocity targets (Kurien, 1990). This last remarks could find evidence in most of the far range application such as the radar and naval applications where accelerations play an important role.

The test of tracking targets in 3D in this chapter, however, is a close range example. Some of these applications, such as the ones in biomechanics motion capture and robot industry, could be characterized as containing high velosities accelerations. In others, such as the medical application for neurosurgery, radiotherapy and medical imaging, the dominating components are the stationary positioning, the unexpected motion detection and low velocity tracking.

Three dimensional reconstruction using DLT and BA Method could help considerably the tracking process. Although, for the successful use of these methods a technique to map the measured space should be developed and eventual optimisation process should be utilised at the calibration stage, they both could be considered efficient for the 1D to 3D transformation.

Linear sensors could be most efficient when multiple active targets are being tracked sequentially by activating them one at a time. The applications with simultaneous passive targets are also feasible, but the level of noise, ambiguity and loss of data should be of primary concern. In this case the filter described in section 6.2.1 could become an important tool to solve the problem of low signal-to-noise ratio, stated in section 4.4.

The probabilistic model could be considered as a very effective tool when passive targets have to be identified by 1D sensor in an environment with significant loss of data. In this direction further fruitful research could be done to determine to what level this model could assist the previously suggested dynamic model. One perspective is to utilise the probabilistic target identification to lower the complexity of the dynamic equations with regard to calculating 3D trajectories, velocities and accelerations.

Chapter Seven

Possible Applications of a Tracking System Using Linear Sensors

The field of tracking targets in three dimensions has grown in the last decades and resulted in creation of a vast number of applications to assist various human activities. Among them, two large subdivisions could be pointed out: non-medical and medical. Initially the applications have focused on direction determination (Reymond and Hidalgo, 1980), with further development in the industrial and machine vision areas. Parallel to this many researchers have used the methods of tracking targets to solve different problems in biomechanics, human motion analysis and in various other medical procedures

7.1 Non-medical examples

7.1.1 Examples in robotic industry and machine vision

Motion detection and tracking have become increasingly recognized for all vision systems which operate in an uncontrolled environment. These systems are used extensively in a variety of industrial applications where robots and other machines carry out automatic motions that are used in the manufacturing process.

In general there are two different approaches to tracking. The first approach is recognition-based tracking, which is a modification of object recognition. This method recognizes the object on successive images and subsequently determines its position. The advantage of this method is that it can be applied to three-dimensional tracking. The main disadvantage is that only recognizable objects can be tracked. The system performance is thus limited by the efficiency of the method, which consists of sophisticated operational tasks and is also defined by the types of objects that are recognizable. The second approach is motion-based tracking, which is based on motion detection and is capable of tracking any moving object (Murray and Basu, 1994).

Video cameras have been used for similar tasks as part of the sensor system of robots since the latter were introduced into industry. Monitoring such areas in three dimensions (3D) is an option included in the functionality of most of the safety and some of the surveillance systems. Many examples in this field could be mentioned where videophotogrammetric methods are at the core of the 3D object position determination. Systems based on linear sensors could be used to register the kinematic and dynamic properties of robots during testing and diagnosing their functionality. Another application of the 3D tracking systems is to use them to digitise shapes and contours of complex objects and machine parts for different Computer

Aided Design (CAD) tasks. By using pointers with attached markers to outline their edges and corners, these objects are presented in 3D and these 3D models could be further used by the CAD software. This technique is also useful to describe complex surfaces of small and large objects by measuring numerous points and then reconstructing the real shape.

Some of the more complex systems for volumetric tracking have the capability of changing their camera parameters during the process of image capture and tracking. Such systems are known under the common name of *Active Camera Systems*, which means that the cameras in the system are moving according to their relation to the environment. Unlike the passive systems which are stationary, Active Camera Systems perform better when doing dynamic tasks and following moving objects of complex shape or objects that can perform unpredictable movements. An example called active binocular head (Shih *et al.*, 1998) describes a dual camera system design with a motorized focus lens (MFL) and adjustable orientation and positioning of the cameras according to four rotation angles (left and right camera convergence angle along the vertical axes of the camera holders, tilt angle along X axis, pan angle along Z axis) and translations along X and Y. If linear sensors are used in such Active Camera System, the amount of outgoing data per sensor will be reduced offering an increase of speed for the successive image captures. The number of sensor parameters for 1D compared to 2D sensors will also be reduced allowing more efficient camera motion control.

The Active Camera Systems can significantly improve tracking capability in cases where the object moves out of sight and adjustable focusing can help the process of recognition at different viewpoints in the object space. Such systems based on linear sensors are feasible in applications where the requirements for object recognition are not very strict. These objects can be described by attaching markers to certain locations outlining their dimensions and position. The camera head could then track them in space according to the marker detection feedback and reconstructing 3D coordinates of the markers. The limitation of the narrow angular view for such a camera is compensated for by its ability to move and follow the object trajectory. However, the complexity of these tasks should be taken into account, which is due to the great number of parameters required to be computed at each step. One potential application of the relevant system based on linear sensors could be in the field of industry where such an application registers the specific motion of processed objects or materials. Another application would be the implementation of a safety system that controls machine parts or robots in motion.

7.1.2 Examples of motion capture

The use of motion capture for computer character animation is relatively new, starting in the late 1970's, and only recently becoming widespread. This process entails capturing the movement of a real object and mapping it onto a computer generated object. Usually, motion

capture is used to create synthetic actors by capturing the movements of real humans. In this case special markers are placed over the joints of the actors. Specialized hardware then takes samples of the position and/or orientation of those markers over time, generating data sets, known as motion curves. This technique has been used in combination with special effects to produce animations in the movies.

The proven advantage of motion capturing over traditional animation techniques such as key-framing and simulation is the capability of real-time visualization and the high quality (true to nature) of the generated animation. Although it has been studied since the 1980's, the present utilization of motion capturing is restricted to the direct mapping of animation parameters. In other words, the movements captured from live subjects are mapped directly onto a virtual character and then the animation is displayed.

During the same period, biomechanics laboratories started using computers to analyze human movement. The techniques and devices used in these studies have found their place in the computer graphics arena. In the 1980's commercial optical tracking systems such as the Op-Eye and Sel-Spot systems were used in computer graphics. In the early 1980's, both the MIT Architecture Machine Group and the New York Institute of Technology Computer Graphics Lab experimented with optical tracking of the human body. Optical trackers typically used small markers attached to the body. These were either flashing LEDs or small reflecting spheres and a series of two or more cameras focused on the performance space. A combination of special hardware and software captured the markers on every camera frame and calculated the three-dimensional position of each marker.

The process to prepare a data set that is useful for character animation includes several steps, of which the first one is the recording of a live subject with markers attached over the body. The next step is signal processing to extract all the kinematic and dynamic parameters for the process in the form of functions of time or frequency variables. The curves in biomechanics have a complex shape, they are not perfectly periodic and quite often noise is introduced. They require user interaction and are not suitable for real time applications. At a further stage, automatic methods are suggested for curve cyclification and cycle detection based on the auto-correlation of 1D signals. As a next step some simple motions, such as the motion of a pendulum, are generated and modelled using modulation on the initial curves. Finally these transformed 1D signals are rendered in 3D to obtain the desired trajectory. In the same way, after processing and simulations, the new motion elements, such as motion of the arm or the head are generated from initially captured data from a live person's motion to be used for the creation of a virtual character.

The aforementioned technology is limited by the speed at which the markers can be observed, which is affected firstly by the number of positions per second that can be captured by occlusion

of the markers by the body, and secondly by the resolution of the cameras, specifically their ability to differentiate between markers located close together. The early systems could only track a dozen markers at a time. The more recent systems can track several dozen markers simultaneously. The occlusion problems can be overcome by the use of more cameras, but aside from this, most current optical systems require manual post-processing to recover the trajectories whenever a marker is not detected. The problem with the resolution involves a trade-off between many variables, including the camera price, the field of view and the space of movement. The greater the need for more resolution, the higher the cost of the camera becomes. The same camera can give a greater motion resolution if it is focused on a smaller field of view, but this then limits the scope of the motion (Sturman, 1997).

7.2 Medical examples

7.2.1 Examples in biomechanics and human movement analysis

In the early years researchers of human movement used photographic and cinematic methods for their recordings, which were later replaced by television and video recordings processed by computers. Since the 1970's optical sensors have become more frequently used tools for capturing the motion from a set of markers attached to the limbs of the studied subject. The device known as Coda 3, mentioned earlier in Chapter 2, was the first to use linear sensors to capture data from passive markers. This technique also included an option for the colour encoding of these reflective markers. Filters mounted in front of each sensor helped to identify them at every scan.

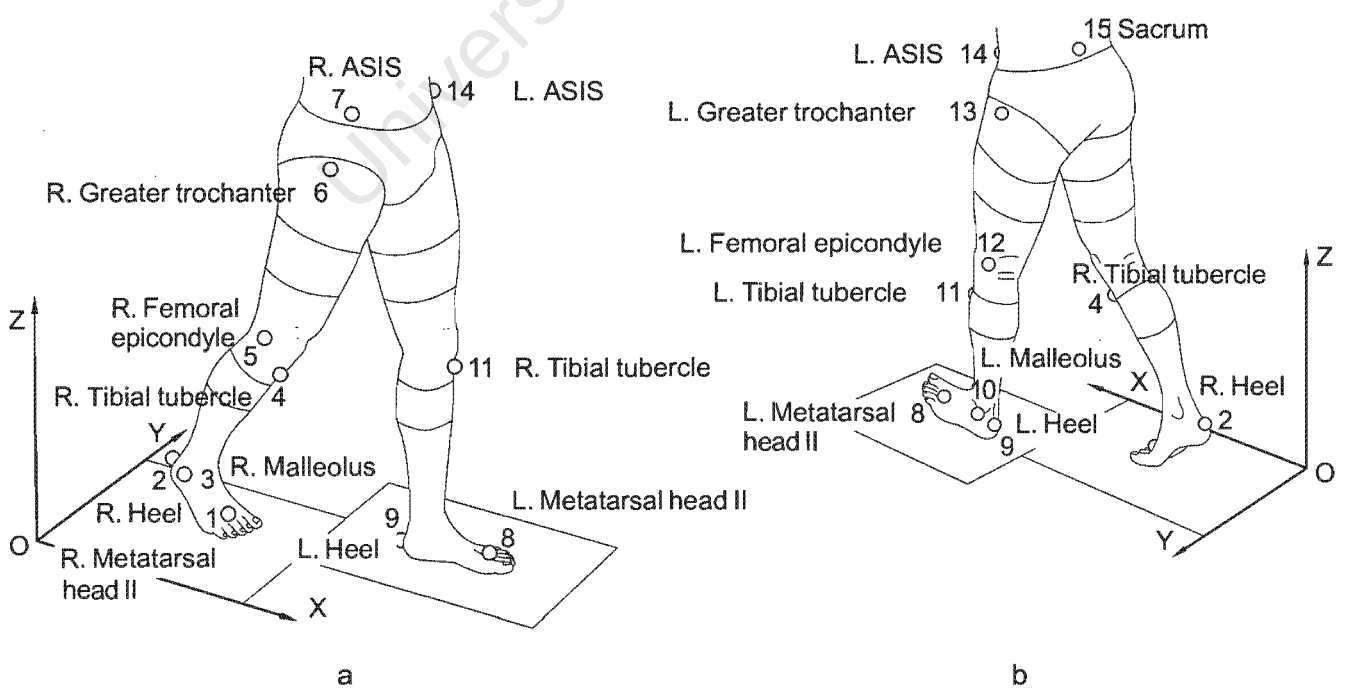


Figure 7.1 The 15 marker system that defines the position of each segment in 3D space, showing the markers placed on anatomic landmarks: (a) anterior view; (b) posterior view (Vaughan *et al.*, 1992).

Other methods preferred to use active LED markers because they provided a higher level of light intensity and were able to be flashed on and off sequentially for easier identification, despite the fact that the subject was required to carry a power supply unit and wires had to be connected to each marker. The camera detectors consisted of multiple (usually two) 2D sensors or pairs of 1D sensors positioned at right angles to each other.

The number of the markers and their arrangement on the subject is also widely discussed in the research literature. One arrangement suggests that the markers be placed directly on the skin over an anatomic landmark and close to the centre of rotation of a joint (Figure 7.1). In this way each limb segment (thigh, calf, foot) is described by three markers positioned on a rigid anatomic structure (femoral bone, tibial bone) (Vaughan *et al.*, 1992; Whittle, 1990).

The other approach to arranging markers is to fix a set of three markers, mounted on a rigid structure, to each limb segment close to its centre of gravity (Figure 7.2). Thus the position and the orientation of the segments relative to each other are determined in 3D space and the kinematic parameters of the joint centres can be derived mathematically. The number of markers for these two arrangements could vary by up to 25, but some authors have proposed a 15 marker system using wands or sticks to attach them to the pelvis, the thighs and the calves (Vaughan *et al.*, 1992).

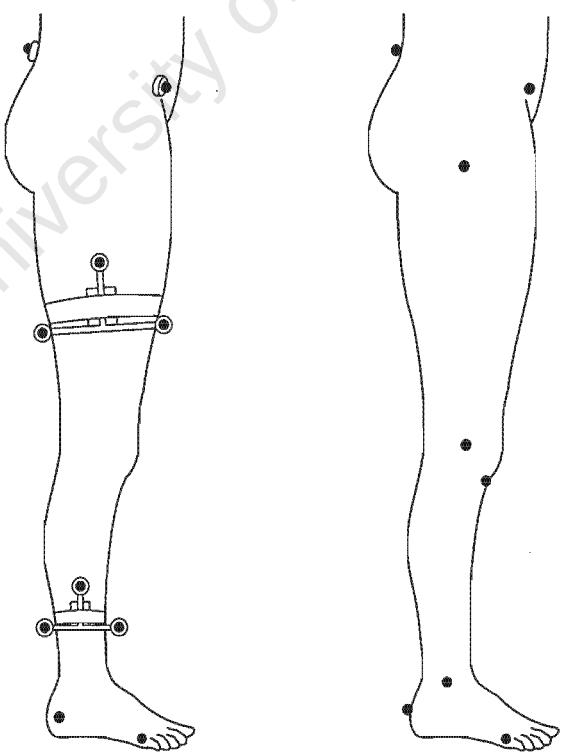


Figure 7.2 Positioning of the markers on the limb segments: (left) using rigid tripods and (right) by placing two markers forming a straight line over a rigid anatomical structure (Whittle, 1990).

The main problem when tracking multiple markers with linear sensors is that the cylindrical lens and linear sensor combination has a limited field of view (a typical lens has a field of view of about 60° at a horizontal angle and about 30° at vertical angle, which is defined by its geometry). To increase the lateral distance of the observed volume of space the subject should be viewed at a further distance from the sensors, which may cause a problem with the light intensity if passive markers are used. In the case of active LEDs a problem may arise from the distribution of the angular intensity resulting in missing markers. However, increasing the angular view of the lens, by changing its geometry, could affect the correct position on the sensor due to lens distortion. If specific measurements for human gait analysis are taken then certain space requirements should be considered, namely that the sensor must have an unobscured view of the walking subject for at least two complete gait cycles, consisting of two steps each. In the event of a problem with the acquisition of continuous data in the space (*e.g.* the cameras can not see the entire cycle), the option is to use several camera detectors with multiple linear sensors which have a specific arrangement for each detector. For the above reasons there would be a considerable trade-off for one to choose between the factors that influence the configuration and the quality of the tracking system performance. This would depend mainly on the nature of the task and the required accuracy tolerance.

7.2.2 Application for radiation therapy

Cancer lesions can be successfully treated by exposure to radiation with the introduction of a radioactive dose to the patient's body. The phenomena that the living tissue cells can survive a certain level of radiation, whereas cancer cells are unable to do so, has made this type of treatment called radiation therapy very effective for certain cancer types at certain locations in the body and at certain stages of development of the disease. During the treatment the patient receives irradiation with X-rays, gamma rays, beta, neutron or proton particles beamed over parts of the body, a process which is carefully calculated by the physicians. The precise positioning of the patient is therefore of prime importance and so is keeping the patient stationary during the procedure in order to achieve the best effect and to reduce the risk of healthy tissue being harmfully exposed to radiation.

The radiation exposure procedure should accurately follow the scheme from the radiation planning in order to minimize planning-to-treatment error. Furthermore the irradiation of exactly the same region throughout every session should be repeated during the radiotherapy course in order to minimize the treatment-to-treatment error. Some of the methods used to immobilize patients for radiation therapy include positioning using stereotactic frames for those cases where the head is the target of the radiation beam. For other cases where the body is the main target, mechanical fixators are used. In most cases patient immobilization techniques are used in combination with control methods. Photogrammetric methods are preferred for this purpose as they are more accurate than the laser control applied to skin landmarks drawn on the body and less expensive compared to the X-ray control of the positioning system.

In cases where brain tumours are treated with radiation, the patient often wears a thermoplastic mask tightly fitted on the head on which markers are attached to determine the position of the mask-head, thereby defining the tumour area within the radiation beam space. Several video or CCD cameras observe the markers and provide data to a 3D positioning system which calculates the space coordinates of the mask (Rogus *et al.*, 1999). Similar systems are also used for the correction of patient movements where retro-reflective sticking markers are attached directly to the body, to masks or marker holders held by bite blocks. The number of markers can vary from several to more than 20, where the minimal number required to define a local coordinate system is three.

Figure 7.3 A thermoplastic mask for proton therapy with 22 attached retroreflective markers to determine the patient head position in 3D.

Figure 7.4 Images from a thermoplastic face mask used for radiation therapy to control patient immobilisation during the exposure procedure. The three images from left, central and right CCD cameras show the markers that could be seen from each camera.

Figure 7.3 shows one example of a face mask for radiation therapy with 22 markers attached and visualized by 3 CCD cameras on Figure 7.4. The large number of markers allows even slight movements to be detected by the system and three cameras are used to ensure that each marker will be seen from at least two cameras to obtain the data needed for a 3D coordinate reconstruction. This system is connected to the radiation system which should

stop the radiation beam whenever a significant variation of the patient position is detected. The necessary corrections must then be sent to the patient positioning system and only after this has been done can the treatment process continue.

