A Fast Webcam Photogrammetric System to support Optical Imaging of Brain Activity

Sen Wong

Department of Civil, Environmental and Geomatic Engineering
University College London (UCL)

Supervisors:
Professor Stuart Robson
Dr. Adam P. Gibson

Thesis submitted for the Degree of Doctor of Philosophy (Ph.D.)
February, 2012

Declaration of authorship

I, Sen Wong, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature……

Date……..

Abstract

Optical topography (OT) is an emerging neuroimaging technique utilising the tight coupling between neural activity and regional cerebral blood flow to monitor relative regional changes of haemoglobin concentration. OT systems are compact, low cost, easily portable, and relatively tolerant of body movements enabling clinical diagnosis, psychological experiments and even monitoring brain activities during daily living. There is a requirement for such systems to present their output functional data in a brain model based coordinate space in order to map to the signal source with brain anatomy. However OT data are obtained from a network of OT sensing devices (optodes) placed in contact with the head surface and cannot capture structural information of the underlying brain which might otherwise be used for registration. An appropriate registration method, widely used in Electroencephalography (EEG), is the 10-20 system which utilises bony landmarks as common points to co-register locations on the scalp with a brain model to a repeatability of a few millimetres in clinical applications to an MRI set of reference points. Inheriting the low cost and portability of OT, this thesis develops and validates a novel registration approach utilising off-the-shelf webcam technology in combination with photogrammetric bundle adjustment techniques in order to reliably coordinate targets on optodes and bony landmarks within the 10:20 reference frame to an accuracy of better than 1mm.

Initial research includes an assessment of the 3D coordination accuracy, precision and stability of a series of low cost webcams in order to prove their suitability for clinical applications. Results demonstrate the capability of a system based on these cameras to reliably coordinate 3D target locations to the order of 0.5mm and better. Difficulties in automated clinical target image extraction due to poor image quality are circumvented through the development of new target image detection methods. Incremental improvements in image quality from successive webcam generations, up to and including the latest HD systems, are shown to increase coordination accuracy by one order of magnitude. The result is a novel webcam photogrammetric system that is able to rapidly and consistently coordinate targets on optodes and bony landmarks to better than 1mm in OT studies and is able to take advantage of the rapid advances being

made in consumer webcam technology. The system is proven in pre-clinical studies to evaluate its coordination accuracy and in simulated clinical OT studies with a head-sized phantom conducted in collaboration with Department of Medical Physics and Bioengineering. Clinical OT studies with human subjects, demonstrate the capability of the system to continuously coordinate targets on optodes and scalp and detect differential movement between optodes and scalp which would invalidate a static registration procedure.

Acknowledgement

First I would like to thank my principal supervisor, Professor Stuart Robson, and my secondary supervisor, Dr. Adam Gibson for their advice and guidance through my PhD research as well as reading my thesis. I would also like to thank Professor Jeremy Hebden for his invaluable advice at the beginning of this research. My special thanks go to Stuart for his advice and assistance for me to develop the prototype of multi-webcam fast photogrammetric tracking system, as well as his great support during my difficult time. His influence of my personal development during the past four years is invaluable for the rest of my life!

I would also like to thank my colleagues from Department of Civil, Environmental and Geomatic Engineering and Department of Medical Physics and Bioengineering. Thanks Dr. Anna Blasi for her help with my visits to the Baby Lab at Birkbeck College of London and comprehensive discussions. Dr. Rob Cooper's help with collecting the optical topography data for studies with both the phantom and the volunteers is greatly appreciated. Thanks also go to Adam and Rob for their participation as volunteers for optical topography studies as well as the optical image reconstruction by Adam. In terms of the system development, I appreciate the help from Mr. Ian Seaton with the infrastructure for building the tracking system and landmarks on the phantom. Thanks also go to Mr. Les Erwin for his help with disassembly of webcams and the preparation of ball bearings for experiments. I would like to express my special thanks to Les for his participation with his head for my tests. Without your help, I would not have been able to complete the experiments and collect the useful data.

Besides the people involved in my research work, I would like to thank Ziyi and Seong for their support and willingness to share during the past three years. Thanks also go to your wives for treating me great meals. I would also like to express my appreciation to my friends in both London and Hong Kong for their encouragement for writing this thesis.

This thesis is dedicated to my parents and my uncle for their financial support as well as inspiration and support in every aspect towards the completion of this thesis.

Contents

List of figures

List of tables

List of flowcharts

Chapter 1 Introduction

1.1 Introduction

This chapter describes the aims of this research project and outlines its key aspects. The work described in this thesis has two motivations. Can increasingly capable consumer webcams be used for accurate, but low cost measurement and secondly can a system comprised of such cameras contribute to a medical physics application where an improvement in the spatial mapping of optical topography (OT) sensing systems is required in order to spatially map haemoglobin changes to cerebral activity. The research carried out involves designing experiments to: validate webcam photogrammetric calibration; to design and implement new algorithms in order to compensate for poor photogrammetric target image quality and; the implementation of a robust rapid multi-camera tracking system which can accommodate variations in lighting and background clutter to spatially reference optical topography sensing pads on the human head. Based on a successful project from Mauren Abreu de Souza [Souza, 2009], the proposed work was initiated following discussions between the Geomatic Research Laboratory and the Biomedical Optics Research Laboratory of UCL in 2007.

1.2 Background and motivation

Neuroimaging techniques to present brain activation data under a consistent coordinate system have received a lot of attention in recent years [Mazziotta *et al.*, 2000, 2001a,b; Toga and Thompson, 2001]. The primary goal is to enhance the comparison between intra- and inter- functional studies. For tomographic brain mapping techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), it has been very common to present functional data under either Talairach [Talairach and Tournoux, 1988] or MNI [Collins *et al.*, 1994] brain template. For transcranial brain mapping techniques, however, such registration methods are still at an early stage [Okamoto *et al*., 2004b; Okamoto and Dan, 2005].

Optical topography is an evolving non-invasive brain mapping technique which has attracted a large amount of attention recently. Compared with many other mapping techniques, optical topography offers much more flexibility in which functional studies

can be performed at patients' bedsides and more flexible experimental protocols can be designed [Gibson *et al.*, 2005]. However, similar to other transcranial mapping methods, the functional data obtained from the scalp surface lacks the corresponding structural information. To comply with the common practice for tomographic mapping techniques, the first step is to register the functional data from the scalp surface onto the cortical surface in order to identify where the activations are generated from an individual brain anatomy.