These observations have shown that large errors of several millimetres may occur between the planning and treatment stage of the radiotherapy process. The error would be within more or less the same limits for the treatment-to-treatment stages. There are also variations of these limits when different anatomic sites are irradiated for treatment, for example brain tumours, breast and lung cancer or other tumours in the pelvic area.

Using a tracking system to control the patient movement will reduce these errors considerably. Some phantom studies have pointed out that it could come down to within 0.3 mm with good repeatability. In actual patient cases there is a systematic error in patient positioning which reaches up to 4 mm and in patient movement an error during treatment was found to range from 0 to 0.15 mm (Rogus *et al.*, 1999).

7.2.3 Application for surgery and neurosurgery

Tracking systems used to assist neurosurgeons when operating on a patient are better known as Surgical Navigation Systems (SNS). These systems started as early as the 1970's when a procedure was established for localizing internal brain tumours by means of Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) images prior to stereotactic operations. The accurate definition of the lesion in 3D space takes place after the patient has been placed in a stationary position and is followed by the determination of the point of intervention. It is a three-dimensional task that is best performed by computers which significantly improve and facilitate the entire operating procedure. In the beginning this was a static implementation which used photogrammetric methods to calculate the three-dimensional position of the lesion in the local (the imaging device) and in the global (the operating theatre) coordinate systems. Images of the patient with the fiducial markers were taken by video cameras and after digitization the data was stored and manipulated by personal computers (PCs) or computer workstations. Later, with the advances of technology, these localizations were implemented online in such a manner that the operating team could be provided with a continuous picture of the treated area as well as the position of the surgical instrument in the close vicinity. This is usually shown on a computer screen where the tip of the instrument is superimposed over the CT or MRI images taken prior to the operation (*cf.* Figure 7.5).

At present the systems for tracking surgical instruments like pointers, forceps, endoscopes, electric knives, *etc.* have become inseparable part of the facilities in operating theatres. They could be based either on rectangular (two-dimensional) CCD sensors or on linear one-dimensional CCD sensors. They typically use infrared (IR) light to avoid interference from the

other light sources in the operating theatre. The targets attached to the instruments can be either Light Emitting Diodes (LEDs) or passive retro-reflective targets that have to be illuminated by an IR light beam. LEDs are easily identifiable when pulsed sequentially and can be distinguished better from the ambient light. Passive targets, unlike the LEDs, can only be detected simultaneously, but they have the advantage of functioning without wire and can pass through standard autoclave sterilization. Figure 7.5 shows an example of a localization procedure with a navigating system based on three linear sensors and IR LED markers which provides a resolution of 1/5 000 and an accuracy of 0.5 mm in all directions within a volume of 1 m^3.

Figure 7.5 A Surgical Navigating System (SNS) including 3D localizer Pixsys FlashPoint from Image Guided Technology (Denver, USA) based on 3 linear sensors and IR LEDs attached to the localizer.

The concept of converting the 3D coordinates of the targets in the local and the global coordinate system is illustrated in Figure 7.6. First the patient undergoes a CT or MRI scan to find the anatomical location of the tumour, which will be operated on. During this scan several fiducial markers, detectable from the scanning system, are attached to the patient, so that the coordinate system xyz is defined and referred to as the local or sometimes the intrinsic coordinate system. The scanning device itself arranges the scans in its own coordinate system $x_s y_s z_s$, which provides the x_s and y_s values from the image. The z_s value can be found from the thickness of the image slice and the scanning step.

The next step is to move the patient into the operating room and to secure the patient in an immobile position. A 3D localizer, which closely observes the patient, is also mounted in the area and calibrated, so that the global (extrinsic) coordinate system XYZ is now defined. This coordinate system may have its origin in the middle of one of the sensors as is shown in Figure 7.6. During the operating procedure a pointer or another instrument is introduced into the measured space and its tip is directed to each of the markers M_1, M_2, M_3. The IR LED markers on the instrument are measured in 3D and from their positions the position of the pointer's tip is calculated.

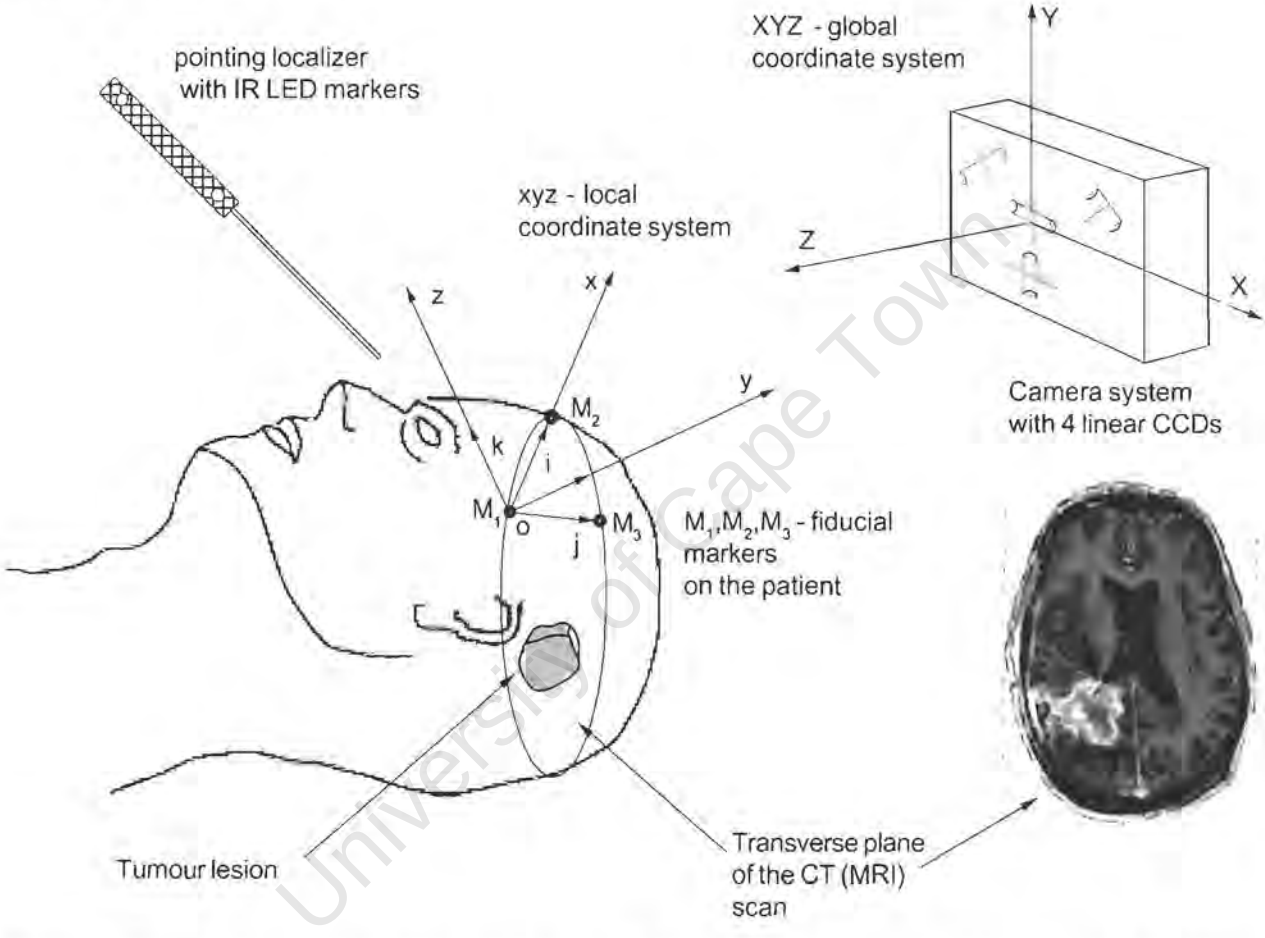


Figure 7.6 Tumour localization procedure in 3D space using a space localizer with IR LED's and a detecting device with 4 linear sensors.

When directed to the markers M_1, M_2, M_3, the pointer is measured from the camera and the positions of M_1, M_2, M_3 are also measured in 3D. From the two unit vectors i and j the coordinate system xyz is determined in the following way:

$$\overrightarrow{M_1M_2} = \vec{i} \qquad \overrightarrow{M_2M_3} = \vec{j}$$

then

$$\vec{x} = \vec{i} \qquad \vec{z} = \vec{k} = \vec{i} \times \vec{j} \qquad \vec{y} = \vec{i} \times \vec{k}$$

This system is further described in terms of XYZ and its translation and rotation values are found. Thus every object with given xyz coordinates can be presented in XYZ coordinates as well. All CT or MRI images lying within the xyz system can be used as reference images from within XYZ. After the markers M_1, M_2, M_3, are found in the XYZ coordinate system, their transformation parameters from xyz into XYZ are calculated and the same is done with the centre of the mass or other reference point in the tumour. The coordinate transformation is also applied to all the images from the pre-operative study, so they can be presented in the XYZ coordinate space. The next procedure entails the definition of the intervention entry point, after which an incision is made and the lesion can be accessed. During the procedure of approaching the lesion, multiple control points are measured for positioning the instrument and establishing the current position of the tumour. This has to be done because the position will change once the skull has been opened due to changes in the intracavital pressure. At each measurement point, after calculating the position of the instrument tip in 3D, the system will call the nearest CT (MRI) scan to show the instrument's relative position superimposed onto the image and to compare this to the original position of the tumour before the operation. Instrumentation based on linear sensors, for tracking targets in a surgical space vies strongly with instrumentation based on rectangular sensors. With proper compensation of the optical disadvantages, and by improving the signal to noise ratio of the linear CCD sensors, the instrumentation based on linear sensors can offer better spatial resolution and greater sampling rate. It can be used very successfully in surgery for tracking instruments and biopsy needles and for rendering them visible over hidden and less accessible anatomic structures. This, together with the methods of medical imaging (*e.g.* MRI, CT and US), has created a new field called Image Guided Surgery (Maciunas, 1993). An example of the use of such an optical instrument is described by Bucholz and Smith (1993) during stereotactic surgery. In this procedure they used three linear sensor cameras mounted on a 1m bar and tracked a forceps, marked with two active targets which were light emitting diodes. The tip of the instrument was shown on a series of CT scans and was navigated throughout the procedure of a tumour resection. The technique allowed surgeons to make smaller incisions by mapping the tumour onto the skull and left a minimum of skull openings needed to expose the lesion.

A research team at the University of Uppsala has used an integrated neuro-navigational system, ViewScope (Figure 7.7) from Elekta and Leica, in the field of spinal cord and brain surgery. This is a surgical microscope which is combined with a tracking localizer. As the tracking device, OPTOTRAK 3000 was incorporated into the system. This facility ensures a real-time 2D and 3D image update and a quick 3D image reconstruction with optional orthogonal views in the main planes (coronal, sagittal and transverse) as well as oblique slices and volumetric visualization.

Figure 7.7 ViewScope neuro-navigational system (Elekta, Stockholm, Sweden) consising of a surgical microscope from Leica (left) optical tracking localizer OPTOTRAK 3000 (right) and the image processing and visualization unit (middle).

There is an extensive discussion on the current state of the use of tracking systems for surgical navigation (Barnett *et al.*, 1998; Maciunas, 1998) pointing out that such systems, based on 1D and 2D CCD sensors, are playing a substantial role in this field, and are taking precedence over the other types such as magnetic and ultrasonic systems. The magnetic systems have the advantage of being able to operate without a direct line of view of the measured space, but they are sensitive to magnetic fields and metal objects which cause interferences. The ultrasonic systems, similar to the optical sensors, require a direct line of view and can be influenced by temperature and humidity changes. Most commercial sonic systems for surgical navigation have already been converted to optical based technologies. Passive targets for digitiser pointing probes are also competitively used in practice, despite their disadvantages related to identification and sterilization. There are some reports on the use of ultraviolet (UV) light together with fluorescent targets, which absorb the light from the near UV range and re-radiate it back at visible wavelengths. These targets can also be coded to re-radiate at different colours for easier identification. However, experience has mainly shown that illumination in the prescribed UV range causes extreme fatigue to the human eye and might be potentially dangerous (Barnett *et al.*, 1998). Among the most used target tracking systems are OPTOTRAK 3D bar camera (Northern Digital, Waterloo, Canada) and Flashpoint 5000 3D localizer (Image Guided Technology, Denver, USA). Similar to the design of the Flashpoint 5000 system, another company Radionics (Burlington, USA) has presented its Optical Tracking System (Figure 7.8). It uses three linear sensors mounted on a straight line, one vertical in the middle and two horizontal on the sides pointing inwards to the field of view.

Figure 7.8 The Optical Tracking System from Radionics (Burlington, USA) showing the arrangement of the linear sensors, the pointing probe with IR LEDs and the 3D localization on the computer screen.

A variety of examples could be pointed out where light emitting optical SNS are used in different surgical and neuro-surgical applications. Stereotactic frame methods are more often replaced by frameless procedures in brain surgery and craniotomy. Recent reports confirm the growing number of cases which involve the use of such SNS in operations of vascular malformations (Zamoranno *et al.*, 1998); intracranial meningiomas (Barnett and Kaakaji, 1998); spinal surgery (Kalfas, 1998); and in cases of brain biopsies and related procedures the SNS allows for tracking trajectories by means of a LED digitizer (Barnett and Miller, 1998). Another example is the intra-operative open MRI device, where the optical localizing system Pixsys (Image Guided Technology) is integrated directly into the surgical space between the two super-conducting magnet tori. It consists of three linear CCD cameras which can provide a 3D mapping space in real-time mode and can define volumes and anatomic structures (Alexander *et al.*, 1998).

More recently, optical digitizers that are able to track targets in 3D space have become an inseparable part of the clinical equipment used in the field of Computer Integrated Surgery and Therapy (Lavallee *et al.*, 2000). A list of some of the major benefits includes: less invasive surgery; better accuracy and improved reliability and safety; shorter periods of intervention; facilitating the procedure for the surgeon; creating the possibility of tele-surgery when ionizing radiation is needed; and reduced risk to medical staff.

Chapter Eight

Conclusions

Linear sensors can provide sufficient linear resolution, especially with the latest technological design, which consists of several thousand pixels. A resolution reaching 1/150 000 can be attained by using the most recent technology and applying centroid algorithm to achieve sub-pixel resolution.

A sensor consisting of 5000 pixels with a clock rate of 40 MHz, can provide sampling rates of more than 8000 lines per second. A rectangular 2D sensor, with resolution 1K x 1K, has a maximum rate of 500 frames per second. Considered from the data processing point of view, 2D sensors, compared to linear sensors, need more storage space and faster links to transfer over one million pixels per sample. The sampling rate is also important when multiple targets have to be identified, especially such targets that have crossing tracks.

A cylindrical lens, in combination with a linear sensor, may cause more distortion in comparison to that which a spherical lens can produce on a rectangular sensor. This problem has to be solved individually for every linear sensor-cylindrical lens element, also taking into account all contributing factors such as the sensor length, the lens geometry and the character of the distortion curve. The cylindrical lens distortion curve can be approximated by a polynomial, similar to the model described for a spherical lens. Corrections can be worked out in the form of a table of coefficients applied to the distance from the optical axis in a 1D case, or as mapped spatial corrections for multiple sensors in a 3D case.

A loss of optical power caused by the cylindrical lens, directly affects the image intensity. The attenuation coefficient, when comparing the level of intensity from a spherical lens to that from a cylindrical lens, can reach an order of one hundred. The signal-to-noise ratio, which is another quantification factor, is more than five times higher for a spherical lens compared to that from a cylindrical lens. This effect should be carefully taken into account when passive targets are imaged in the presence of a strong ambient light reflection. To improve the signal-to-noise ratio when using passive targets, it might be necessary to reduce the distance to the sensors to achieve stronger reflection from the targets. In this case the object space to be measured will become smaller and more sensors at different positions might be needed.

Active targets are more convenient for use in terms of their higher light intensity and their capability of being controlled sequentially for easier identification. To enable this latter function, the active targets need additional sophisticated electronic control circuits and have to either be connected with wires or controlled remotely. For some applications such as surgery, they

may need to be sterilized under special conditions. Targets such as light emitting diodes have a limited angle of irradiance.

Passive targets, on the other hand, suffer from poor light reflection and they have to be identified simultaneously. However, if necessary, they can be prepared more easily for different experiments in terms of attaching, positioning and sterilizing.

At the signal processing stage, filtering and centroid calculation significantly improve linear resolution, but some filters, such as most convolutional types, may require additional phase shift compensation of the calculated position. Infinite Impulse Response (IIR) filters have a nonlinear phase response which could complicate this process.

Multiple targets can be detected separately on the linear sensor, however, targets located close together may also be observed. These targets can be separated at a certain stage, but in a single projection they might still be identifiable as one target, although they are separated from one another in 3D space.

The reconstruction of 3D coordinates of targets can be achieved using either Direct Linear Transformation (DLT) or Bundle Adjustment (BA) Method based on the Least Squares approach, applied to an overdetermined system of equations for $n \geq 3$ sensors. For both methods an appropriate calibration calculating the camera parameters should be completed prior to using them for coordinate reconstruction. DLT is more sensitive to changes in the sensor parameters and sometimes it is difficult to obtain convergence of the principal point. It is also less stable than the BA Method at the reconstruction stage. Both methods work well with data generated using the Perspective Projection Theory for imaging targets on the sensor. The error from the reconstruction is observed to have the same magnitude in the directions parallel to the plane of the sensors and greater magnitude in the direction corresponding to the depth of the measured space. Increasing the number of sensors used in the reconstruction process minimizes this error.

Optimization of the camera parameters leads to significant improvement of the accuracy of the 3D reconstruction. This process however is slow and requires sufficient computational power.