Since the functional data is obtained from source-detector fibre bundle (optodes), in order to accurately register the functional data from the scalp surface to the cortical surface, accurate determination of optodes' positions is essential. However, optodes' positions are commonly measured manually with a 3D digitizer arm by taking the average of three measurements for each optode. It is inevitable that the measurement accuracy is significantly lowered by this manual process and therefore the accuracy of optodes' positions is no better than a few millimetres [Souza, 2009]. Considering possible mis-registration between functional data and structural data caused by the positional inaccuracy of optodes, it is necessary to increase the positioning accuracy of optodes for optical topography studies, preferably with an automated measurement method.

Moreover, Gibson *et al.* (2005) reported that light leakage resulted from differential movement between the scalp surface and the optodes has been a recurring problem for optical imaging studies. Artefacts in the reconstructed optical images can easily be resulted by this effect. Conventionally tedious manual post-processing of optical signal data is required to identify and eliminate this effect from image reconstruction. The differential movement is likely to cause light leakage frin sources, instability of detectors or even worse, the failure to collect useful data. For optical imaging studies, the only changeable variable allowed is the optical property representing brain activities, but not the scalp location in contact with an optode [Gibson *et al.*, 2005].

Photogrammetrists have directed research efforts towards medical measurement since the earliest days of photogrammetry, mainly because of the benefits that photogrammetry can offer to humanity as a painless and non-invasive means of

providing medical practitioners with spatial measurement relating to the human body. Various techniques of body measurements (examples given in Section 3.6) have been developed in response to the demands for specialized spatial measurement tools for a wide variety of medical ends [Mitchell and Newton, 2002]. While Ahmad *et al.* (2004) discussed the experience of calibrating a consumer digital camera using low-cost digital close range photogrammetric software, the results showed that all the camera calibration parameters could be recovered without difficulty and could be used to capture images suited to medical photogrammetry applications. Ahmad *et al.* (2004) suggested from their study that consumer digital cameras have the potential to be used in many close range photogrammetric applications, particularly when the budget is limited.

Consumer digital cameras are gaining in popularity and capability with increasing numbers of pixels but reducing price. The speed, convenience and repeatability of digital image access together with rapid image processing development by the computer vision community, have attracted a lot of attention on the potential applications of photogrammetry. There are three advantages of digital cameras over the traditional optical cameras. First, digital cameras allow faster and repeatable image access; second, there is no need for expensive and specific device and trained professional technicians; third, the result of measurement is in form of digital data, which simplifies processing and offers the potential for real time data analysis. However, low cost digital cameras intended for use as webcams compromise on lens quality and stability when compared to more expensive digital SLR models typically used for photogrammetric measurement. The challenge is to deliver consistent photogrammetric measurement accuracy with the much poorer image quality of low cost digital cameras under natural lighting in environments where there is no control of lighting or image content.

1.3 Research aims

Motivated by the status (both optical imaging and photogrammetry) at the commencement of this research, the research aims to develop and validate a novel photogrammetric registration approach between a set of optical sensing optodes and brain anatomy utilising off-the-shelf webcam technology. The primary aims are listed as follows:

- To assess the suitability of low cost webcams for photogrammetric measurement;
- To develop target image detection methods suited to relatively poor webcam image quality for effective feature extraction against cluttered backgrounds;
- To accurately coordinate targets over a spherical head;
- To reliably coordinate targets on optodes and bony landmarks on a close to real time basis;
- To detect differential movement between optodes and scalp during optical studies.

1.4 Research objectives

The research objectives are formulated as follows:

1. **Effective camera system calibration**

   3D measurements require at least two imaging devices. Accurate 3D measurements rely on precise recovery of camera parameters from a calibration process. Since individual camera calibration may represent overhead for each optical imaging study, an efficient system calibration taking all available imaging devices into account is considered to be necessary for clinical work;

2. **Verification of webcam stability**

   Consistent coordination accuracy is required for continuous positioning of optodes during optical studies. However, the stability of low cost imaging

devices such as webcams remains unknown due to the compromise between their built quality and their price range. Regarding this the variation of camera parameters over different time intervals as well as the associated coordination accuracy needs to be verified to ascertain that the coordination accuracies over time are consistent and meanwhile within the accuracy requirement;

3. **Minimisation of the infrastrature and cost for building a camera system**

   To inherit the advantage of low cost from optical topography, the goal of building a camera system for future optical imaging studies is with a total cost under £500. Meanwhile, considering the flexibility and portability of optical topography, the infrastrature to construct such an imaging system needs to be light but rigid enough to ascertain the stable imaging geometry for 3D measurements;

4. **Target positioning accuracy**

   To increase the positioning accuracy of optodes previously measured with a digitizer arm, the goal of this research is to coordinate targets better than 1mm accuracy in 3D space. The 1mm positioning accuracy is considered to be appropriate for optical topography. This is because the estimation error introduced from the subsequent registration of optodes' positions can be minimised by increasing the initial surface measurement accuracy. To achieve this, accurate target image extraction is required from webcam imageries, under test scenes containing background clutters without lighting control;

5. **Fast and automated target coordination**

   A fully automated 3D positioning of optodes is necessary to eliminate the measurement error introduced from human interventions. This demands automation for all processes including target image detection, target correspondence matching and 3D coordinate computation. The automation of target image measurements demands effective method(s) to eliminate background noises and meanwhile accurate recognition of true targets, in order to minimise the target ambiguities for correspondence matching. In terms of speed, the goal of the coordination frequency is ~1Hz so that any differential

movement leading to image artefacts can be detected immediately during the study. This level of frequency should also allow the reconstruction of target movement for any selected period of time during an optical study.

1.5 Research methodologies

Research methodologies can be formulated based on three main chronological experimental works for this research:

1. Assessment of webcam photogrammetric performance

- Two camera models representing the mainstream low cost webcams, Logitech QuickCam Pro 4000 and Logitech C500, are selected for this research. Comparisons can be made between the two generations of webcams to ascertain the photogrammetric improvement in webcam technology during this three year project;
- A 6MP Nikon D100 camera fitted with a fixed focus 28mm Nikon lens is used as an established comparative standard for the assessment of accuracy;
- For effective calibration method, results from individual webcam calibrations can be compared with those from a self-calibrating bundle adjustment of all webcams;
- 2D and 3D target coordinate uncertainties as well as 3D target coordinate accuracy between data sets can be compared to derive the setting of image resolution with the highest precision and accuracy.