It is suggested that corrections for the target position on the sensor can be worked out by creating a map of the measured space, using large number of points, but in this instance a problem might be caused by the ambiguity, from the projection point of view, when targets at different positions in a space produce the same sensor values.

It is possible to track a single target in a space by using a dynamic model to evaluate its velocity and acceleration. This process is usually divided into two stages namely single sensor (1D) tracking and 3D tracking. The latter includes coordinate (position) and trajectory reconstruction. Multiple targets tracked simultaneously should use the dynamic model to evaluate velocity and acceleration parameters. This should be done in combination with a probabilistic model to evaluate the state of the targets, which determines the targets' association with the existing tracks.

A system based on linear sensors could be a successful solution to a variety of spatial measurements for medicine, sport, the robotic industry and computer vision.

Appendix A

Tests to Determine the Distortion of a Cylindrical Lens

Data from the test for a lens distortion shown on Figure 3.15 where the target and the linear sensor-cylindrical lens system were simulated and linear displacement s were measured along the sensor.

distance (mm)	measured position on the sensor (pixels)	predicted position (pixels)	distance (mm)	measured position on the sensor (pixels)	predicted position (pixels)
0.0	1025.0	1025.0	87.5	880.3	866.9
3.5	1018.3	1018.7	91.0	875.5	860.6
7.0	1012.8	1012.4	94.5	870.6	854.3
10.5	1005.4	1006.1	98.0	866.3	848.0
14.0	1000.2	999.7	101.5	861.2	841.7
17.5	993.7	993.4	105.0	856.5	835.3
21.0	986.6	987.1	108.5	852.1	829.0
24.5	980.3	980.8	112.0	847.5	822.7
28.0	974.1	974.4	115.5	843.1	816.4
31.5	967.3	968.1	119.0	838.5	810.0
35.0	961.3	961.8	122.5	834.2	803.7
38.5	955.3	955.5	126.0	830.3	797.4
42.0	949.3	949.2	129.5	825.6	791.1
45.5	943.8	942.8	133.0	821.5	784.7
49.0	938.8	936.5	136.5	817.2	778.4
52.5	932.3	930.2	140.0	812.8	772.1
56.0	926.9	923.9	143.5	808.9	765.8
59.5	921.4	917.5	147.0	804.8	759.5
63.0	916.2	911.2	150.5	800.8	753.1
66.5	910.9	904.9	154.0	796.9	746.8
70.0	905.3	898.6	157.5	793.1	740.5
73.5	900.3	892.2	161.0	789.2	734.2
77.0	895.3	885.9	164.5	785.3	727.8
80.5	890.2	879.6	168.0	781.7	721.5
84.0	885.3	873.3	171.5	777.8	715.2

Table A.1 Horizontal space intervals (column 1) at 3.5 mm of a simulated target, its predicted pixel values (column 2) and its measured pixel positions along the CCD sensor (column 3).

Data from the test where lens distortion was measured according to Figure 3.16 where the one target was positioned at the centre of the lens and the other was moved parallel to the linear sensor.

Left side			Right side		
distance (mm)	measured position on the sensor (pixels)	predicted position (pixels)	distance (mm)	measured position on the sensor (pixels)	predicted position (pixels)
0	1024.00	1024.00	0	1024.00	1024.00
15	912.20	912.40	15	925.20	910.80
20	876.90	875.30	20	885.20	873.10
25	844.00	838.10	25	847.50	835.30
30	811.10	800.90	30	809.90	797.60
35	778.10	763.70	35	769.90	759.90
40	738.10	726.50	40	732.20	722.20
45	707.50	689.30	45	694.60	684.40
50	674.60	652.20	50	656.90	646.70
55	644.00	615.00	55	626.40	609.00
60	611.10	577.80	60	591.10	571.20
65	578.10	540.60	65	555.80	533.50
70	545.20	503.40	70	522.80	495.80
75	512.20	466.20	75	487.50	458.00
80	481.60	429.10	80	454.60	420.30
85	451.10	391.90	85	421.60	382.60
90	420.50	354.70	90	388.70	344.80
95	387.50	317.50	95	351.10	307.10
100	356.90	280.30	100	318.10	269.40
105	327.50	243.10	105	285.20	231.70
110	296.90	205.90	110	249.90	193.90
115	264.00	168.80	115	216.90	156.20
120	228.70	131.60	120	184.00	118.50
125	198.10	94.40	125	155.80	80.70
130	169.90	57.20	130	125.20	0.00
135	141.60	0.00			

Table A.2 Horizontal space intervals (column 1) at 0.5 mm of a measured target, its predicted pixel values (column 2) and its pixel positions along the CCD sensor (column 3).

This data, presented in pixels could be easily converted into millimetres by multiplying the error value by the CCD pixel size, which is 0.014 mm.

Appendix B

Theory of the Magnitude and Phase Response of FIR and IIR Filters

For FIR filter the transfer function could be derived from the impulse response after a Fourier transform is applied:

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n)e^{j\omega n} \qquad [B.1]$$

This complex function could be written in an exponential form:

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\theta(\omega)} \qquad [B.2]$$

where $|H(e^{j\omega})|$ is the magnitude response of the FIR filter and $\theta(\omega)$ is its phase response. For certain constraints on the impulse response, $\theta(\omega)$ is exactly linear (Rabiner and Gold, 1975) and it can be written in the form:

$$\theta(\omega) = -\alpha\omega \qquad -\pi \leq \omega \leq \pi \qquad [B.3]$$

where α is a constant phase delay in samples. Thus equation [B.3] will become:

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{-j\alpha\omega} \qquad [B.4]$$

and for the real and imaginary part it could be written:

$$Re[H(e^{j\omega})] = |H(e^{j\omega})|\cos(\alpha\omega) = \sum_{n=0}^{N-1} h(n)\cos(\omega n) \qquad [B.5]$$

$$Im[H(e^{j\omega})] = -|H(e^{j\omega})|\sin(\alpha\omega) = -\sum_{n=0}^{N-1} h(n)\sin(\omega n) \qquad [B.6]$$

from where the phase response could be obtained in the form:

$$\theta(\omega) = -\alpha\omega = -\arctg\left\{\frac{\sum_{n=0}^{N-1} h(n)\sin(\omega n)}{\sum_{n=0}^{N-1} h(n)\cos(\omega n)}\right\} \qquad [B.7]$$

According to Rabiner and Gold (1975) one solution for α from equation [B.7] can be expressed by

$$\alpha = \frac{N-1}{2} \qquad [B.8]$$

where N is the number of the sample points of the filter given by its order.

In the case of the IIR filter there are three important functions which can be derived from the filter transfer function

$$H(e^{j\omega}) = H(e^{j\omega})e^{j\varphi(\omega)} = H(z)|_{z=e^{j\omega}} \qquad [B.9]$$

These are: the magnitude square response

$$A(\omega)^2 = |H(e^{j\omega})|^2 = |H(z)H(z^{-1})|\,|_{z=e^{j\omega}} \qquad [B.10]$$

described as a change in the magnitude of the input signal with the frequency when passed through the filter and it is important for determining the approximation function *e.g.* Butterworth, Chebyshev, Bessel; the phase response

$$\varphi(\omega) = \text{arctg}\{\frac{\text{Im}[H(z)]}{\text{Re}[H(z)]}\}|_{z=e^{j\omega}} \qquad [B.11]$$

which is a change of the phase angle with the frequency; and the group delay

$$\tau(\omega) = \frac{d\varphi(\omega)}{d\omega} \qquad [B.12]$$

which is the measure of the average delay of the filter as a function of frequency.

One typical characteristic of these filters is that the requirement to have a linear phase response will arrange the poles and the zeros of the transfer function H(z) in such a way as to render the filter unstable. To determine what the exact phase shift will be in samples for a discrete signal, it is recommended that this parameter is linearized for the frequencies in the pass band from 0 to ω_{cutoff}, and then as in the case of a FIR filter to assume that the phase shift in samples is constant and equal to the group delay.

$$\tau(\omega) = \frac{d(-\alpha\omega)}{d\omega} = -\alpha \qquad [B.13]$$

Appendix C

Linear Resolution Tests for a Cylindrical Lens and a Rectangular Sensor

As it was described in Chapter 4, multiple tests to determine the linear resolution of a cylindrical lens-linear sensor system were performed. However, in these experiments rectangular, instead of a linear sensor was used as a detecting element. This rectangular sensor had 512 pixels along the horizontal side, compared to the linear sensor, which had 2048. The objective of these tests was to find out whether 512 pixels would be sufficient to distinguish a target performing very small position changes in a direction parallel to the direction of the sensor.

The tests were performed with an active target (infrared light emitting diode) 2m away from the camera. Different displacement steps were experimented to determine the lower limit at which two separate positions could be identified correctly. The data processing stage included filtering with different types of filters, followed by calculation of the centroid value of the detected signal as described in section 4.2.2. These tests also showed the effectiveness of the filters using various parameters for different step intervals following a decreasing order.

The results are presented in Tables C.1 through C.4, where the displacement intervals were 5 mm, 1 mm, 0.5 mm, 0.1 mm respectively. All tables include a section presenting the raw data and different sections presenting the use of a median, Infinite Impulse Response (IIR, Butterworth) and Finite Impulse Response (FIR) filters. The order and the cutoff frequency of the filters were tested with different values to find out which set of these parameters could identify the target positions properly. Each section includes pixel data representing the position of the maximum pixel value on the sensor and centroid data representing the centre of gravity of those pixels corresponding to the target position. The pixel values from some of the sections containing filtered data have increased their values by certain number, because of the phase shift that the particular type of filter introduces during the processing stage.

Raw data			Filtered data									
5mm			median		Butterworth n=4				Butterworth n=11			
			n = 3		fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
mm	pixels	Centroid	pixels	Centroid	pixels	Centroid	pixels	Centroid	pixels	Centroid	pixels	Centroid
0	310	309.42	309	309.40	309	310.43	311	311.42	312	312.33	315	314.53
5	317	316.32	316	316.28	316	317.43	318	318.38	318	319.29	322	321.50
10	322	323.43	322	323.09	324	324.39	325	325.32	325	326.23	328	328.12
15	329	329.98	330	330.03	330	331.36	332	332.29	332	333.21	335	335.06
20	337	336.99	336	337.03	337	337.99	339	338.95	339	339.54	342	342.32
25	345	343.50	342	344.16	344	344.93	346	345.92	346	346.81	349	349.04
30	352	350.83	349	351.21	351	351.91	352	352.89	353	353.80	356	356.01
35	356	357.82	357	357.99	358	358.89	359	359.55	360	360.44	363	362.95
40	363	364.84	364	364.90	365	365.56	366	366.52	367	367.42	370	369.93
45	371	371.61	372	371.65	372	372.55	373	373.53	374	374.41	377	376.95
50	378	380.17	379	379.96	380	380.86	381	381.84	382	382.42	385	384.95
55	386	385.31	383	385.62	385	386.42	387	387.39	387	388.32	390	390.51
60	391	392.41	392	392.78	393	393.44	394	394.37	394	395.27	397	397.48
65	398	399.38	399	399.19	399	400.40	401	401.01	401	402.23	404	404.16
70	405	405.96	406	406.09	406	407.29	408	407.92	408	408.77	411	411.08
75	414	412.81	411	412.71	413	414.31	415	414.94	415	416.19	418	418.09
80	421	419.81	418	419.73	420	420.90	421	421.90	422	422.76	425	425.05
85	426	426.65	426	426.58	427	428.25	428	428.89	429	429.77	432	432.04
90	433	433.95	432	433.57	434	434.80	435	435.80	436	436.69	439	438.93
95	440	440.88	441	440.61	441	441.80	442	442.78	443	443.35	446	445.89
100	447	449.08	448	448.81	449	449.43	450	450.49	451	451.35	454	453.97
105	455	455.85	455	455.51	456	456.36	457	457.45	458	457.98	461	460.92
110	461	463.08	462	462.56	463	463.75	464	464.42	465	464.97	468	467.58
115	469	469.43	469	469.39	470	470.35	471	471.06	472	471.57	475	474.53
120	476	476.38	476	476.36	477	477.02	478	478.06	479	478.59	482	481.53
125	483	483.41	482	483.36	484	483.97	485	485.03	486	485.56	489	488.50
130	490	490.33	489	490.33	490	491.24	492	491.92	492	492.47	495	495.10
135	496	497.21	495	496.88	497	497.85	499	498.89	499	499.42	502	502.41
140	502	504.19	502	504.19	504	504.79	505	505.82	506	506.35	509	508.99
145	510	510.35	509	510.09	511	511.00	512	511.57	512	512.00	512	401.80

Table C.1 Measurements from a linear resolution test at 5 mm steps using median (n - aperture in samples), Butterworth and FIR filter (n - filter order; fc - cutoff frequency). Data are presented with their pixel values and centroid values.

Raw data			Filtered data							
5mm			FIR n=4				FIR n=11			
			fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
mm	pixels	Centroid	pixels	Centroid	pixels	Centroid	pixels	Centroid	pixels	Centroid
0	310	309.42	312	311.44	312	311.45	315	314.96	315	314.98
5	317	316.32	319	318.38	319	318.40	323	321.62	322	321.62
10	322	323.43	324	325.07	324	325.38	328	328.59	328	328.57
15	329	329.98	331	332.31	331	332.33	334	335.52	335	335.52
20	337	336.99	339	338.99	339	338.99	343	342.48	342	342.49
25	345	343.50	347	346.22	347	346.24	348	349.42	349	349.46
30	352	350.83	354	353.23	354	352.95	355	356.40	356	356.45
35	356	357.82	358	359.90	358	359.90	362	363.38	363	363.41
40	363	364.84	367	366.85	367	366.85	369	370.35	370	370.38
45	371	371.61	373	373.58	373	373.57	376	377.36	377	377.37
50	378	380.17	380	381.88	380	381.88	384	385.35	385	385.39
55	386	385.31	388	387.39	388	387.41	390	390.95	390	390.97
60	391	392.41	393	394.40	393	394.41	397	397.90	397	397.94
65	398	399.38	400	401.02	400	401.35	404	404.86	404	404.58
70	405	405.96	407	407.96	407	407.96	410	411.48	411	411.49
75	414	412.81	414	415.24	414	415.27	417	418.48	418	418.51
80	421	419.81	421	421.89	421	421.91	424	425.42	425	425.46
85	426	426.65	428	428.65	428	429.28	432	432.45	432	432.46
90	433	433.95	435	435.90	435	435.57	438	439.34	439	439.36
95	440	440.88	442	442.54	442	442.54	445	446.30	446	446.33
100	447	449.08	450	450.90	450	450.81	453	454.34	454	454.38
105	455	455.85	457	457.51	457	457.51	460	461.32	461	461.35
110	461	463.08	463	464.84	464	464.46	467	468.29	468	468.03
115	469	469.43	471	471.43	471	471.41	474	474.95	474	474.95
120	476	476.38	478	478.41	478	478.39	481	481.93	481	481.94
125	483	483.41	485	485.41	485	485.39	488	488.92	488	488.93
130	490	490.33	492	492.31	492	492.30	495	495.78	495	495.49
135	496	497.21	498	499.24	498	499.26	502	502.43	502	502.47
140	502	504.19	505	506.20	505	506.20	508	509.36	509	509.39
145	510	510.35	512	511.55	512	511.55	510	293.78	509	299.26

Table C.1 Continued from previous page

Raw data			Filtered data											
1mm			median								Butterworth n=4			
			n = 3		n = 4		n = 7		n = 8		fc = 0.6		fc = 0.4	
mm	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
0	312	309.58	311	309.55	311	310.37	309	309.51	309	310.03	309	310.93	311	311.90
1	311	310.52	310	310.90	311	311.40	309	310.91	310	311.46	311	311.59	312	312.56
2	310	312.27	311	312.31	312	312.51	310	312.35	312	312.92	312	312.96	314	313.96
3	314	313.38	313	313.66	313	313.95	311	313.83	312	314.36	313	314.49	315	315.46
4	314	315.03	315	315.15	316	315.56	314	314.31	314	315.08	315	315.95	316	316.59
5	317	316.30	315	316.35	316	317.20	314	316.36	315	316.95	316	317.36	318	318.34
6	317	317.57	318	317.57	318	318.01	317	317.29	317	318.03	318	318.50	319	319.47
7	318	319.05	318	318.69	320	319.52	318	318.69	318	319.19	319	320.31	320	321.26
8	321	320.31	318	320.66	320	320.92	318	320.72	320	321.01	320	321.39	322	322.37
9	321	322.05	322	321.78	323	322.53	321	320.99	321	321.75	322	322.93	323	323.60
10	322	323.43	322	323.09	324	323.63	322	323.10	322	323.53	324	324.39	325	325.32
11	325	324.76	324	324.78	324	325.01	322	324.84	323	325.40	324	325.49	326	326.46
12	326	326.02	325	326.08	326	326.51	324	326.11	325	326.61	326	327.27	328	328.26
13	329	327.48	328	327.40	328	327.58	326	327.10	326	327.56	328	328.43	329	329.38
14	328	328.91	329	328.94	329	329.45	328	329.00	328	329.18	329	329.88	330	330.85
15	329	329.72	329	330.27	331	330.49	329	329.99	329	330.48	331	331.35	332	332.30
16	332	331.37	331	331.33	331	331.94	330	331.83	331	332.36	331	332.49	333	333.44
17	332	333.03	333	332.78	334	333.53	331	332.66	332	333.15	333	334.29	334	334.92
18	335	334.32	333	334.31	334	334.52	332	334.32	334	334.92	334	335.32	336	336.31
19	336	335.79	335	335.80	335	336.04	334	335.91	334	336.44	336	336.53	337	337.49
20	335	337.22	336	337.35	337	337.51	335	337.37	337	337.93	337	337.97	338	339.26
21	339	338.34	336	338.33	338	338.92	336	338.70	338	338.97	338	339.44	340	340.41
22	338	340.22	339	340.34	340	340.52	338	339.83	340	340.63	340	340.90	341	341.89
23	342	341.33	340	341.35	341	341.53	339	341.37	341	341.91	341	342.33	343	343.30
24	342	342.60	343	342.66	344	343.11	342	342.54	342	343.01	343	343.53	344	344.51
25	343	344.06	344	343.75	345	344.56	343	343.72	343	344.17	344	345.29	346	345.92
26	345	345.17	346	345.18	346	345.95	345	345.12	345	345.58	346	346.41	347	347.38
27	346	346.60	346	346.55	348	347.40	346	346.22	346	346.98	347	347.88	348	348.53
28	347	347.99	347	347.94	349	348.49	347	347.93	347	348.42	348	349.32	350	350.28
29	350	349.38	349	349.69	349	349.98	347	349.82	348	350.37	349	350.48	351	351.46
30	350	350.59	350	350.54	352	351.38	350	350.50	350	351.00	351	351.88	352	352.54
31	351	351.81	351	352.35	353	352.88	350	352.38	351	352.87	353	353.43	354	354.37
32	353	353.66	354	353.63	355	354.13	353	353.24	353	353.71	354	354.93	355	355.56
33	354	355.08	354	354.68	356	355.53	354	354.96	354	355.44	355	356.32	357	357.26
34	355	356.37	356	356.76	356	357.00	355	356.83	355	357.37	356	357.48	358	358.43
35	357	357.56	357	357.51	359	358.08	357	357.49	357	358.01	358	358.87	359	359.51

Table C.2 Measurements from a linear resolution test at 1 mm steps using median (n - aperture in samples), Butterworth and FIR filter (n - filter order; fc - cutoff frequency). Data are presented with their pixel values and centroid values.

Raw data			Filtered data											
1mm			Butterworth n=11				FIR n=4				FIR n=11			
			fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
mm	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
0	312	309.58	312	312.51	315	314.96	312	311.91	312	311.92	315	315.18	316	315.46
1	311	310.52	313	313.45	316	315.97	313	312.55	313	312.56	317	316.37	316	316.39
2	310	312.27	314	314.84	317	317.10	315	314.30	314	314.30	318	317.49	317	317.52
3	314	313.38	316	316.36	319	318.87	316	315.44	316	315.46	319	319.29	319	319.32
4	314	315.03	317	317.47	320	319.98	316	316.98	316	316.65	319	320.44	320	320.43
5	317	316.30	318	319.26	321	321.15	319	318.35	319	318.37	322	321.56	322	321.59
6	317	317.57	320	320.36	323	322.89	319	319.53	319	319.52	322	323.31	323	323.01
7	318	319.05	321	322.18	324	324.04	320	321.00	320	321.32	324	324.49	324	324.50
8	321	320.31	322	323.30	325	325.45	323	322.38	323	322.39	326	325.89	326	325.91
9	321	322.05	324	324.48	327	327.00	323	323.99	323	323.65	327	327.44	327	327.42
10	322	323.43	325	326.23	328	328.12	324	325.07	324	325.38	328	328.59	328	328.57
11	325	324.76	327	327.36	330	329.88	327	326.78	327	326.79	330	330.30	330	330.33
12	326	326.02	328	328.85	331	331.04	328	328.02	328	328.02	332	331.47	331	331.50
13	329	327.48	329	330.28	332	332.47	329	329.45	329	329.44	332	332.63	333	332.61
14	328	328.91	331	331.43	334	333.97	330	330.91	330	330.91	333	334.37	334	334.39
15	329	329.72	332	332.87	335	335.09	331	332.05	331	332.07	335	335.55	335	335.54
16	332	331.37	334	334.36	337	336.54	334	333.44	334	333.46	338	337.00	337	337.01
17	332	333.03	335	335.50	338	338.01	334	334.99	334	334.98	337	338.46	338	338.46
18	335	334.32	336	337.22	339	339.13	337	336.34	337	336.35	340	339.52	339	339.55
19	336	335.79	338	338.39	341	340.90	338	337.81	338	337.82	340	341.31	341	341.36
20	335	337.22	339	340.19	342	342.03	340	339.28	340	339.29	343	342.46	342	342.50
21	339	338.34	340	341.33	344	343.52	341	340.40	341	340.42	344	343.94	344	343.96
22	338	340.22	342	342.79	345	345.03	340	342.25	342	341.97	344	345.41	345	345.45
23	342	341.33	343	344.23	346	346.10	344	343.35	343	343.36	346	346.81	346	346.55
24	342	342.60	345	345.41	348	347.91	344	344.57	344	344.56	347	348.37	348	348.37
25	343	344.06	346	346.82	349	349.04	345	346.00	345	345.99	349	349.48	349	349.47
26	345	345.17	347	348.28	350	350.48	347	347.49	347	347.47	350	350.62	350	350.93
27	346	346.60	349	349.42	352	351.94	348	348.58	348	348.57	351	352.39	352	352.39
28	347	347.99	350	351.20	353	353.04	349	350.01	349	350.02	353	353.52	353	353.51
29	350	349.38	352	352.37	355	354.89	352	351.44	352	351.46	354	355.29	355	355.02
30	350	350.59	353	353.42	356	355.95	352	352.58	352	352.58	356	356.39	356	356.39
31	351	351.81	354	355.27	357	357.45	353	354.46	353	354.45	357	357.62	357	357.91
32	353	353.66	356	356.45	359	358.97	355	355.62	355	355.61	358	359.43	359	359.42
33	354	355.08	357	357.83	360	360.06	356	357.02	356	357.33	360	360.51	360	360.50
34	355	356.37	359	359.35	362	361.85	357	358.76	357	358.45	361	362.28	362	362.01
35	357	357.56	360	360.41	363	362.91	359	359.55	359	359.55	362	363.38	363	363.38

Table C.2 Continued from previous page

Raw data			Filtered data											
0.5mm			median								Butterworth n=4			
			n = 3		n = 4		n = 7		n = 8		fc = 0.6		fc = 0.4	
mm	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
0.0	309	310.42	310	309.85	311	310.63	309	310.02	309	310.20	311	311.37	312	311.98
0.5	310	310.94	311	310.94	311	311.43	309	310.33	310	311.45	312	311.92	313	312.91
1.0	311	311.61	311	311.63	313	312.11	311	311.62	311	312.40	312	312.52	313	313.51
1.5	311	312.06	311	312.09	312	312.61	311	312.72	311	313.27	313	313.33	314	314.02
2.0	313	313.26	313	312.93	313	313.44	311	313.31	313	313.86	313	313.92	315	314.91
2.5	315	313.80	314	313.74	314	313.98	312	313.56	313	314.71	314	314.51	315	315.80
3.0	314	314.12	315	314.42	315	314.97	313	314.76	313	314.98	316	315.40	316	316.37
3.5	315	315.40	315	315.03	315	315.54	313	315.03	315	315.89	316	315.96	317	316.96
4.0	316	315.85	315	315.55	316	316.07	314	316.23	316	316.45	316	316.55	317	317.53
4.5	315	318.27	316	316.15	316	316.94	315	316.16	315	317.27	317	317.09	318	318.10
5.0	317	316.75	316	316.70	318	317.54	316	317.33	316	317.55	318	317.97	319	319.00

Filtered data										
Butterworth n=11				FIR n=4				FIR n=11			
fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
312	312.88	315	315.08	311	312.06	311	312.04	315	315.56	315	315.53
313	313.81	316	316.03	312	312.94	313	312.95	316	316.42	316	316.44
314	314.06	317	316.93	313	313.56	313	313.55	317	317.01	317	317.34
314	314.90	317	317.17	313	314.36	314	314.37	317	317.57	318	317.57
315	315.80	318	318.03	315	315.27	315	315.27	318	318.43	318	318.45
316	316.37	319	318.91	315	315.53	315	315.53	318	319.30	319	319.34
316	316.96	319	319.47	316	316.45	316	316.44	320	319.61	320	319.61
317	317.86	320	320.07	317	317.35	317	317.33	320	320.49	320	320.50
318	318.41	321	320.94	318	317.85	318	317.52	321	321.34	321	321.36
319	318.97	322	321.53	318	318.40	318	318.41	322	321.91	322	321.94
319	319.86	322	322.15	319	319.04	319	319.03	322	322.53	322	322.53

Table C.3 Measurements from a linear resolution test at 0.5 mm steps using median (n - aperture in samples), Butterworth and FIR filter (n - filter order; fc - cutoff frequency). Data are presented with their pixel values and centroid values.

Raw data			Filtered data											
0.1mm			median								Butterworth n=4			
			n = 3		n = 4		n = 7		n = 8		fc = 0.6		fc = 0.4	
mm	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
50.1	379	379.88	380	379.29	380	380.07	378	379.30	379	380.08	380	380.46	381	381.46
50.2	379	379.57	380	379.90	380	380.12	379	379.27	379	379.78	380	380.50	381	381.51
50.3	379	379.94	379	379.67	380	380.17	379	380.25	379	380.45	380	380.84	381	381.85
50.4	379	380.00	379	379.72	381	380.51	379	379.72	379	380.45	381	380.90	382	381.90
50.5	380	380.38	380	380.03	380	380.21	380	380.02	380	380.24	380	380.95	382	381.95
50.6	380	380.41	380	380.37	380	380.25	380	379.75	380	380.26	381	381.32	382	382.01
50.7	379	378.16	379	380.73	380	380.95	379	380.79	379	381.04	381	381.37	382	382.06
50.8	380	380.43	380	380.41	380	380.63	380	380.46	380	380.97	381	381.40	382	382.07
50.9	380	378.57	379	380.43	380	381.26	379	380.78	379	381.32	381	381.45	382	382.13
51.0	381	380.54	380	380.49	382	381.37	380	380.79	380	381.35	381	381.82	382	382.84

Filtered data										
Butterworth n=11				FIR n=4				FIR n=11			
fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4		fc = 0.6		fc = 0.4	
Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid	Pixels	Centroid
382	382.01	385	384.89	381	381.52	381	381.51	385	385.29	385	385.31
382	382.39	385	384.93	381	381.54	381	381.54	385	385.33	385	385.35
382	382.40	385	384.97	381	381.59	381	381.58	385	385.36	385	385.38
382	382.80	385	385.03	381	381.97	381	381.64	385	385.43	385	385.44
382	382.84	385	385.07	382	382.35	382	382.33	385	385.49	385	385.50
383	382.88	385	385.14	382	382.38	382	382.38	385	385.54	385	385.56
383	382.94	386	385.50	382	382.41	382	382.41	385	386.23	386	385.93
383	382.93	386	385.52	382	382.43	382	382.42	385	385.90	386	385.93
383	383.00	386	385.88	382	382.46	382	382.47	386	386.26	386	386.29
383	383.41	386	385.95	383	382.55	383	382.56	386	386.37	386	386.38

Table C.4 Measurements from a linear resolution test at 0.1 mm steps using median (n - aperture in samples), Butterworth and FIR filter (n - filter order; fc - cutoff frequency). Data are presented with their pixel values and centroid values.

Appendix D

An Algorithm for Peak Separation by Linear Slope Extension

This algorithm was developed to separate targets close to one another, when their peaks are overlapped. To describe this technique two separate targets were captured. The targets occupied neighbouring positions from the linear resolution test with a 0.1 mm displacement interval. Their original peak profiles, shown in Figure D.1 were superimposed and the profile in Figure D.2 was obtained.

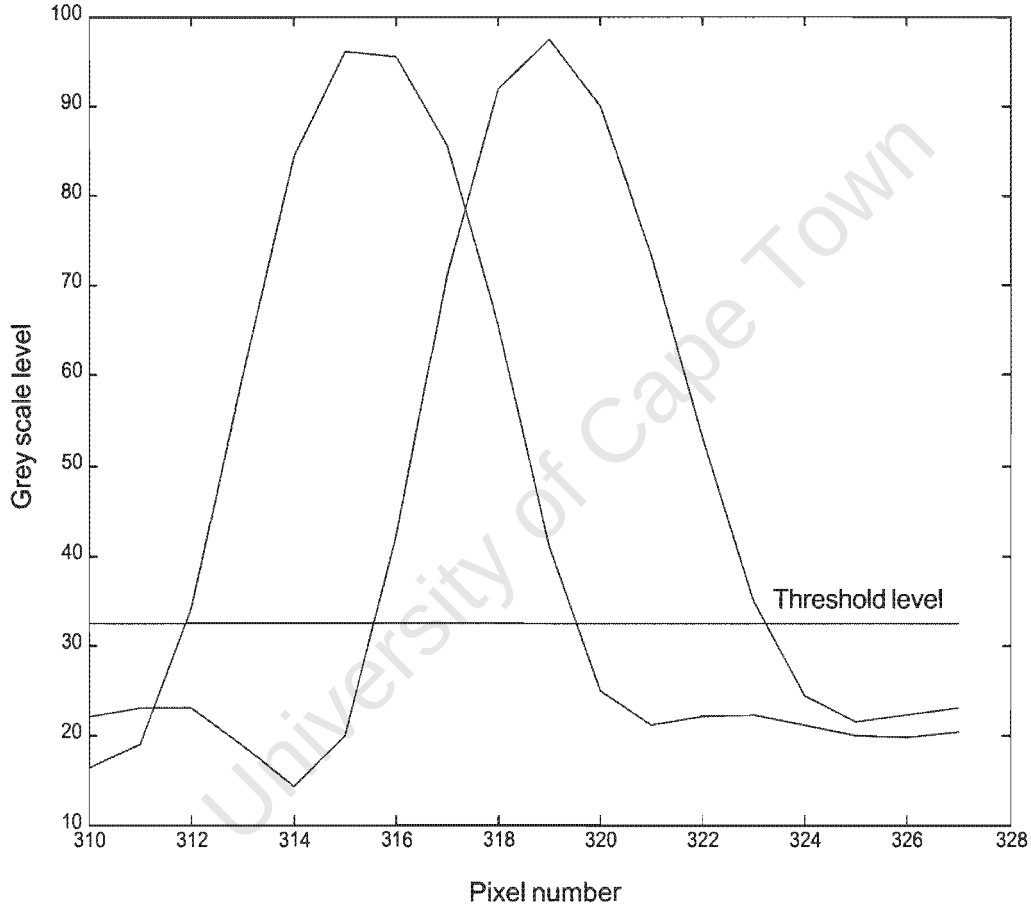


Figure D.1 Two closely spaced separate peaks with their real shapes overlapping each other.

The main objective of the above algorithm is to separate the two peaks and to evaluate the results by comparing the centroid values of the original peaks to those obtained after the separation.

This algorithm calculates an extension to the side of the peak which is hidden by the other peak. This process is illustrated for the one side of the peak on the left, but it could be applied to both, or to all overlapped peaks that are detected on the profile.

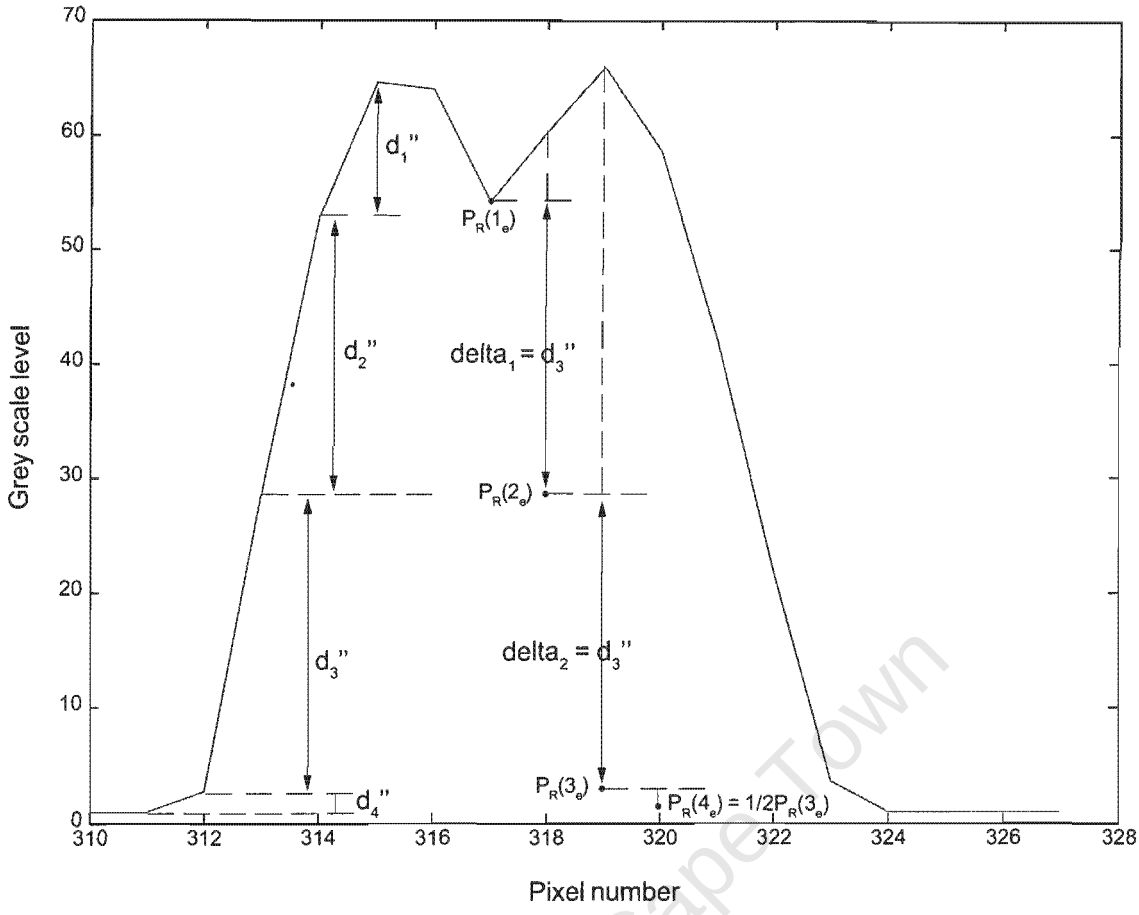


Figure D.2 Peaks from Figure D.1 wrapped together by superimposing their pixel profiles and after the threshold was applied. Extension segments delta on the right hand side of Peak 1 show which section from the other side of the peak is used to restore the missing part by recalculating the new pixel values for those points covered by the other peak.