2. Development of the fast tracking system

- Development of the multi-webcam tracking system should take convergent network geometry and relative webcam positions into account to allow the key positions (optodes) to be intersected by three or more camera rays;
- Difficulties in terms of feature extraction for webcam imageries should be identified;
- Relationship between USB connection and image qualities for webcams should also be tested;
- Tests with a variety of scenes is necessary to derive efficient target image detection methods for webcam images suited to optical imaging studies;
- Edge patterns of detected target images and the associated 3D coordinates (of the

same target) from different detection methods should be compared to address the characteristics of each method.

3. Applications of the camera system

- To coordinate optodes and bony landmarks, experiments can start with a minimum of two webcams. Further studies are required to derive the number of webcams for sufficient target coverage and coordination accuracy;

- Both absolute (independent of photogrammetric comparison) and indirect photogrammetric assessment of positioning accuracy can be used to verify the coordination accuracy of the camera system;

- The phantom optical imaging studies can be used to assess the 3D coordination performance of the camera system, while the volunteer studies can be used to assess the capability of the system in terms of differential movement detection;

- The 10-20 system bony landmarks can be utilised as control points to establish a common 3D relationship between the optical sensing pad (optodes) and internal brain structure;

- Different approaches should be used to co-register the positions of optodes with the internal heating locations, based on the available structural information.

1.6 Research tools

A set of tools have been utilised throughout this research. Investigation of the suitability of webcams in medical studies including image measurement and reference data processing were implemented through the in-house photogrammetric tool VMS 8.0. The fast multi-camera tracking system was developed and tested in Microsoft's Visual Studio 2005 Professional in C/C++ and a series of open source tools including Microsoft DirectShow, OpenCV, OpenGL and VideoInput (Appendix 1, 2, 3) were also utilised. For the assessment of the developed tracking system, a 3D similarity transform was implemented in Visual Studio whilst the visualisation of the tracked targets and the pattern of their coordinate discrepancies across video frames were performed in the Mathworks Matlab environment [Mathworks, 2007].

1.7 Thesis overview

This thesis is composed of six main chapters supported by references and appendices. Chapter 2 forms the core medical background whilst Chapter 3 provides the core photogrammetric background to this research work, Chapter 4 discusses the development of the proposed fast multi-camera tracking system and its associated target detection methodology for the purposes of this research work, Chapter 5 outputs and analyses the results from both laboratory and optical imaging studies and finally Chapter 6 concludes this research work and discusses directions for future research.

Chapter 2

This chapter comprises the medical background of this research. Fundamentals of tissue optics in Near-Infrared Spectroscopy and the working principle as well as applications of optical topography are first described, followed by the overview of atlases in neuroimaging community which consists of image registration, overview of functionality of human cerebral cortex, the international 10-20 system used in Electroencephalography and the normalization methods in brain mapping. The last section describes state of arts concerning registration techniques of functional Near-Infrared Spectroscopy from the scalp surface into the cortical surface.

Chapter 3

This chapter reviews the photogrammetric and imaging principles. Fundamental photogrammetric principles are first described, followed by comparisons between conventional high cost cameras and webcams from different perspectives and the associated calibration methods adopted for webcams in this research. Target properties and lighting suited to clinical use are then discussed, followed by the fundamentals of automated target image measurement suited to webcam photogrammetry. Following the image based measurement is the 3D correspondence solution with its optimisation in favour of a fast coordination application. This chapter finally considers the development of a fast multi-webcam photogrammetric coordination system.

Chapter 4

This chapter describes the research work. Accuracy and precision assessment of the off-the-shelf low cost webcams is first reported, followed by a discussion comparing webcam technology from different generations and outlining factors giving rise to target measurement inaccuracy. Self-calibrating bundle adjustment experiments with a larger calibration object are described and results are reported based on the findings. Studies on webcam stability are then reported confirming the consistency of target coordination over time. Research work towards a fast tracking system with low cost webcams, which is composed of the design of the tracking system, proposed target detection method and comprehensive comparison of target detection result, is then discussed.

Chapter 5

This chapter presents the validation of the developed fast multi-webcam tracking system in a series of chronological experiments. First, an investigation on how many cameras are sufficient in terms of tracking targets located on the scalp and on optical topographic pads is given. Taking these minimum configurations into account, studies with a polystyrene head, OT sensing pad and targeted calipers provide a proof of concept together with an accuracy assessment. Drawing upon conclusions from previous studies, optical imaging studies involving both a phantom head and volunteer human subjects are reported, where both the accuracy assessment of the tracking system and differential movement between the optical sensing pad and head surface during the study are reported in detail.

Chapter 6

The final chapter of this thesis first summarises the research objectives and activities. Research data analysis and conclusions are then presented, followed by suggestions of future research directions. A final remark is presented at the end of this chapter to conclude the entire thesis.

Chapter 2 Optical Imaging and Brain Mapping

2.1. Optical Topography

2.1.1 Overview of Near-Infrared Spectroscopy (NIRS)

The absorption of water and haemoglobin is relatively low in the near-infrared region of the spectrum, 650–1050 nm, corresponding to the visible and mid-infrared regions. This enables the light to travel deep into most tissues and therefore makes NIRS a suitable tool for *in vivo* measurements.

A NIR imaging system is able to reveal information about the functional and metabolic activity of tissues by providing maps of oxygenation status of tissue within a living brain or other part of the body [Obrig and Villringer, 2003]. An advantage of using NIR light to create an imaging system is that it is non-invasive, low cost and portable, and so the data acquisition process can be performed repeatedly at the bedside. This makes NIR a useful tool to monitor the progression of conditions and the response to treatment.

2.1.2 Fundamentals of tissue optics

2.1.2.1 Optical properties

The transmission of light (photons) through tissues can be modelled by both light absorption and light scattering, where they affect the light intensity transmitted through tissues.