The input data for the algorithm contain the maximum value of the peak, the pixel position of the maximum and the two positions defining the beginning and the end of the peak. These positions are determined from the sign of the first derivative of the profile, when it changes from negative to positive on both sides of the maximum, or if the signal reaches zero. As the threshold was applied prior to separation, all non-overlapping peaks are separated by zero values. If close peaks are found they are separated at the point of intersection. Next, the algorithm checks in which of the three intervals below the point of intersection falls.

1) $0 < P(n_s) < 1/2\, P_{max}$

2) $1/2\, P_{max} < P(n_s) < 3/4\, P_{max}$

3) $3/4\, P_{max} < P(n_s) < P_{max}$

where

P = value of the peak;

n = current pixel number;

$P(n_s)$, $P(n_e)$ = value of the start point (n_s) and end point (n_e);

P_{max} = value of the peak maximum.

Depending on which interval the intersection point falls into, the extention segment delta is determined to be used for completing the peak shape. According to 1), 2) or 3) delta is:

1* delta = $|P(n_s+1) - P(n_s)|$

2* delta = max[D'] where $[D'] = [d_1', d_2' \ldots d_{n-1}']$
$d_1', d_2' \ldots d_{n-1}'$ differences between each two points on the same side of the peak, denoted by prime.

3* delta = max[D''] where $[D''] = [d_1'', d_2'' \ldots d_{n-1}'']$
$d_1'', d_2'' \ldots d_{n-1}''$ differences between each two points on the opposite side of the peak, denoted by second

This is an iterative process which continues until the value at the point of intersection reaches a level, small enough to distinguish the peak shape, which was covered. Figure D.3 shows the result after the separation of the two peaks from Figure D.1. The difference in their new shapes occured because the calculations were not applied to the intersection point and it retained its value from the superposition. Additional calculations including this point would improve the new peak shapes, however, the evaluation after comparison of the centroid values

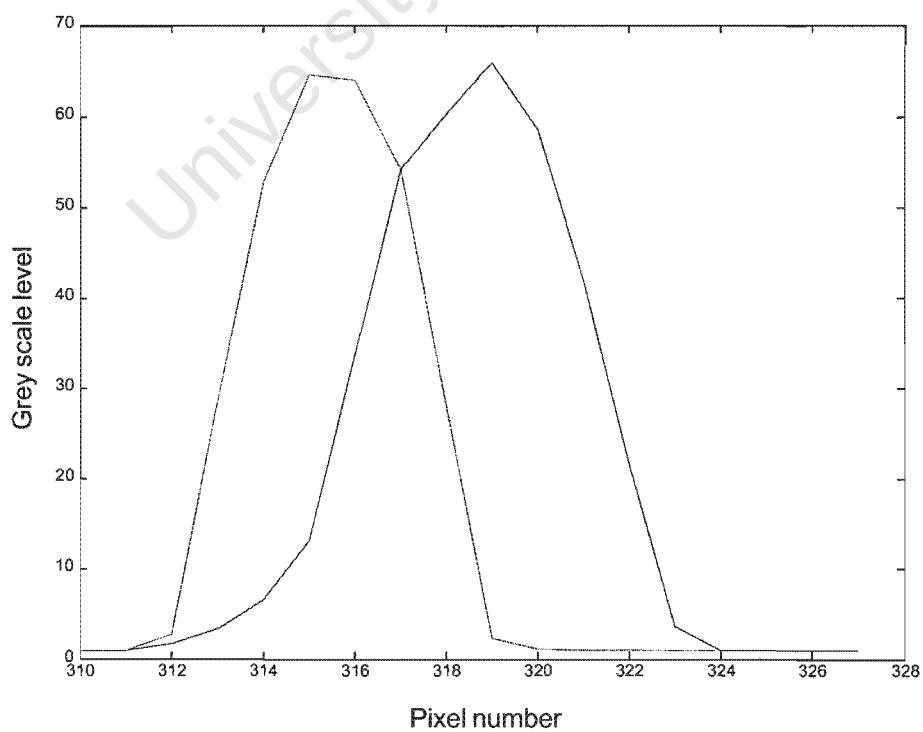


Figure D.3 Separated peaks from Figure D.2 using the algorithm for linear slope extension. The new calculated points are connected together to construct the new shape of each peak.

between the original and the separated shapes showed that the algorithm is sufficient at this level. The block diagram in Figure D.4 shows the main steps of the described algorithm.

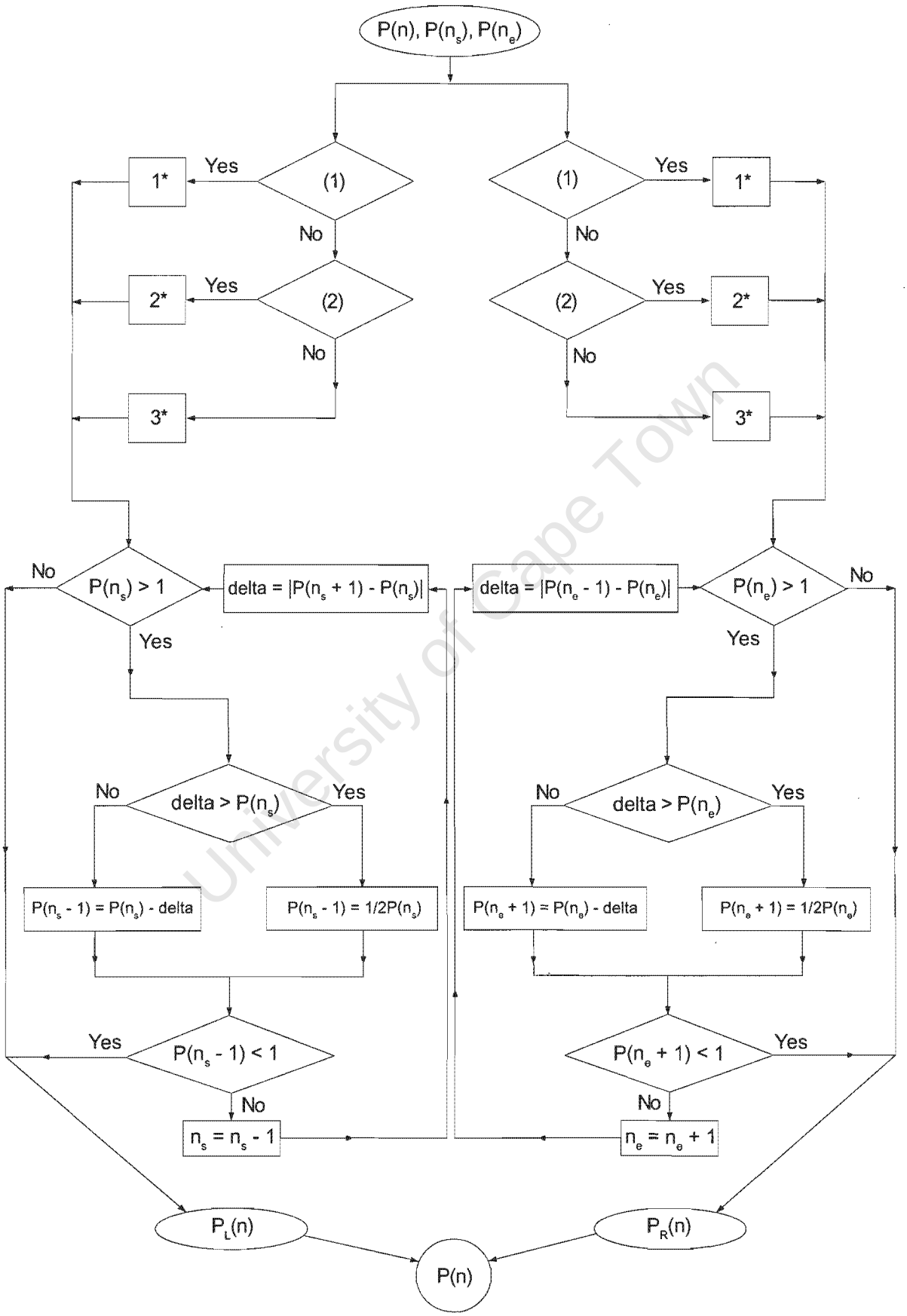


Figure D.4 Block diagram of the algorithm for peak separation by linear slope extension.

Appendix E

Theory and Principles of Operation of the Optical Play Ground Software

This software is a simulator for optical measurements in 3D. The measurement system includes an optical system consisting of different types of lenses and a detector, which could be a rectangular or a linear array of sensitive elements. One or multiple source targets could be positioned in the 3D space and then measured from the lens-detector system, also called a camera. In the process of simulation several of these cameras with different positions and orientations could be created and the measured values from the source targets could be used for testing 3D coordinate reconstruction algorithms, such as the Direct Linear Transformation and the Bundle Adjustment Methods described in Chapter 5.

As a first step, an external XYZ coordinate system in the cartesian space is chosen in which all the simulated components will be positioned. Some of the calculation steps in the process, however, are executed in a local xyz coordinate system. The transformation from the external to the local coordinate system and vice versa includes 3 translations and 3 rotations along the coordinate axes.

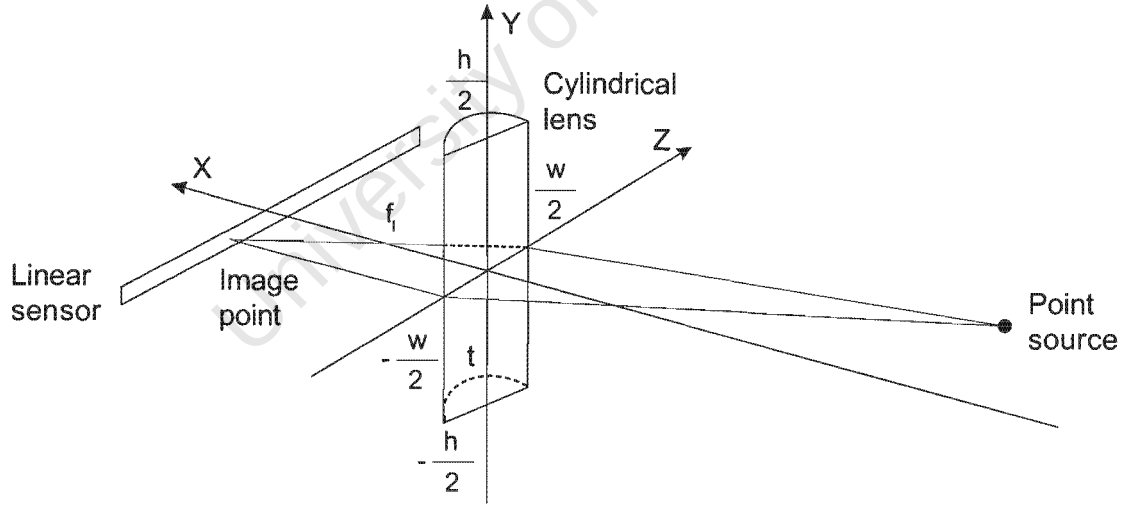


Figure E.1 The external XYZ coordinate system and an image of a point source through a combination of linear sensor and cylindrical lens created by the OPG software.

Figure E.1 shows one example of a cylindrical lens-linear sensor camera, with the centre of the flat surface of the lens positioned at the origin of the XYZ system and a point source lying in the X-Z plane. The image point on the sensor is calculated using ray tracing methods and the Fresnel equations of reflection and refraction.

I. Construction of the cylindrical lens-linear sensor system and the source.

To simulate the above components the software completes the following steps:

1. The cylindrical lens is constructed as an object consisting of an ellipsoid surface with one infinite radius, intersected by a plane that represents the flat surface of a plano-convex cylindrical lens. Equation [E.1] describes the ellipsoid with a centre at the origin of the external coordinate system.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \qquad [E.1]$$

According to the above condition to form a cylinder, the ellipsoid radii are introduced, so that $r_x = r_z$ and $r_y >> r_x$.The orientation of the ellipsoid is provided by the rotation angles $[\delta_x;\delta_y] = [0;0]$. Then the centre of the ellipsoid is displaced at a distance $r_x - t$ along X, where t is the lens thickness, moving the centre of the flat surface to the origin of XYZ. The lens geometry is further defined in 3D by the parameters: t; -w/2; w/2; -h/2; h/2, where w is the lens width and h is the lens height. A frame around the lens to stop the light is also constructed, using a coefficient k = 1.5, applied to the lens geometry in a horizontal and vertical direction.

2. A detector is constructed as a parallelogram using three pairs of intersecting planes, with the front, lower, left corner positioned at [-1;-12.29;-5] and the back, upper, right corner at [1;12.29;5]. Then the detector is displaced along X, so that the cross section parallel to the Y-Z plane is centred at $P_d = [f_l;0;0]$, where f_l is the focal distance of the lens. The orientation of the detector is given by the 3 rotation angles $[\psi_x;\psi_y;\psi_z] = [0;0;0]$. The size of the detector is a linear array of 2048 elements along the Z axis with a total length of 24.58 mm, representing a linear sensor.

3. The lens and the detector are combined to form a camera, which will later provide the measured pixel values from multiple source points positioned in the 3D space. To measure these points and to provide the input data for the tests described in Chapter 5 (Table 5.1) and Appendix F (Test 5.1), four cameras are created with their positions coinciding with the positions of the flat surfaces of the lenses as shown on Figure 5.2. The orientation of the cameras is given by the orientation of the detector *i.e.* [0;0;0]. The external rotation of the cameras at angle φ_k, where k represents the camera number, is provided by rotating the lens, the sensor and the source, so that the light beam traced to the lens always lies in a plane orthogonal to the main axis of the lens. For cameras 2,3 and 4 $\varphi = [\varphi_x;\varphi_y;\varphi_z]$ is [-90°;0;0], [45°;0;0]and [135°;0;0] respectively.

4. The source is constructed as a semispherical surface and positioned at various locations to simulate multiple target points with coordinates $P_i = [X_i;Y_i;Z_i]$. Before the procedure of ray tracing has started, each source position is transformed from the external XYZ into the local xyz coordinate system of each camera, so that the flat surface of the lens is always at [0;0;0], using the following equation:

$$\mathbf{r}_i' = (\mathbf{r}_i - \mu)\mathbf{R}_x\mathbf{R}_y\mathbf{R}_z \qquad [E.2]$$

where:

$\mathbf{r}_i' = [x_i;y_i;z_i]$ = vector of the transformed 3D coordinates in the local system xyz

$\mathbf{r}_i = [X_i;Y_i;Z_i]$ = vector of the original 3D coordinates in the external system XYZ

$\mu = [\mu_x;\mu_y;\mu_z]$ = translation vector defining the camera position

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta_x & -\sin\beta_x \\ 0 & \sin\beta_x & \cos\beta_x \end{bmatrix} = \text{rotation matrix around x}$$

$$\mathbf{R}_y = \begin{bmatrix} \cos\beta_y & 0 & \sin\beta_y \\ 0 & 1 & 0 \\ -\sin\beta_y & 0 & \cos\beta_y \end{bmatrix} = \text{rotation matrix around y}$$

$$\mathbf{R}_z = \begin{bmatrix} \cos\beta_z & -\sin\beta_z & 0 \\ \sin\beta_z & \cos\beta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} = \text{rotation matrix around z}$$

$\beta = [\beta_x;\beta_y;\beta_z]$ = rotation angles around xyz

In this particular case $[\beta_x,\beta_y,\beta_z] = [\varphi_x,\varphi_y,\varphi_z]$.

II. Ray tracing method.

The next steps of the process of ray tracing are described in the two main projection planes separately. For a spherical lens the calculations are considered in the horizontal X-Z plane and in the vertical X-Y plane. The cylindrical lens has an optical power only in the X-Z plane and although the elliptical suface has a curvature along the Y axis, the effect of the light conversion in this direction should be considered negligible.

Every source target, after its position has been transformed in relation to the camera, has

occupied a new position $P_{io}(X_{io};Y_{io};Z_{io})$. From this point, the chief ray is drawn connecting the point and the centre of the lens. This ray subtends an angle α_z with the X axis, which is the angle of the horizontal projection.

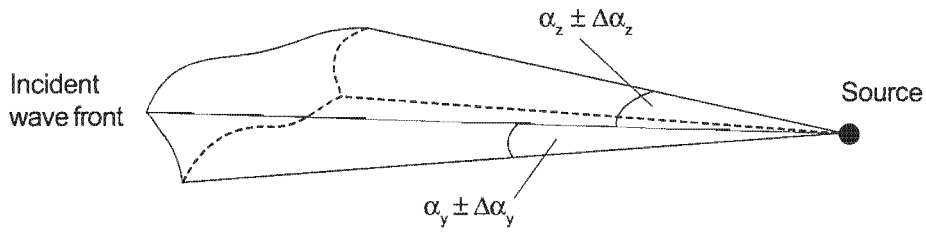


Figure E.2 The angles of the horizontal and the vertical projections from the source to the lens which determine the range of the bundle of rays to be directed to the lens surface.

As long as one dimensional conversion will be considered for the further calculations, only the Z_k component will be used. The projections in the Y direction will be constructed later in a similar way, using the vertical projection angle α_y.

$$\alpha_z = \text{arctg}\frac{Z_{io}}{X_{io}} \qquad \alpha_y = \text{arctg}\frac{Z_{io}}{Y_{io}} \qquad [E.3]$$

To construct multiple rays in a shape of a fan, as shown in Figure E.2, an angular range is defined as $\alpha_z \pm \Delta\alpha_z$, so that the rays projected from the source could intersect with the lens surface. They will hit the flat surface, each of them following a small angular step, at a point $P_k(0;Y_k;Z_k)$, away from the centre. The procedure of this ray tracing is presented in Figure E.3 together with an illustration of the refraction of the ray according to Snell's Law. For the further calculation the distance from the centre in Y and Z direction is given by:

$$Z_k = Z_{io} \pm \Delta Z_k = X_{io}\text{tg}(\alpha_z \pm \Delta\alpha_z)$$
$$Y_k = Y_{io} \pm \Delta Y_k = X_{io}\text{tg}(\alpha_y \pm \Delta\alpha_y) \qquad [E.4]$$

and from this point on the flat surface and the position of the source, the first ray is drawn as a line passing through two points in the space

$$\frac{x - X_{io}}{X_{io} - X_k} = \frac{y - Y_{io}}{Y_{io} - Y_k} = \frac{z - Z_{io}}{Z_{io} - Z_k} \qquad [E.5]$$

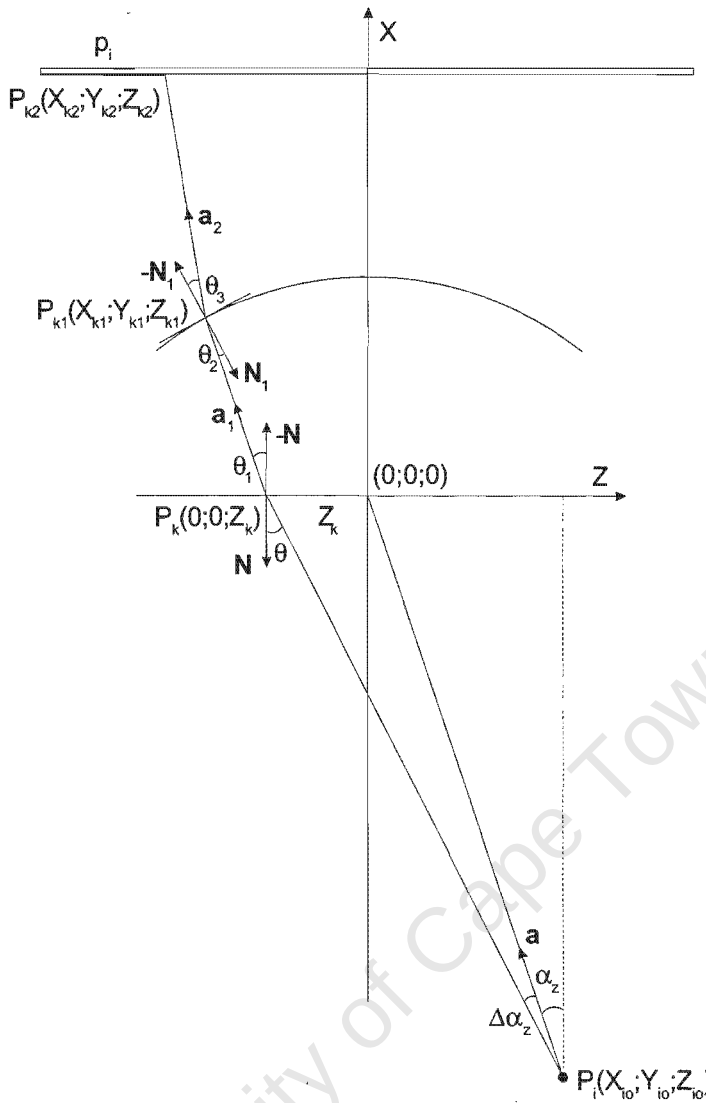


Figure E.3 A projection of a single ray released from a point source in the space towards a cylindrical lens. The ray is refracted first from the flat surface and then from the elliptical surface and directed to the linear sensor behind the lens.

This line could also be represented by its direction vector **a**, which subtends with the normal vector of the flat surface **N** an angle θ, coinciding with the angle of incidence. This angle could be determined from the equation of the cross product between the two vectors,

$$\sin\theta = \frac{\mathbf{a} \times \mathbf{N}}{|\mathbf{a}||\mathbf{N}|} \qquad [E.6]$$

from where Snell's Law could be applied to find out the angle of the refracted ray θ_1.

$$\sin\theta_1 = \frac{n}{n_1} \sin\theta \qquad [E.7]$$

and

$$\cos\theta_1 = \sqrt{1 - \sin^2\theta_1} \qquad [E.8]$$

$$\cos\theta_1 = -\frac{\mathbf{a}_1 \times \mathbf{N}}{|\mathbf{a}_1||\mathbf{N}|} \qquad [E.9]$$

The direction vector of the refracted ray $\mathbf{a}_1$ could be determined as a line defined by two intersecting planes: the horizontal X-Z plane; and the vertical plane, passing through P_k, in a direction defined by the angle θ_1, and perpendicular to the X-Z plane.

As a next step, the intersection point between the refracted ray and the elliptical surface is calculated using the transformed equation [E.1] and the line defined by $\mathbf{a}_1$.

$$\frac{x^2}{r_x^2} + \frac{z^2}{r_x^2} = 1 \qquad \begin{aligned} x &= X_k + A_{1x}t \\ z &= Z_k + A_{1z}t \end{aligned} \qquad [E.10]$$

This point $P_{k1}(X_{k1};Y_{k1};Z_{k1})$ is the point of incidence for the elliptical surface at which the second refraction is calculated. The normal vector to the curved surface $\mathbf{N}_1$ is given by:

$$\frac{x - X_{k1}}{\frac{\partial F}{\partial x}} = \frac{y - Y_{k1}}{\frac{\partial F}{\partial y}} = \frac{z - Z_{k1}}{\frac{\partial F}{\partial z}} \qquad [E.11]$$

where F(x,y,z) is found from [E.1]. Then repeating the steps from [E.6] through [E.8] it could be obtained:

$$\sin\theta_2 = \frac{\mathbf{a}_1 \times \mathbf{N}_1}{|\mathbf{a}_1||\mathbf{N}_1|} \qquad [E.12]$$

$$\sin\theta_3 = \frac{n_1}{n}\sin\theta_2 \qquad [E.13]$$

$$\cos\theta_3 = \sqrt{1 - \sin^2\theta_3} \qquad [E.14]$$

The normal $\mathbf{N}_1$ could also be represented by its direction cosines: $\cos\gamma_x$, $\cos\gamma_y$, $\cos\gamma_z$. The angle γ_x could then be used to calculate the angle which the ray from the second refraction, given by the vector $\mathbf{a}_2$, subtends with the X axis. Now, working in the X-Z plane, the line that determines the refracted ray could be defined as:

$$x - X_{k1} = k(z - Z_{k1})$$
$$k = tg(\gamma_x - \theta_3) \qquad [E.15]$$

After this line is constructed the intersection point with the midline of the detector, parallel to Z, is found at $P_{k2}(X_{k2};Y_{k2};Z_{k2})$. The distance from the beginning of the sensor to the point P_{k2}, denoted as p_i, is used to calculate the position of the projected point on the linear sensor.

After multiple rays are cast to the lens following the solid angle between α_y and α_z, the refracted rays are summed at the projected locations on the sensor, representing the positions of the sensor elements. The result from this summation is such that the majority of the refracted rays are falling into several adjacent elements and few are being dispersed at larger distance, corresponding to the effect of focusing from the lens, as shown in Figure 5.1 in Chapter 5. Then by dividing the distance p_i to the size of the sensor element gives the sensor value in pixels. These pixel values are rounded off to the nearest integer.

During the calculation of the amplitudes of the refracted rays the impact of the magnetic and electric vectors is accounted for and to reduce the calculation time the software uses only the refracted rays, ignoring the effects from high order reflections.

This numerical model, simulating a lens and a CCD sensor and applying ray tracing methods to measure point sources positioned in 3D space, is implemented in Matlab 5.1 and 5.2 code where the final data are available in a matrix form for further processing.

Appendix F

3D Reconstruction of Simulated Data Points Using DLT and BA Methods

This example uses 12 points with known XYZ coordinates, which were used to test the DLT and the BA Methods. The first test uses OPG software to simulate the points and to measure their pixel values according to the setup in Figures 5.2 and 5.3.

Test 5.1 Simulated points by Optical Playground. DLT algorithm used for calibration and reconstruction.

The OPG output data for each of the 4 cameras, where the pixel values were multiplied by the pixel size (0.012 mm) and presented in millimetres.

Camera values simulated by the software (mm)

Cam1	Cam2	Cam3	Cam4
12.32	12.28	10.24	14.38
11.27	11.71	9.96	13.16
10.17	11.08	9.84	11.75
9.33	10.60	9.75	10.65
13.60	10.74	12.45	14.00
14.42	9.74	13.98	13.64
13.80	12.77	11.02	15.71
15.21	13.18	11.90	16.79
16.54	13.58	12.74	17.83
10.29	14.55	7.43	14.34
9.72	15.19	6.68	14.29
8.89	16.11	5.59	14.22

Next, according to section 5.1, the algorithm performs the calibration calculating the L parameters and the distance to the principal point on the sensor, where the sensor length is 24.58 mm.

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
11.12	12.12	11.38	19.95

This is followed by the coordinate reconstruction and the errors were calculated as a difference between the original and the reconstructed coordinate values. For error evaluation the mean,

the standard deviation, the absolute mean, the absolute standard deviation and the absolute maximum error from the reconstruction are given.

Differences (mm)

X	Y	Z
0.33	0.18	0.13
1.74	-0.12	-0.08
3.86	0.19	-0.07
-0.87	0.30	0.04
-0.02	0.00	-0.11
0.03	0.00	0.03
-1.24	0.10	0.00
1.04	0.34	0.04
-0.30	0.04	-0.02
1.19	-0.16	0.19
0.79	-0.02	-0.01
0.27	0.09	0.03

mean (mm)

0.57	0.08	0.02

standard deviation (mm)

1.28	0.15	0.08

absolute mean (mm)

0.97	0.13	0.06

absolute standard deviation (mm)

1.01	0.11	0.05

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
-1.5000	-0.0010	0.0010	-1.4997	-0.0008	0.0011
-1.5500	0.0490	0.0280	-1.5483	0.0489	0.0279
-1.7300	0.1120	0.0640	-1.7261	0.1122	0.0639
-1.9000	0.1715	0.0980	-1.9009	0.1718	0.0980
-1.6700	-0.0660	0.0790	-1.6700	-0.0660	0.0789
-1.9000	-0.1235	0.1480	-1.9000	-0.1235	0.1480
-1.5500	-0.0710	-0.0220	-1.5512	-0.0709	-0.0220
-1.7100	-0.1510	-0.0460	-1.7090	-0.1507	-0.0460
-1.9000	-0.2460	-0.0745	-1.9003	-0.2460	-0.0745
-1.6600	0.1010	-0.1135	-1.6588	0.1008	-0.1133
-1.7500	0.1370	-0.1540	-1.7492	0.1370	-0.1540
-1.9000	0.1970	-0.2215	-1.8997	0.1971	-0.2215

At the end of the test are given the values of the DLT calibration parameters used in the equation for the reconstruction .

L values

0.1782	0.1801	0.2315	0.1060
0.4544	-0.0015	0.4303	0.1565
0.0003	0.4637	-0.4092	0.2682
10.4885	11.1544	68.6524	-9.6496
0.0146	0.0147	0.0188	0.0085
-0.0005	-0.0001	0.0005	-0.0019
0.0001	-0.0001	0.0013	0.0062

Test 5.2 Generated points according to Perspective projection theory using XYZ coordinates from Test 5.1.

The second test generated the so-called "perfect camera values" using the XYZ coordinates and the camera parameters as described in section 5.3. The first five parameters were measured in millimetres, while the next six were the sensors' direction cosines related to the external XYZ system. The meanings of the parameters was explained in section 5.3. This data generation did not take into account any optical effect, such as focusing and lens distortion and was done according to a perfect geometric projection.

Generated camera values (mm)

Cam1	Cam2	Cam3	Cam4
12.27	12.27	11.85	12.76
13.46	11.62	10.57	12.36
14.69	10.92	9.25	11.94
15.63	10.38	8.24	11.62
10.83	10.54	11.69	14.95
9.89	9.41	11.61	16.37
10.60	12.82	13.44	13.54
9.02	13.29	14.92	14.28
7.50	13.74	16.36	15.00
14.54	14.82	12.09	9.30
15.19	15.55	12.17	8.31
16.13	16.60	12.28	6.87

Camera parameters

37	37	37	37
12.29	12.29	12.29	12.29
0	0	0	0
0	0	-95	95
-80	0	70	70
0	0	0	0
-1	0	0.707	0.707
0	1	0.707	-0.707
-1	-1	-1	-1
0	0	0	0
0	0	0	0

As a next step, the DLT algorithm was used for calibration and reconstruction of the generated data and the results below follow the same order as for Test 5.1.

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
12.29	12.21	12.02	12.26

Differences (mm)

X	Y	Z
0.0172	-0.0015	0.0017
0.0281	0.0045	-0.0017
-0.0919	-0.0043	0.0015
-0.0336	-0.0013	0.0031
0.1751	0.0076	-0.0112
-0.0875	-0.0066	0.0084
0.1416	0.0030	0.0023
-0.0562	-0.0116	-0.0004
0.0231	0.0073	0.0000
-0.3852	0.0204	-0.0247
-0.3314	0.0258	-0.0295
0.4835	-0.0487	0.0557

mean (mm)

-0.0098	-0.0004	0.0004

standard deviation (mm)

0.2167	0.0178	0.0200

absolute mean (mm)

0.1545	0.0119	0.0117

absolute standard deviation (mm)

0.1522	0.0132	0.0162

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
-1.5000	-0.0010	0.0010	-1.5000	-0.0010	0.0010
-1.5500	0.0490	0.0280	-1.5500	0.0490	0.0280
-1.7300	0.1120	0.0640	-1.7301	0.1120	0.0640
-1.9000	0.1715	0.0980	-1.9000	0.1715	0.0980
-1.6700	-0.0660	0.0790	-1.6698	-0.0660	0.0790
-1.9000	-0.1235	0.1480	-1.9001	-0.1235	0.1480
-1.5500	-0.0710	-0.0220	-1.5499	-0.0710	-0.0220
-1.7100	-0.1510	-0.0460	-1.7101	-0.1510	-0.0460
-1.9000	-0.2460	-0.0745	-1.9000	-0.2460	-0.0745
-1.6600	0.1010	-0.1135	-1.6604	0.1010	-0.1135
-1.7500	0.1370	-0.1540	-1.7503	0.1370	-0.1540
-1.9000	0.1970	-0.2215	-1.8995	0.1970	-0.2214

L values

22.84889	19.58045	-8.21726	-11.86928
-68.76753	-0.02110	-17.55147	-25.26191
-0.00059	58.89446	-17.38823	25.29174
11.78704	7.99225	-422.85967	647.52837
1.85917	1.59343	-0.66875	-0.96593
0.00011	-0.00179	-0.00336	0.00156
-0.00005	-0.00333	0.01021	0.00057

Finally, the Bundle Adjustment Method was used for reconstruction of the generated data for verification of the results obtained with the DLT Method. Below are presented the differences from the original and reconstructed data and the error evaluation.

Differences (mm)*1.0e-10

X	Y	Z
-0.0045	0.0001	0.0001
0.0455	-0.0014	-0.0011
-0.4911	0.0318	0.0210
-0.0364	0.0045	0.0034
0.0364	0.0011	-0.0023
-0.2910	-0.0205	0.0273
-0.1137	-0.0053	-0.0009
-0.5411	-0.0475	-0.0102
-0.4366	-0.0568	-0.0145
0.3092	-0.0205	0.0182
-0.2547	0.0159	-0.0227
-0.9459	0.1091	-0.1046

mean (mm)*1.0e-10

-0.2270	0.0009	-0.0072

standard deviation (mm)*1.0e-10

0.3390	0.0421	0.0340

absolute mean (mm)*1.0e-10

0.2922	0.0262	0.0189

absolute standard deviation (mm)*1.0e-10

0.2792	0.0321	0.0287

Appendix G

3D Reconstruction from a Calibration Frame Using DLT and BA Methods

Tests with a calibration frame in Figure 5.8 were performed using rectangular cameras equipped with cylindrical lenses and positioned according to the setup in Figure 5.9. The frame was measured at two distances - 3 m and 1 m from the cameras. The camera and the lens were orientated, so the the lens was orthogonal to the horizontal sensor direction for camera position (2) in Figure 5.9 and to the vertical sensor direction for camera positions (1), (3) and (4). The CCD sensor size was measured 6.78 mm in the horizontal and 5.08 mm in the vertical direction.

Test 5.3 Measured points from the calibration frame placed at 3m distance from the cameras. DLT algorithm used for calibration and reconstruction.

Similar to the tests in Appendix F the camera values were converted in millimetres by multiplying the pixel value to the pixel size.

Camera values measured from the calibration frame (mm)

Cam1	Cam2	Cam3	Cam4
3.48	1.87	4.51	4.90
3.47	2.27	4.77	4.67
3.46	2.60	5.00	4.46
3.25	2.89	5.02	4.13
2.79	2.79	4.63	3.82
2.38	2.73	4.26	3.53
2.15	2.11	3.68	3.69
2.15	2.52	3.94	3.46
2.89	1.63	3.91	4.59
2.43	1.72	3.63	4.16
3.40	2.16	4.44	4.68
3.41	2.49	4.68	4.47
3.07	1.80	3.91	4.58
2.73	1.86	3.72	3.84
2.35	1.91	3.50	3.94
2.97	2.91	4.61	4.27
2.59	2.83	4.30	3.58
2.21	2.47	3.81	3.50
2.20	2.16	3.58	3.70

The DLT algorithm was used for calibration and reconstruction of the measured data and the data report below includes: distance to the principal point; differences between the original and reconstructed coordinates; error evaluation for the entire data set; the original and the reconstructed values and the values of the L parameters.