2.1.2.1.1 Light absorption

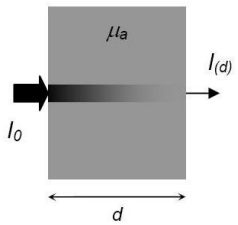


Figure 2 - 1. *Decrease of light intensity when light passes through a medium where only absorption occurs.*

If a light source with intensity I_0 and wavelength λ passes through a non-scattering medium, the intensity of the transmitted light $I_{(d)}$ will be:

$$I_{(d)} = I_0 e^{-\mu_a(\lambda)d} \qquad 2 - 1$$

where $\mu_a(\lambda)$ is the absorption coefficient of the medium and d is the width of the medium as illustrated in Figure 2 - 1. The aborption coefficient represents the probability in which a photon is absorbed per unit length and may be due to a number of absorbing substances (chromophores) mixed together. The individual extinction coefficients of each chromophore represent their absorption at a particular concentration. Therefore for a mixture of chromophores its absorption coefficient can be expressed as the sum of the products of chromophore concentrations c_n with their respective extinction coefficients ε_n:

$$\mu_a(\lambda) = \sum_n \varepsilon_n(\lambda) c_n \qquad 2 - 2$$

2.1.2.1.2 Light scattering

In a non-scattering medium, the travelling path of light is straight. However, this is no longer the case in a scattering medium, where the light photons collides with the particles in the medium and change their paths. The scattering coefficient represents the probability in which a photon changes its travelling path per unit length.

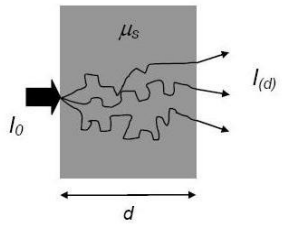


Figure 2 - 2. *Decrease of light intensity when light passes through a medium where only scattering occurs.*

Although there is no energy loss, elastic scattering can still cause attenuation of a light beam by changing the initial path of photons. For a light source of intensity I_0 passing through a scattering medium, the intensity of transmitted light I(d) representing the non-scattered component can be expressed as:

$$I_{(d)} = I_{(0)} e^{(-\mu_s \cdot d)} \qquad 2 - 3$$

where μ_s is the scattering coefficient of the medium and d is the width of the medium as illustrated in Figure 2 - 2.

2.1.2.1.3 Anisotropy

In order to correctly define the directions of light scattering in tissue, the probability of a photon being scattered in a given direction by a particular scatterer must be considered. When light scatters off a particular scatterer, the light generally emerges in a preferential direction, relative to its incident angle as shown in Figure 2 - 3:

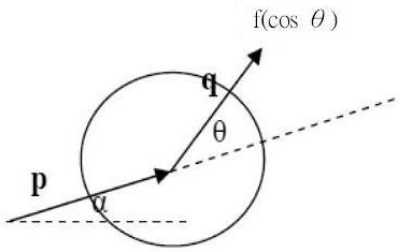


Figure 2 - 3. *The scattering phase function f(cos θ).*

Such a preferential direction is dependent on a number of factors including the light wavelength, the size of the particle it collides with as well as the refractive indices of different media the photon is travelling through. The phase function can be represented as a function of the cosine of the scattering angle θ between the incident light *p* and the scattered light *q* under the assumption that the scatterer is random and non-structured, as in the case of biological tissue. In general, the new direction does not occur with equal probability i.e. anisotropic scattering. The degree of the anisotropy can be characterized by the mean cosine of the scattering angle θ, commonly refer to the anisotropy factor g, which provides a measure of the mean direction of the scattering:

$$g = \int_{4\pi} p(\theta) \cdot \cos(\theta) dq \qquad 2 - 4$$

where g=0 if scattering is equally distributed over all angles, i.e. perfectly isotropic scattering.

The scattered intensity distribution increases in the forward direction as the particle size increases, i.e. a small angle in θ will result in g moving towards unity, which indicates a more forward-peaked scattering of the incident wave. Although it is a fact that photons predominantly scatter in a forward direction in biological tissue, typical values of scattering coefficient ensure that photons travelling through a few millimeters of tissue

lose all of their original directionality and become isotropically distributed. Therefore, it is often appropriate to assume isotropic scattering where the scattering coefficient has been reduced by the factor (1-g):

$$\mu_s' = \mu_s(1-g) \qquad 2-5$$

where μ_s' represents the transport scattering coefficient or effective number of isotropic scatters per unit length.

The attenuation of light as a result of absorption and scattering effects can then be described as a function of μ_a and μ_s'. The diffusion equation is a good mathematical model to describe the photon distribution since light propagates through layers of highly scattering tissue and therefore it is possible to utilize this equation to calculate the expected total attenuation due to μ_a and μ_s' [Arridge *et al.*, 1993]. In the NIR region, typical values for μ_a and μ_s' and in most tissues lie in the range of 0.005-0.02 mm^{-1} and 0.5-10 mm^{-1} respectively [Cheong *et al.*, 1990].

2.1.2.2 Absorption by tissue chromophores

Chromophores are tissue compounds with their own absorption spectra at different wavelengths. The absorption spectra of some common chromophores are shown in Figure 2 - 4. In the near-infrared range (700-900nm), water and haemoglobin are the two main chromophores in soft tissues. Niemz (1999) reported that photons travel in diffuse directions since the effect of scattering is much larger than that of absorption within the near-infrared region (grey area in Figure 2 - 4).

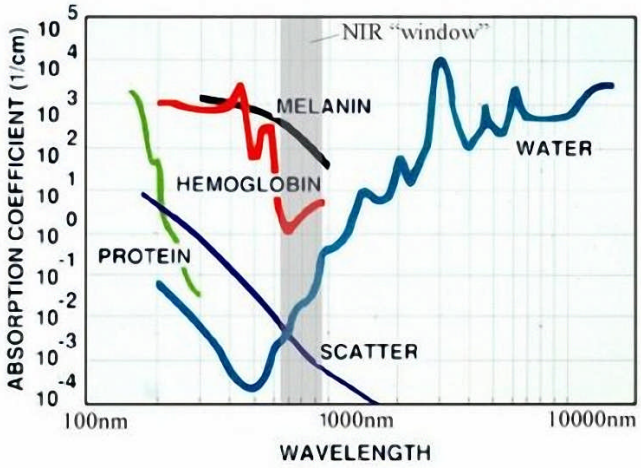


Figure 2 - 4. *Overview of the optical properties of tissues. (reproduced from Hillman (2002))*

While the concentration of water and melanin remains almost constant, the concentrations of oxygenated and deoxygenated haemoglobin, which vary with time, provide clinically useful physiological information.

2.1.2.2.1 Haemoglobin

The typical value for haemoglobin concentration in adult brain tissue is about 84 μmolar [Cope, 1991]. Oxygen molecules bind to the iron atoms in the haemoglobin to form oxy-haemoglobin (HbO_2) and deoxy-haemoglobin (HHb) is produced when the oxygen molecules separate from the iron atom. Both HbO_2 and HHb makes up the total haemoglobin (HbT), which is approximately proportional to the blood volume. Figure 2 - 5 shows the absorption spectra of the two chromophores in the NIR region.