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
11.14	2.12	3.81	4.43

Differences (mm)

X	Y	Z
-14.59	-8.15	-146.61
-9.53	-3.36	-25.95
0.21	-0.49	6.06
-5.34	-1.69	-102.75
0.93	-8.44	-31.26
5.31	4.07	8.01
3.16	19.38	-7.33
2.56	31.14	-15.56
0.88	-3.76	-113.42
-13.66	-8.40	-105.43
0.12	3.13	1.04
8.12	-3.18	-130.65
7.04	-0.35	19.70
0.73	-11.04	-15.24
1.51	1.57	8.20
2.84	-0.96	126.33
1.20	-9.95	195.07
-6.33	-36.50	-98.66
10.94	-34.51	-14.53

mean (mm)

-0.21	-3.76	-23.31

standard deviation (mm)

6.70	14.68	82.55

absolute mean (mm)

5.00	10.00	61.67

absolute standard deviation (mm)

4.46	11.39	59.62

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-2.7480	0.0574	0.7489	-2.8946
-0.0100	0.7570	-2.7470	-0.0195	0.7536	-2.7730
-0.0780	0.7570	-2.7450	-0.0778	0.7565	-2.7389
-0.1380	0.7990	-2.7540	-0.1433	0.7973	-2.8567
-0.1170	0.8960	-2.7650	-0.1161	0.8876	-2.7963
-0.1000	0.9830	-2.7880	-0.0947	0.9871	-2.7800
0.0230	1.0350	-2.7940	0.0262	1.0544	-2.8013
-0.0570	1.0360	-2.7930	-0.0544	1.0671	-2.8086
0.1240	0.8770	-2.7650	0.1249	0.8732	-2.8784
0.1050	0.9730	-2.7860	0.0913	0.9646	-2.8914
0.0490	0.7580	-3.0600	0.0491	0.7611	-3.0590
-0.0310	0.7580	-3.0600	-0.0229	0.7548	-3.1906
0.1280	0.8370	-3.0420	0.1350	0.8366	-3.0223
0.1130	0.9150	-3.0280	0.1137	0.9040	-3.0432
0.0970	1.0010	-3.0120	0.0985	1.0026	-3.0038
-0.1250	0.8570	-3.0390	-0.1222	0.8560	-2.9127
-0.1090	0.9350	-3.0250	-0.1078	0.9251	-2.8299
-0.0290	1.0330	-3.0100	-0.0353	0.9965	-3.1087
0.0410	1.0330	-3.0090	0.0519	0.9985	-3.0235

After the calibration the DLT algorithm calculated the following parameters

L values

0.00025	0.00052	-0.00075	0.00125
-0.00716	-0.00051	-0.00138	-0.00094
-0.00055	0.00084	0.00106	0.00123
7.69656	2.58458	5.83349	4.40665
0.00006	-0.00008	0.00009	0.00028
-0.00059	-0.00022	-0.00004	-0.00021
-0.00019	0.00032	0.00021	0.00028

The same steps were completed for the test which measured the calibration frame at the second distance.

Test 5.4 Measured points from the calibration frame placed at 1m distance from the cameras. DLT algorithm used for calibration and reconstruction.

Camera values measured from the calibration frame (mm)

Cam1	Cam2	Cam3	Cam4
3.73	0.86	3.97	4.78
3.73	1.50	4.47	4.38
3.72	2.06	4.86	4.03
3.35	2.52	4.90	3.42
2.55	2.37	4.26	2.94
1.84	2.26	3.69	2.51
1.40	1.34	2.72	2.83
1.42	1.93	3.17	2.42
2.67	0.50	2.98	4.27
1.88	0.67	2.55	3.58
3.49	1.60	3.57	4.10
3.48	2.07	3.95	3.78
2.98	1.09	2.90	4.07
2.45	1.15	2.64	3.67
1.84	1.22	2.35	3.19
2.84	2.64	3.98	2.94
2.24	2.54	3.57	2.58
1.65	2.03	2.86	2.51
1.63	1.59	2.51	2.80

The DLT algorithm was used for calibration and reconstruction of the measured data and the data report below includes: distance to the principal point; differences between the original and reconstructed coordinates; error evaluation for the entire data set; the original and the reconstructed values and the values of the L parameters.

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
9.32	3.81	2.72	9.07

Differences (mm)

X	Y	Z
11.79	-11.83	-12.14
2.91	-9.79	-18.78
-0.84	-12.68	-33.85
-3.07	3.18	-7.34
-0.29	8.08	-2.59
-1.03	2.73	24.32
-0.58	0.38	-6.38
0.41	-0.44	2.14
-7.85	4.43	26.59
-8.27	8.89	23.41
-1.37	4.88	8.53
0.92	-5.42	-2.25
-3.85	0.53	-3.53
-2.24	1.24	-0.20
6.04	-0.12	-34.29
0.66	-0.46	15.76
1.77	5.65	14.22
5.95	-0.68	-21.62
8.07	-6.03	-46.81

mean (mm)

0.48	-0.39	-3.94

standard deviation (mm)

4.86	6.07	20.10

absolute mean (mm)

3.57	4.60	16.04

absolute standard deviation (mm)

3.32	3.97	12.74

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-1.0040	0.0838	0.7452	-1.0161
-0.0100	0.7570	-1.0030	-0.0071	0.7472	-1.0218
-0.0780	0.7570	-1.0010	-0.0788	0.7443	-1.0348
-0.1380	0.7990	-1.0100	-0.1411	0.8022	-1.0173
-0.1170	0.8960	-1.0210	-0.1173	0.9041	-1.0236
-0.1000	0.9830	-1.0440	-0.1010	0.9857	-1.0197
0.0230	1.0350	-1.0500	0.0224	1.0354	-1.0564
-0.0570	1.0360	-1.0490	-0.0566	1.0356	-1.0469
0.1240	0.8770	-1.0210	0.1161	0.8814	-0.9944
0.1050	0.9730	-1.0420	0.0967	0.9819	-1.0186
0.0490	0.7580	-1.3160	0.0476	0.7629	-1.3075
-0.0310	0.7580	-1.3160	-0.0301	0.7526	-1.3183
0.1280	0.8370	-1.2980	0.1242	0.8375	-1.3015
0.1130	0.9150	-1.2840	0.1108	0.9162	-1.2842
0.0970	1.0010	-1.2680	0.1030	1.0009	-1.3023
-0.1250	0.8570	-1.2950	-0.1243	0.8565	-1.2792
-0.1090	0.9350	-1.2810	-0.1072	0.9406	-1.2668
-0.0290	1.0330	-1.2660	-0.0231	1.0323	-1.2876
0.0410	1.0330	-1.2650	0.0491	1.0270	-1.3118

L values

0.0001	-0.0130	0.0137	0.0073
-0.0079	-0.0013	0.0073	-0.0089
-0.0010	-0.0050	0.0045	-0.0008
8.1547	-1.5186	-8.4484	11.5121
0.0000	-0.0001	0.0008	0.0003
-0.0007	-0.0007	-0.0008	-0.0005
-0.0004	-0.0013	0.0021	-0.0007

Test 5.5 Generated points according to Perspective projection theory using XYZ coordinates from Test 5.3.

This test projected the XYZ positions through the perspective centre using the Perspective projection, which is the forward problem of the Bundle Adjustment Method and produced the following camera values.

Camera values generated by the software (mm)

Cam1	Cam2	Cam3	Cam4
5.29	9.30	9.56	11.43
5.30	7.80	10.63	10.37
5.30	6.55	11.52	9.49
4.52	5.44	11.72	8.15
2.74	5.82	10.16	7.15
1.16	6.10	8.77	6.21
0.23	8.32	6.52	7.11
0.21	6.88	7.53	6.08
3.09	10.21	7.30	10.50
1.34	9.82	6.28	8.97
4.99	8.32	9.10	10.24
4.99	7.00	10.03	9.31
3.70	9.65	7.29	10.28
2.41	9.43	6.57	9.23
0.97	9.20	5.76	8.05
3.37	5.46	10.02	7.09
2.07	5.74	8.95	6.38
0.43	7.09	6.87	6.19
0.43	8.27	6.05	7.02

For the camera parameters was used the setup from Figure 5.9, where the first five parameters in the list are measured in millimetres, while the next six are direction cosines between the local coordinate system of the sensor and the external XYZ system.

Camera parameters

50	50	50	50
2.54	3.39	2.54	2.54
250	-250	291	-291
907	907	1079	1079
-25	-25	-25	-25
0	1	-0.707	0.707
-1	0	-0.707	-0.707
0	0	0	0
0	0	0	0
0	0	0	0
1	1	1	1

The meanings of the camera parameters was explained in section 5.3.

Next, the DLT algorithm was used for calibration and reconstruction of the generated data and the data report below includes: the distance to the principal point; the differences between the original and reconstructed coordinates; error evaluation for the entire data set; the original and the reconstructed values and the values of the L parameters..

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
2.54	3.39	2.54	2.52

Differences (mm)

X	Y	Z
0.0002	0.0001	0.0003
-0.0008	0.0005	-0.0005
0.0001	-0.0030	-0.0020
0.0001	-0.0015	-0.0022
0.0002	0.0011	0.0008
-0.0004	-0.0004	-0.0082
-0.0001	-0.0005	0.0076
0.0015	0.0000	-0.0026
0.0004	0.0016	0.0030
-0.0004	-0.0010	0.0044
-0.0005	-0.0008	0.0027
0.0006	0.0013	0.0029
-0.0008	-0.0008	0.0019
0.0013	-0.0006	0.0079
0.0008	0.0003	-0.0102
0.0003	0.0023	-0.0033
-0.0001	0.0004	-0.0017
-0.0021	0.0000	-0.0013
-0.0001	-0.0002	-0.0036

mean (mm)*1.0e-03

0.0120	-0.0615	-0.2108

standard deviation (mm)

0.0008	0.0012	0.0045

absolute mean (mm)

0.0006	0.0009	0.0035

absolute standard deviation (mm)

0.0005	0.0008	0.0028

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-2.7480	0.0720	0.7570	-2.7480
-0.0100	0.7570	-2.7470	-0.0100	0.7570	-2.7470
-0.0780	0.7570	-2.7450	-0.0780	0.7570	-2.7450
-0.1380	0.7990	-2.7540	-0.1380	0.7990	-2.7540
-0.1170	0.8960	-2.7650	-0.1170	0.8960	-2.7650
-0.1000	0.9830	-2.7880	-0.1000	0.9830	-2.7880
0.0230	1.0350	-2.7940	0.0230	1.0350	-2.7940
-0.0570	1.0360	-2.7930	-0.0570	1.0360	-2.7930
0.1240	0.8770	-2.7650	0.1240	0.8770	-2.7650
0.1050	0.9730	-2.7860	0.1050	0.9730	-2.7860
0.0490	0.7580	-3.0600	0.0490	0.7580	-3.0600
-0.0310	0.7580	-3.0600	-0.0310	0.7580	-3.0600
0.1280	0.8370	-3.0420	0.1280	0.8370	-3.0420
0.1130	0.9150	-3.0280	0.1130	0.9150	-3.0280
0.0970	1.0010	-3.0120	0.0970	1.0010	-3.0120
-0.1250	0.8570	-3.0390	-0.1250	0.8570	-3.0390
-0.1090	0.9350	-3.0250	-0.1090	0.9350	-3.0250
-0.0290	1.0330	-3.0100	-0.0290	1.0330	-3.0100
0.0410	1.0330	-3.0090	0.0410	1.0330	-3.0090

L values

-0.000014	-1.999745	1.374570	-1.405644
2.034587	0.000015	1.374527	1.405584
0.103338	0.135591	0.098839	0.100992
-1842.841257	-496.527878	-1880.421885	-1923.208923
-0.000005	-0.000002	0.000005	0.000006
-0.000002	0.000001	-0.000002	-0.000013
0.040685	0.039995	0.038893	0.039763

The Bundle Adjustment Method was also used for reconstruction of the generated data and the data report includes: the differences between the original and the reconstructed coordinates; the mean; the standard deviation; the absolute mean and the absolute standard deviation, calculated from the differences.

Differences (mm)*1.0e-11

X	Y	Z
0	0	0
-0.0114	0	0
0.0057	0	0
0	0.0455	-0.1819
0.0057	0	0
-0.0028	0.0341	0.1819
0.0057	0	-0.1819
0.0057	0	0
0.0227	0	-0.1819
0	0.0114	0
-0.0227	0	0.3638
0	0.0455	0
0.0227	0.0455	0
0	0.0227	0.1819
0	0	0.1819
-0.0043	0.0455	0
0.0014	0	0
0.0057	0	0.0909
0	0	0

mean (mm)*1.0e-12

0.0180	0.1316	0.2393

standard deviation (mm)*1.0e-11

0.0100	0.0194	0.1350

absolute mean (mm)*1.0e-12

0.0613	0.1316	0.8138

absolute standard deviation (mm)*1.0e-11

0.0080	0.0194	0.1089

Test 5.6 Generated points according to Perspective projection theory using XYZ coordinates from Test 5.4.

Similar to the Test 5.5, this test generates "perfect camera values" for the XYZ coordinates used in Test 5.4, where the calibration frame is 1 m away from the cameras.

Camera values generated by the software (mm)

Cam1	Cam2	Cam3	Cam4
10.20	19.84	22.07	27.27
10.21	15.66	25.06	24.34
10.22	12.20	27.57	21.92
8.02	9.08	27.98	18.08
3.09	10.07	23.52	15.21
-1.19	10.75	19.43	12.50
-3.70	16.71	13.30	14.89
-3.76	12.81	16.04	12.10
4.05	22.17	15.64	24.44
-0.70	20.84	12.69	19.99
8.31	14.97	17.96	20.64
8.31	11.87	20.15	18.45
5.29	18.24	13.79	20.90
2.22	17.81	12.14	18.49
-1.24	17.35	10.28	15.79
4.51	8.31	20.30	13.34
1.43	9.00	17.85	11.72
-2.54	12.29	12.97	11.31
-2.54	15.12	10.98	13.32

Camera parameters

50	50	50	50
2.54	3.39	2.54	2.54
250	-250	291	-291
907	907	1079	1079
-25	-25	-25	-25
0	1	-0.707	0.707
-1	0	-0.707	-0.707
0	0	0	0
0	0	0	0
0	0	0	0
1	1	1	1

the DLT algorithm was used for calibration and reconstruction of the generated data and the data report below includes: the distance to the principal point; the differences between the original and reconstructed coordinates; error evaluation for the entire data set; the original and the reconstructed values and the values of the L parameters..

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
2.5406	3.3895	2.5411	2.5388

Differences (mm)

X	Y	Z
0.0001	-0.0004	0.0006
-0.0001	-0.0001	0.0001
0.0001	0.0000	0.0000
0.0000	0.0001	-0.0006
-0.0004	0.0003	0.0008
0.0000	-0.0002	0.0007
0.0001	0.0000	-0.0007
0.0003	-0.0001	0.0007
0.0004	-0.0004	0.0012
0.0002	-0.0007	0.0001
0.0000	0.0003	0.0006
-0.0001	-0.0002	0.0016
0.0000	0.0001	0.0002
-0.0005	-0.0003	0.0008
-0.0002	0.0002	0.0031
0.0001	0.0000	-0.0002
-0.0001	0.0005	0.0003
0.0005	0.0002	-0.0008
0.0005	0.0006	0.0000

mean (mm)*1.0e-3

0.0487	-0.0037	0.4460

standard deviation (mm)*1.0e-3

0.2610	0.3117	0.8759

absolute mean (mm)*1.0e-3

0.1989	0.2490	0.6942

absolute standard deviation (mm)*1.0e-3

0.1760	0.1875	0.6959

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-1.0040	0.0720	0.7570	-1.0040
-0.0100	0.7570	-1.0030	-0.0100	0.7570	-1.0030
-0.0780	0.7570	-1.0010	-0.0780	0.7570	-1.0010
-0.1380	0.7990	-1.0100	-0.1380	0.7990	-1.0100
-0.1170	0.8960	-1.0210	-0.1170	0.8960	-1.0210
-0.1000	0.9830	-1.0440	-0.1000	0.9830	-1.0440
0.0230	1.0350	-1.0500	0.0230	1.0350	-1.0500
-0.0570	1.0360	-1.0490	-0.0570	1.0360	-1.0490
0.1240	0.8770	-1.0210	0.1240	0.8770	-1.0210
0.1050	0.9730	-1.0420	0.1050	0.9730	-1.0420
0.0490	0.7580	-1.3160	0.0490	0.7580	-1.3160
-0.0310	0.7580	-1.3160	-0.0310	0.7580	-1.3160
0.1280	0.8370	-1.2980	0.1280	0.8370	-1.2980
0.1130	0.9150	-1.2840	0.1130	0.9150	-1.2840
0.0970	1.0010	-1.2680	0.0970	1.0010	-1.2680
-0.1250	0.8570	-1.2950	-0.1250	0.8570	-1.2950
-0.1090	0.9350	-1.2810	-0.1090	0.9350	-1.2810
-0.0290	1.0330	-1.2660	-0.0290	1.0330	-1.2660
0.0410	1.0330	-1.2650	0.0410	1.0330	-1.2650

L values

0.00000400	-1.99925400	1.41413000	-1.41292700
2.00105600	0.00000300	1.41415100	1.41289200
0.10165200	0.13555900	0.10161800	0.10151000
-1812.41885900	-496.41971500	-1934.82808100	-1933.14942400
0.00000100	0.00000100	0.00000003	-0.00000040
0.00000100	0.00000020	0.00000100	-0.00000200
0.04002100	0.03998600	0.04000500	0.03996900

The Bundle Adjustment Method was used for reconstruction of the generated data and the data report includes: the differences between the original and the reconstructed coordinates; the mean; the standard deviation; the absolute mean and the absolute standard deviation, calculated from the differences.