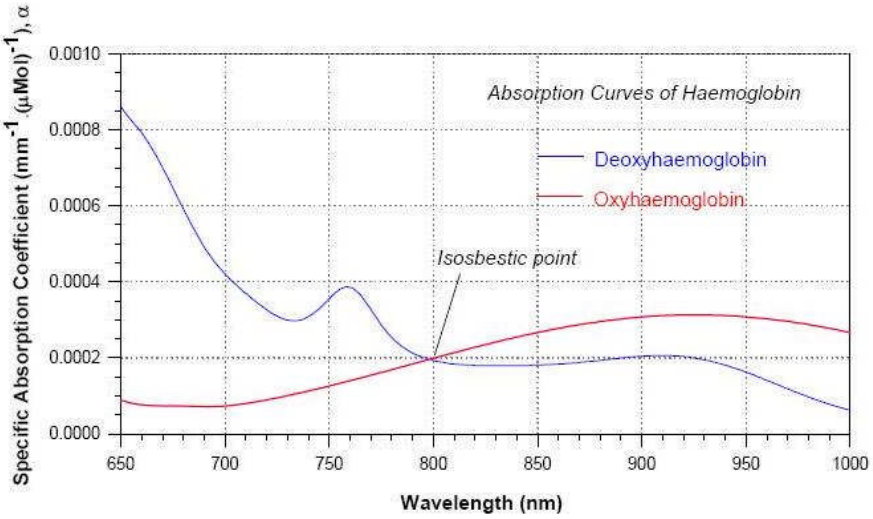


Figure 2 - 5. *Absorption spectra for HbO2 and HHb. [Cope, 1991]*

From the above figure, the absorption spectra of HbO_2 and HHb are different particularly in the NIR region (700-900 nm), with the absorptions crossing at about 800 nm [Cope, 1991]. Making simultaneous measurements at a minimum of two wavelengths allows HbO_2 and HHb to be monitored separately.

2.1.3 Basics of Optical Topography

2.1.3.1 How Optical Topography works

2.1.3.1.1 Configuration

Optical topography is a functional Near-Infrared Spectroscopy (fNIS) technique which is capable of measuring optical changes occurred below the tissue surface. The measurement is made, either simultaneously or serially, through optical fibre bundle (optodes) in contact with tissue surface, where each optode contains a pair of source and detector. The smaller the distance between the source and detector, the higher the measured signal can be obtained. Hebden (2003) reported that fast data acquisitions between source-detector pair can be achieved in this way to a frequency of 10Hz.

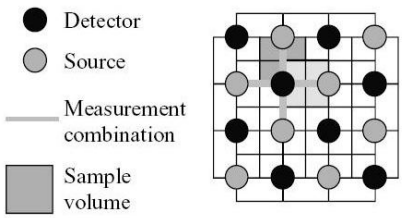


Figure 2 - 6. *Typical arrangement for arrays of sources and detectors for optical topography. [Yamashita et al., 1996]*

Figure 2 - 6 shows a typical source-detector arrangement for topography of the brain's cortex. The grid array allows measurement of changes of cortex close to the surface of the tissue if close source-detector pairs are chosen.

2.1.3.1.2 Data acquisition

Data acquisition of optical topography can be either serial or parallel. While serial acquisition uses all available detectors for one source at a time, parallel acquisition utilises all sources modulated with different frequencies. Everdell *et al.* (2005) developed an efficient method which demultiplexes the parallel sources by Fourier transform, while Koizumi *et al.* (2003) used lock-in amplifiers.

2.1.3.1.3 Image reconstruction of Optical Topography

By making multiple NIRS measurement simultaneously while ensuring that the measurement areas overlap spatially, optical topography produces two-dimensional maps of haemoglobin changes from specific regions of the cerebral cortex and adopts linear image reconstruction.

Arridge *et al* (1993) proposed the finite element mesh as the numerical method to model the complex geometries for optical imaging. Arridge and Schweiger (1995), Arridge (1999) and Arridge *et al* (2000) explained this method in detail and subsequently refine this modelling technique based on a finite element mesh built for the purpose of image reconstruction. The forward problem can be formulated as

$$y = F(x) \qquad 2 - 6$$

The forward problem involves in calculating simulated data y from the optical properties x where the absorption and the scattering coefficients may be included [Gibson *et al.*, 2005]. Arridge (1999) proposed a novel approach by solving the inverse problem of equation 2-6 in order to reconstruct an image representing the change of optical properties. The inverse problem is formulated as:

$$x = F^{-1}(y) \qquad 2 - 7$$

Since in optical topography studies, two optical measurements are acquired, one before and one after a small change of optical properties, such data acquisition method enables the linearization of the non-linear inverse problem, where the measurements before the optical change can be effectively used as the approximated values for Taylor series expansion. However, the linearization can be unreliable if the optical changes are relatively large. Other error contributors include the selection of regularization parameter [Gibson *et al.*, 2005] and the quantization error from the assumption of uniform optical properties (which is not in practice) [Hebden *et al.*, 2008].

2.1.3.2 UCL Optical Topography System

The UCL optical topography system [Everdell *et al.,* 2005] currently consists of 16 laser diode sources (8 at 785 nm and 8 at 850 nm) and eight avalanche photodiode detectors. The source and detector optodes are arranged to record simultaneously from 30

channels (source–detector pairs). The depth from the surface interrogated is approximately proportional to the source detector spacing. A 1 mm diameter multimode optical fibre is used to couple each source and detector to the scalp. Parallel data acquisition is used by modulating sources with frequencies ranged from 2 kHz to 4 kHz, followed by demultiplexing signals received from detectors with Fourier transform at a frequency of 10Hz.

Figure 2 - 7. *UCL Optical Topography System.*

2.1.3.3 Applications of Optical Topography

Optical topography creates a map of properties which demonstrates whether changes are occurring in areas where activation is expected, by simultaneously monitoring areas where no variation should be seen, as indicated by Figure 2 - 8:

Figure 2 - 8. *The topographical map for total haemoglobin seen over the left central sulcus (dotted line) for left and right finger tapping. [Yamashita et al., 1996]*

The application of optical topography includes detections of functional activation on cortical surface representing brain cognition (Table 2 - 1), as well as investigations of blood supply on superficial tissues such as forearm muscles [Vaithianathan *et al.,* 2004]. However, problems exist with the technique of optical topography. This includes the calibration of individual detectors and variations in the coupling coefficients for each

source and detector due to bad contact and hair on the head. Complete light scattering resulting in no data often occurs especially for adults with thicker skulls as compared with babies. Due to the nature of scattered light, the spatial resolution of optical imaging usually decreases with the depth below the surface. A cylinder with 8mm diameter and 8mm height within a head-sized phantom (10mm below the surface) was reported as an approximate spatial resolution for current optical imaging [Hebden *et al.*, 2008].

2.2 Atlas

2.2.1 Overview of image registration

The movement of the subject during magnetic resonance (MR) imaging is one of the fundamental issues in medical imaging research. Image registration is a technique developed in order to register serially acquired MR images with each other, through seeking the corresponding points across images and subsequently deriving the appropriate transformation. Detailed review of image registration techniques can be found in [Brown, 1992; Maintz, 1998]. Applications of image registration can be found in many areas, such as motion correction in functional MR imaging [Jiang *et al.*, 1995].

2.2.2 Human brain anatomy

2.2.2.1 Cerebral cortex

The cerebral cortex constitutes a superficial layer of grey matter and internally the white matter, as shown in Figure 2 - 9. A high proportion of nerve bodies exist in the grey matter, while white matter is composed of nerve fibres, or axons, transmitting signals between the nerve cell nuclei that compose the grey matter. The axons scatter light so that the white matter appears white.

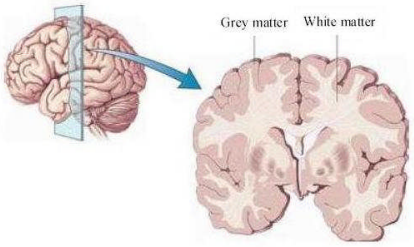


Figure 2 - 9. *Illustration of cross section of the adult brain. (reproduced from Crossman (1995))*

2.2.2.2 Division of brain lobes

During brain evolution, the volume of the cerebral cortex increased much faster than the cranial volume resulting in convolution of the surface and the folding of the total structure of the cortex. The grooves (sulci) are further separated by their magnitudes of folding; the more elevated gyri (bumps) are separated by the deep sulci and these landmarks are used to divide the cerebral cortex. Despite the same fundamental features of gyri/sulci, there exists a large variation of their precise locations from one person to another. Figure 2 - 10 shows four major lobes. The frontal lobe is located at the anterior part of the brain. The parietal lobe is separated by the central sulcus. The occipital lobe is separated from the parietal lobe by the parietal-occipital sulcus. The lateral sulcus separates the frontal lobe from the temporal lobe.

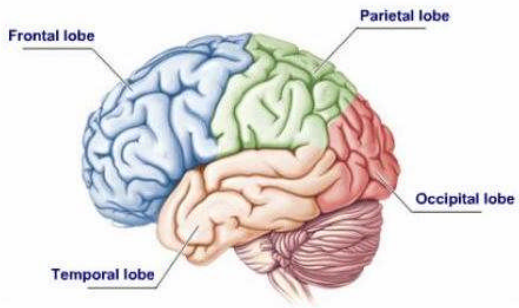


Figure 2 - 10. *The anatomical division of the brain. [Marieb and Hoehn, 2006]*

2.2.2.3 Functionality of cerebral cortex

The cerebral cortex, which constitutes a 2-4 mm thick layer of grey matter, represents a highly-developed structure concerned with the higher order functions associated with the brain. A cortical mapping system produced by Brodmann defines the specific regions of the cerebral cortex and their functionalities based on the cellular composition

of structures within the brain [Brodmann, 1905]. Various function areas of the cerebral cortex and their functionalities are shown in Figure 2 - 11 and Table 2 - 1 respectively.

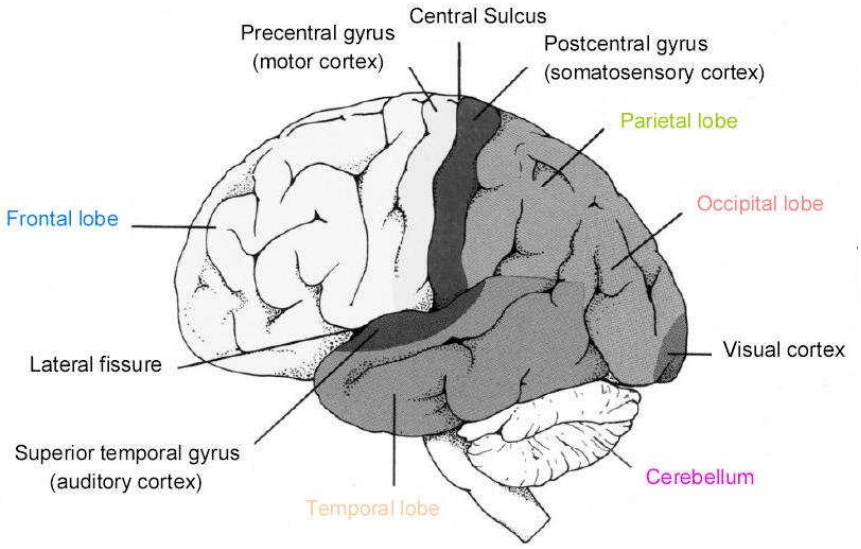


Figure 2 - 11. *Functional regions in cerebral cortex. (reproduced from Shier et al. (2001))*

Cortex Areas	Functionality
Frontal lobe	*Reasoning, planning, parts of speech, emotions, problem solving*
Motor cortex	*Co-ordination of complex movement, initiation of voluntary movement*
Somatosensory cortex	*Receives tactile information from the body*
Parietal lobe	*Perception of stimuli related to touch, pressure, temperature and pain*
Occipital lobe	*Visual processing*
Temporal lobe	*Perception and recognition of auditory stimuli and memory*
Cerebellum	*Learned processes, co-ordination*

Table 2 - 1. *Functionality of main cortex regions.*

2.2.3 International 10-20 system

When cortical areas are active, potential differences are generated between electrodes positioned on the scalp surface. Electroencephalography (EEG) records these potential differences as a function of time. The electrodes are placed in standard positions.

Figure 2 - 12 illustrates the 10-20 international system for electrode placement [Jasper, 1958], which we name 10-20 system here for simplicity. The electrode positions are found by dividing the line between the nasion and inion, and the line between the preauricular points, into intervals which are 10% or 20% of the original length.