Differences (mm)*1.0e-11

X	Y	Z
-0.0227	0	0.0909
0	0	0
0.0114	0	-0.1819
-0.0057	0.0455	-0.0909
-0.0028	0.0455	0.1364
0	0.0227	0.0455
0.0114	0	0
-0.0085	0.0455	0
-0.0227	0	0.0909
0.0114	0.0114	0
-0.0455	0.1364	0.3638
0	0.0909	0
0	0.0455	-0.1819
0.0227	-0.0227	-0.0909
0.0114	0.0227	0.0455
-0.0057	0	0.1819
0.0085	0.0114	0
0	0	0
0.0057	0.0227	-0.0455

mean (mm)*1.0e-12

-0.0165	0.2513	0.1915

standard deviation (mm)*1.0e-11

0.0155	0.0376	0.1253

absolute mean (mm)*1.0e-12

0.1032	0.2752	0.8138

absolute standard deviation (mm)*1.0e-12

0.1146	0.3581	0.9536

Appendix H

Calibration and Reconstruction of Targets from a Calibration Frame at Different Distances

The objective of the following two tests is to find out how the DLT Method performes on data which were not included in the calibration procedure. The generated data from one test was used for calibration, to determine the principal distance and the L parameters, and then the data from the other test with different XYZ positions and camera values was used for coordinate reconstruction.

Test 5.7 Calibration with generated data from Test 5.5 (3 m from the cameras) and reconstruction with the data from Test 5.6 (1 m from the cameras).

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
2.5376	3.3930	2.5447	2.5232

Differences (mm)

X	Y	Z
0.0378	-0.0610	-0.1027
0.0198	-0.0539	-0.0704
0.0178	-0.0533	-0.0754
0.0085	-0.0309	-0.0254
-0.0083	0.0000	0.0982
-0.0221	0.0083	-0.0705
0.0108	0.0004	-0.0283
0.0172	0.0034	0.0069
-0.0409	-0.0036	-0.0846
0.0189	0.0034	0.0276
0.0277	-0.0387	-0.1019
0.0152	-0.0337	-0.0684
-0.0011	-0.0122	0.0135
0.0235	0.0021	-0.0217
0.0021	0.0031	0.0059
0.0041	-0.0071	0.0179
-0.0039	0.0038	0.0867
0.0094	0.0047	0.0304
0.0032	0.0018	0.0024

mean (mm)

0.0073	-0.0139	-0.0189

standard deviation (mm)

0.0176	0.0227	0.0575

absolute mean (mm)

0.0154	0.0171	0.0494

absolute standard deviation (mm)

0.0112	0.0203	0.0351

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-1.0040	0.0720	0.7569	-1.0041
-0.0100	0.7570	-1.0030	-0.0100	0.7569	-1.0031
-0.0780	0.7570	-1.0010	-0.0780	0.7569	-1.0011
-0.1380	0.7990	-1.0100	-0.1380	0.7990	-1.0100
-0.1170	0.8960	-1.0210	-0.1170	0.8960	-1.0209
-0.1000	0.9830	-1.0440	-0.1000	0.9830	-1.0441
0.0230	1.0350	-1.0500	0.0230	1.0350	-1.0500
-0.0570	1.0360	-1.0490	-0.0570	1.0360	-1.0490
0.1240	0.8770	-1.0210	0.1240	0.8770	-1.0211
0.1050	0.9730	-1.0420	0.1050	0.9730	-1.0420
0.0490	0.7580	-1.3160	0.0490	0.7580	-1.3161
-0.0310	0.7580	-1.3160	-0.0310	0.7580	-1.3161
0.1280	0.8370	-1.2980	0.1280	0.8370	-1.2980
0.1130	0.9150	-1.2840	0.1130	0.9150	-1.2840
0.0970	1.0010	-1.2680	0.0970	1.0010	-1.2680
-0.1250	0.8570	-1.2950	-0.1250	0.8570	-1.2950
-0.1090	0.9350	-1.2810	-0.1090	0.9350	-1.2809
-0.0290	1.0330	-1.2660	-0.0290	1.0330	-1.2660
0.0410	1.0330	-1.2650	0.0410	1.0330	-1.2650

The parameters after the DLT calibration was completed are printed below.

L values

-0.000014	-1.999745	1.374570	-1.405644
2.034587	0.000015	1.374527	1.405584
0.103338	0.135591	0.098839	0.100992
-1842.841257	-496.527878	-1880.421885	-1923.208923
-0.000005	-0.000002	0.000005	0.000006
-0.000002	0.000001	-0.000002	-0.000013
0.040685	0.039995	0.038893	0.039763

Test 5.8 Calibration with generated pixel data from Test 5.6 (1 m from the cameras) and reconstruction with pixel data from Test 5.5 (3 m from the cameras)

This test was done in the same way changing the data sets for the calibration and the reconstruction procedure to check whether there would be a change in the accuracy. The data report follows the same order as for Test 5.7.

Distance to the principal point (mm)

Cam1	Cam2	Cam3	Cam4
2.5406	3.3895	2.5411	2.5388

Differences (mm)

X	Y	Z
-0.0016	0.0009	-0.016
0.0026	-0.0008	-0.0166
0.0007	0.0008	0.0056
-0.0026	0.0001	0.0265
-0.0009	0.0006	-0.0148
0.0009	-0.0005	0.0127
0.0001	0.0002	0.0197
-0.0006	0.0009	0.0151
0.0012	0.0009	0.009
-0.0002	-0.0013	0.0101
0.0012	-0.001	0.0173
-0.0028	-0.0007	0.0245
0.0001	-0.0015	0.0143
0.0017	-0.0013	-0.0008
0.0013	-0.0004	-0.0022
-0.0018	-0.0004	0.0002
0.0029	0.0005	0.0373
0.0044	0.0001	-0.0328
-0.0026	-0.0004	-0.0252

mean (mm)

0.0002	-0.0002	0.0044

standard deviation (mm)

0.0019	0.0008	0.0182

absolute mean (mm)

0.0016	0.0007	0.0158

absolute standard deviation (mm)

0.0011	0.0004	0.0100

The original X,Y,Z coordinates and the reconstructed values in (mm)*1.0e03

X	Y	Z	Xc	Yc	Zc
0.0720	0.7570	-2.7480	0.0720	0.7570	-2.7480
-0.0100	0.7570	-2.7470	-0.0100	0.7570	-2.7470
-0.0780	0.7570	-2.7450	-0.0780	0.7570	-2.7450
-0.1380	0.7990	-2.7540	-0.1380	0.7990	-2.7540
-0.1170	0.8960	-2.7650	-0.1170	0.8960	-2.7650
-0.1000	0.9830	-2.7880	-0.1000	0.9830	-2.7880
0.0230	1.0350	-2.7940	0.0230	1.0350	-2.7940
-0.0570	1.0360	-2.7930	-0.0570	1.0360	-2.7930
0.1240	0.8770	-2.7650	0.1240	0.8770	-2.7650
0.1050	0.9730	-2.7860	0.1050	0.9730	-2.7860
0.0490	0.7580	-3.0600	0.0490	0.7580	-3.0600
-0.0310	0.7580	-3.0600	-0.0310	0.7580	-3.0600
0.1280	0.8370	-3.0420	0.1280	0.8370	-3.0420
0.1130	0.9150	-3.0280	0.1130	0.9150	-3.0280
0.0970	1.0010	-3.0120	0.0970	1.0010	-3.0120
-0.1250	0.8570	-3.0390	-0.1250	0.8570	-3.0390
-0.1090	0.9350	-3.0250	-0.1090	0.9350	-3.0250
-0.0290	1.0330	-3.0100	-0.0290	1.0330	-3.0100
0.0410	1.0330	-3.0090	0.0410	1.0330	-3.0090

The parameters after the DLT calibration was completed are printed below.

L values

0.00000400	-1.99925400	1.41413000	-1.41292700
2.00105600	0.00000300	1.41415100	1.41289200
0.10165200	0.13555900	0.10161800	0.10151000
-1812.41885900	-496.41971500	-1934.82808100	-1933.14942400
0.00000100	0.00000100	0.00000003	-0.00000040
0.00000100	0.00000020	0.00000100	-0.00000200
0.04002100	0.03998600	0.04000500	0.03996900

Appendix I

Abstracts of the published papers

Three Dimensional Reconstruction with Linear CCD Cameras

Roumen Georgiev*, Patrick Gross+, Christopher Vaughan*

* Department of Biomedical Engineering, University of Cape Town, South Africa
+ Cambridge University, Cambridge, United Kingdom

Abstract: Computerized tracking systems based on rectangular CCD cameras currently exist and a variety of commercial systems are available. One example application is the location of a surgical instrument on a CT or MRI image during brain surgery. The present study is based on developing a system of linear CCD sensors for tracking passive markers for the needs of neurosurgery. First the spatial resolution of such sensors is better, since they consist of up to 4096 pixels and second the amount of data to be processed is many times less than with rectangular CCDs. Our experiments showed good linear resolution of 0.1mm using a 1D sensor. The signal processing before passing the data for reconstruction also includes extraction of peaks from the signal and determining the peak positions along the sensor. An algorithm to supress the background is suggested, together with a routine to separate peaks located close to each other. The measurements taken from passive targets illuminated by infrared (IR) light and active IR emitting diodes demonstrate that the passive targets are well detected by the system. A three dimensional (3D) reconstruction algorithm, based on the Direct Linear Transformation (DLT), calculated (X,Y,Z) coordinates of the target points. The absolute maximum error observed in this case was 2mm. We conclude that an improvement of the accuracy is necessary at this final step before it is possible to use the system in combination with a CT or MRI image.

Reconstructing Three Dimensional Coordinates from One Dimensional Sensors

Christopher Vaughan*, Roumen Georgiev*, Patrick Gross+

* Department of Biomedical Engineering, University of Cape Town, South Africa
+ Cambridge University, Cambridge, United Kingdom

Abstract: The most widely used method to track the 3D position of targets on the human body is based on two or more rectangular CCD cameras, and systems are commercially available (*e.g.* Oxford Metrics, Motion Analysis Corporation). Rectangular CCDs suffer from two main drawbacks: limited spatial resolution (typically 512x512 pixels) and, because a large number of pixels must be processed (over 250,000), limited spatial resolution. Linear CCDs, used in fax machines and photocopiers, have a much greater spatial resolution and far fewer pixels to process (typically 4096 pixels). The use of linear CCDs to reconstruct 3D position is not new (Macellari, 1983) and during the past decade commercial systems have become available (e.g. Northern Digital, Charnwood Dynamics). These systems use active sequentially-pulsed infrared targets and the instruments include just 3 linear CCDs. We introduce in this paper a system based on passive markers with 4 linear CCDs located on a single printed circuit board.

References

Abdel-Aziz Y I, Karara H M, "Direct Linear Transformation from comparator coordinates to object-space coordinates", *Close Range Photogrammetry*, American Society of Photogrammetry, Falls Church, Virginia, pp 1-18, 1971.

Alexander E, Kikinis R, Black P, Jolesz F, "Intraoperative magnetic resonance imaging, In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 215-227, 1998.

Allard P, Stokes I, Blanchi J, *Three Dimentional Analysis of Human Movement*, Human Kinetics, pp30-32,1995.

Barnett G H, Kaakaji W, "Intracranial meningiomas", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 87-100, 1998.

Barnett G H, Miller D J, "Brain biopsy and related procedures", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 181-191, 1998.

Barnett G H, Steiner C P, Roberts D W, "Surgical navigation system technologies", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 17-32, 1998.

Bar-Shalom Y, Chang K, Blom H, "Automatic track formation in clutter with recursive algorithm", In: *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, Massachusetts, pp 25-42, 1990.

Bianchi G, Gazzani F, Macellari V, "The COSTEL system for human motion measurement and analysis", SPIE vol.1356, First World Congress of Biomechanics, San Diego, California, pp 38-50, 1990.

Bopp H, Krauss H, " An Orientation and calibration method for non-topographic applications", *Photogrammetric Engineering and Remote Sensing*, vol.44, No 9, pp 1191-1196, 1978

Born M, Wolf E, *Principles of Optics*, Pergamont Press, New York, 1959.

Brown D C, "Decentering distortion of lenses", *Photogrammetric Engineering*, XXXII, 3, May, 1966.

Brown D C, *Advanced methods for the calibration of metric cameras.* Final Report, Part 1,U.S. Army Contract: DA-44-009-AMC-1457, (X), DBA Systems Inc., Florida. 117pages, 1968.

Brown D C, "Close-range camera calibration", *Photogrammetric Engineering*, 37(8), pp 855 –866, 1971.

Brown D C, "Calibration of close-range cameras", *International Archives of Photogrammetry*, 19(5): 26 pages, 1972

Bucholz R D, Smith K R, "A comparison of sonic digitizers versus light emitting diode-based localization", In: *Interactive Image-Guided Neurosurgery*, American Association of Neurological Surgeons, Charlottesville, Virginia, pp 179-200, 1993.

Crouch D, "OPTOTRAK – At last a system with resolution of ten microns", SPIE vol.1356, First World Congress of Biomechanics, San Diego, California, 1990.

Dainis A, "Linear direction sensor cameras for position measurement", US Patent, 1990.

Dufour F, Mariton M, "Passive sensor data fusion and maneuvering target tracking", In: *Multitarget-Multisensor Tracking: Applications and Advances*, Volume II, Artech House, Norwood, Massachusetts, 1990.

Freyer J G, "Camera calibration", In: *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, Caithness, Scotland, pp 156-179, 1996.

Georgiev R H, Gross P, Vaughan C L, "Three dimensional reconstruction with linear CCD cameras", VIII Mediterranean Conference on Medical and Biological Engineering and Computing, June 14-17, 1998, Lemesos, Cyprus.

Greaves J O, "Instrumentation in video-based three-dimensional systems", In: *Three Dimensional Analysis of Human Movement*, Human Kinetics, Champaign, Illinois, pp 41-55, 1995.

Gruen A, "Fundamentals of videogrammetry", International Summer School "Three Dimensional Analysis of Human Tracking", Grenoble, France, June, pp 1-24, 1996.

Gush K, "Software development calibration and optimisation of a 3D space marker detector system", Thesis Project, Department of Electrical Engineering, University of CapeTown, 1996.

Hager W W, *Applied numerical linear algebra*, Prentice Hall, Englewood Cliffs, New Jersey 07632, 1988.

Kalfas I, "Spinal surgery", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 117-134, 1998.

Karara H M, "Non-metric cameras", In: *Developments in close range photogrammetry*, Applied Science Publishers Ltd, London, pp 63-80, 1980.

Kurien T, "Issues in the design of practical multitarget tracking algorithms", In: *Multitarget-multisensor tracking: Advanced applications*, Artech House, Norwood, Massachusetts, pp 43-83, 1990.

Lavallee S, Bainville E, Bricault I, "An overview of computer-integrated surgery and therapy", In: *3D Imaging in Medicine*, Second Edition,CRC Press, Boca Raton, 2000.

Leo T, Macellari V, "Online microcomputer system for gait analysis data acquisition, based on commercially available optoelectronic devices", In *Biomechanics VII-B*, pp 163-169, 1981.

Leotta D F, Detmer P R, Martin R W, "Performance of a miniature magnetic position sensor for three-dimensional ultrasound imaging", *Ultrasound in Medicime and Biology*, vol.23, No 4, pp 597-609, 1997.

Litwiller D, "CCD vs CMOS: facts and fiction", *Photonics Spectra*, Lawrin Publishing Co.Inc. January, 2001.

Macellari V, "CoSTEL: A computer peripheral remote sensing device for 3-dimensional monitoring of human motion", *Medical and Biological Engineering and Computing*, 21, pp 311-318, 1983.

Maciunas R J, "Yesterday's tomorrow: the clinical relevance of interactive image-guided surgery", In: *Interactive Image-Guided Neurosurgery*, American Association of Neurological Surgeons, Charlottesville, Virginia,1993.

Maciunas R J, "Pitfalls", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 43-60, 1998.

Magill A, "Variation of distortion with magnification", *Journal of Research of the National Bureau of Standards*, Vol.54, No 3, pp 135-142, 1955.

Marzan G T, Karara H M, "Rational Design for Close Range Photogrammetry", Report UILU-ENG-76-2001, pp 67-72, Department of Civil Engineering, University of Illinois, Urbana, Illinois, 1976.

Mitchelson D L, *Techniques for analysis of human movement*, pp 59-64, Lepus Books, London, 1975.

Murray D, Basu A, "Motion tracking with an active camera", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.16,No 5, May, pp 449-459, 1994.

Owen C P, "An improved camera and marker system for real-time three-dimensional tracking", Thesis Project, Department of Biomedical Engineering, University of Virginia, 1995.

Pedrotti F L, Pedrotti L S, *Introduction to Optics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Rabiner L R, Gold B, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

Reymond J C, Hidalgo J L, "Arrangement for locating radiating sources" US Patent, 1980.

Rogus R D, Stern R L, Kubo H D, "Accuracy of a photogrammetry-based patient positioning and monitoring system for radiation therapy", *Medical Physics*, Vol.26, No 5, May, pp 721-728, 1999.

Shih S W, Hung Y P, Lin W S, "Calibration of an active binocular head", *IEEE Trans. on Systems, Man and Cybernetics – Part A: Systems and Humans*, Vol.28, No 4, July, pp 426-442, 1998.

Shortis M R, Beyer H A, "Sensor technology for digital photogrammetry and machine vision", In: *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, Caithness, Scotland, pp 106-155, 1996.

Sturman D J, "A Brief History of Motion Capture for Computer Character Animation", 1997, http://www.css.tayloru.edu/instrmat/graphics/hypgraph/animation/motion_capture/history1.htm

Vaughan C L, Davis B L, O'Connor J C, "Integration of antropometry, displacements, and ground reaction forces", *Dynamics of Human Gait*, Human Kinetics Publishers, Box 5076, Champaign, Illinois, 1992.

Vaughan C L, Georgiev R H, Gross P, "Reconstructing three dimensional coordinates from one dimensional sensors", 5th International Symposium on the 3-D Analysis of Human Movement, July 2-5, 1998, Chattanooga, Tennessee, USA

Whittle M W, "Methods of gait analysis", *Gait Analysis,* Butterworth-Heinemann Ltd, Halley Cort, Jordan Hill, Oxford OX2 8EJ, 1991.

Zamorano L, Lis-Planells M, Zhaowei J, Nolte L, Kadi A, Diaz F, "Vascular malformations", In: *Image Guided Neurosurgery: Clinical Applications of Surgical Navigation*, Quality Medical Publishing Inc, St. Louis, Missouri, pp 105-116, 1998.