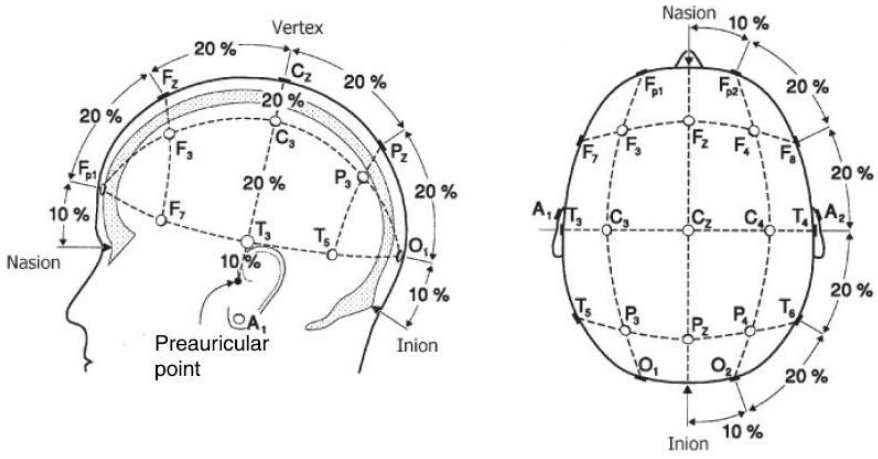


Figure 2 - 12. *The 10-20 international system electrode placement.*

[Malmivuo and Plonsey, 1995]

Following its introduction 50 years ago [Jasper, 1958], the 10-20 system has become the standard approach of electrode application in EEG research. Nineteen standard electrode positions were defined in the original version. Since then, several supplementary guidelines have been published by American EEG Society in 1991 and 1994 to allow for a standardized application of additional electrodes. One basic feature of the 10-20 system is that its proportional measurement strategy, which is based on individual measurements of head circumference, allows accommodating for interindividual differences in head size and head shape, and thus yields electrode locations which are standardized between and within subjects.

For EEG, the accuracy of source localization critically depends on an accurate and reliable placement of the electrodes onto the 10-20 coordinates. Test-retest measurement errors of up to 7 mm, and between subject variability of up to 7.7 mm have been reported even when electrodes were applied by an experienced senior registered EEG technologist [Towle *et al.*, 1993]. In addition, the amount of placement error was dependent on the position of the electrodes, with more lateral electrodes

displaying more error than electrodes placed on the midline. Also, no anatomical landmark could be determined with less than 5 mm of error. Additional within- and across-subject error might result from the experience and reliability of the experimenter, and the length and type of a subject's hair. To conclude, an average electrode placement error of about 1 cm seems to be a realistic estimate [Kavanagh, 1978; Böcker *et al.*, 1994]. For EEG, this value is likely to increase with the use of electrode caps. Although the arrangement of electrodes in such caps usually follows the rules of the 10-20 system, the limited flexibility of the cap fabric and the inability to reposition separate electrodes can result in considerable electrode misplacement. Head movements of the subject during the experiment might also result in additional electrode displacement.

2.2.4 Normalization methods in brain mapping

No two brains are physically identical, however the basic anatomy (topology) is identical, and functions are tied to a particular gyrus (bumps). Since corresponding voxels in images of two different subjects will not correspond to the same brain gyrus, some form of anatomical normalization of brains into a unified coordinate system is necessary to rectify this before analysis of images across subjects on a voxel-by-voxel basis can be carried out. Currently, two kinds of normalization methods are available and they will be discussed in the following sections.

2.2.4.1 Stereotactic atlas

2.2.4.1.1 Definition

To standardise the study of neuroanatomy, various brain atlas have been proposed. Talairach and Tournoux (1988) dissected a human brain, photographing transverse slices onto which a co-ordinate system was overlaid and the brain gyri and lobes labelled. This brain defined the Talairach coordinate system which has become the reference brain to describe brain anatomical locations. Hence, even with studies of an individual, anatomical normalization is necessary to enable locations to be reported in standard Talairach co-ordinates.

Currently all spatially normalized images are transformed to align with the Talairach system [Talairach and Szikla, 1967; Talairach and Tournoux, 1988]. The Talairach space represents the human brain stereotactic atlas where its position is defined by human anatomical landmarks [Brett *et al.*, 2002].

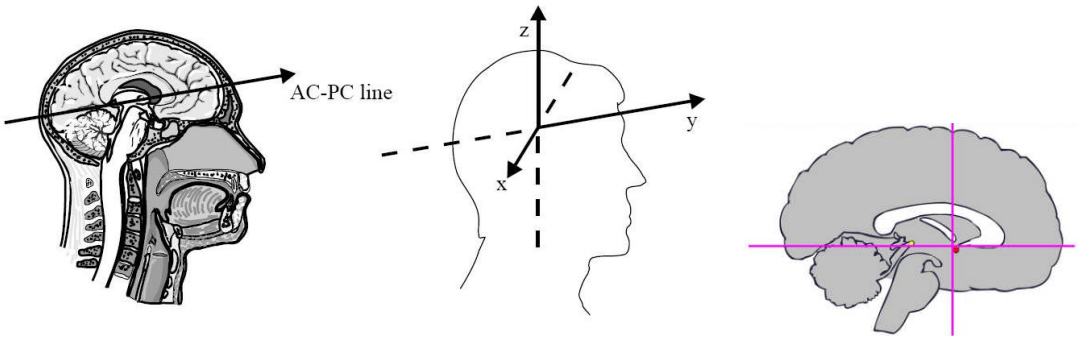


Figure 2 - 13. *The inter-commissural line [left], Talairach axes [middle] and brain schematic showing the anterior commissure (red spot) and posterior commissure (yellow spot) [right].*

The Talairach co-ordinate system is based on an imaginary line joining the centres of the anterior and posterior fibre tracts connecting two brain hemispheres. The midpoint of this line is the origin of the Talairach co-ordinate system (Figure 2 - 13). Co-ordinates are given in millimetres from the origin in the axial directions. The Talairach coordinate system is illustrated in Figure 2 - 14a. Talairach and Tournoux (1988) also compared and numbered cortical areas with Brodmann areas [Brodmann, 1909], as shown in Figure 2 - 14b, so that the stereotactic atlas of Talairach space can be mapped to the corresponding Brodmann areas and the functional areas being activated can be found. The contributions of building of the Talairach space can be summarized as follows:

1. Anatomical landmarks are associated to brain areas by the coordinate system;
2. Brains from different subjects can be mapped to the standard brain atlas through spatial transformation;
3. Every anatomical position corresponds to specific Brodmann area so that researchers can easily find out where the functional area(s) are in the brain through the coordinate system.

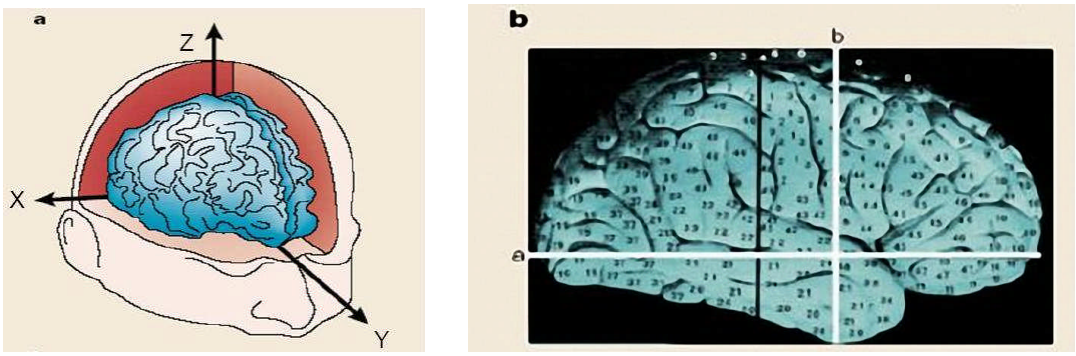


Figure 2 - 14 (a). *Talairach coordinate system;* **(b).** *Brodmann areas. [Brett et al., 2002]*

Compred with a single brain used for the Talairach atlas, 250 MRI scans were used to define the MNI template. Two procedures were used in order to approximately match the MNI brain with the Talairach brain:

1. Landmarks were manually defined for all 250 MRI scans in order to produce a line similar to the imaginary line used for the Talairach brain (Figure 2 - 13). This follows by scaling of each brain to align these landmarks defined in MNI brain with those in the Talairach brain;
2. Evans *et al.* (1992) used a linear method to register an extra 55 images to the database with the 250 scans. The first MNI template, called the MNI 305 atlas, was made by averaging the automatic registered 55 brains with those manually registered 250 brains.

The same averaging procedure was taken for an extra 152 MRI scans on top of the 305 scans to define the current MNI template called ICBM152, which is used by the International Consortium for Brain Mapping as the standard template.

2.2.4.1.2 Difference between Talairach and MNI coordinates

The two 3D space coordinate systems for brain, defined by Talairach (Talairach Deamon Database) and Montreal Neurological Institute (MNI) (The MNI brain and the Talairach atlas), are currently the world wide recognized brain atlases. The Talairach atlas is a coordinate system described by Talairach and Tournoux (1988) after a real brain dissection while the MNI atlas is the coordinate system defined by Evans *et al.* (1992) after they scanned a large number of subjects by Magnetic Resonance Imaging (MRI), followed by superposing scanned images with different head shapes and

smoothing them [Evans *et al.*, 1992; 1993; 1994].

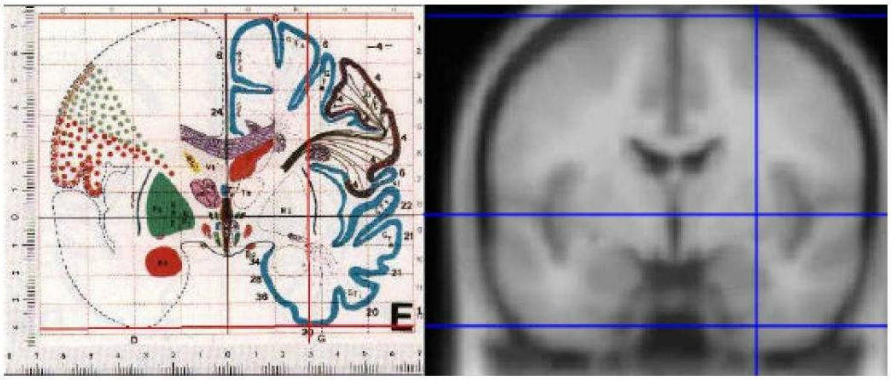


Figure 2 - 15. *The difference between the Talairach and MNI coordinate systems. [MNI brain and the Talairach atlas]*

Figure 2 - 15 shows the difference between the two coordinate systems. In order to make the two systems compatible, some transformations are needed. As we can see from Figure 2 - 15, the Talairach coordinate system is smaller than the MNI system, plus the distance between the centre and the top of the Talairach system is longer. Regarding this, there are several approaches to convert the MNI coordinates to Talairach coordinates. A common approach was proposed by Andreas Meyer-Lindenberg (1998) to apply an affine transform to map the Talairach compatible template onto the MNI compatible template. By assuming very small values of affine transform parameters, the transformation can be formulated as:

$$X' = 0.88X-0.8$$
$$Y' = 0.97Y-3.32$$
$$Z' = 0.05Y+0.88Z-0.44 \qquad 2 - 8$$

where X, Y, Z are the coordinates of the Talairach system, X', Y', Z' are the MNI coordinates after the transformation [MNI brain and the Talairach atlas].

However, this transformation with fixed parameters is not capable of matching every part of the brain from one template to the other due to their different shapes. Regarding this different approaches have been proposed to apply non-linear transformation where different parameters were used for different parts of the brain. Examples of this can be found in [Duncan *et al.*, 2000; Calder *et al.*, 2001].

2.2.4.1.3 Problems of Talairach atlas

The most well known method, proposed by Talairach and Tournoux, normalizes different brains to a standard brain template by translation, rotation, and scaling. However, the problem of this method is that it does not consider individual differences in detailed structures. Besides, the original template is based on an old European person. Imagine there is a subject with a smaller head size than the brain template, after the spatial normalization and transformation, the brain cortex of the subject will be enlarged in order to accommodate the template and this may result in a decrease of the activation on the cerebral cortex and an unnecessary increase of effects caused by some voxels [Li *et al.*, 2004].

2.2.4.2 Registration of EEG and MRI

In order to achieve higher mapping accuracy between the topographic EEG activity with individual head, the techniques for combining EEG and MRI have received a lot of attention. Compared with the conventional mapping methods between EEG induced activities and a 2D scalp surface, mapping of brain activities onto a 3D structural head image demands highly accurate transformation of the EEG electrodes' positions into the MRI coordinate system. Regarding this various approaches have been proposed for accurate mapping such as matching of fiducials between EEG and MRI coordinate systems [Towle *et al.*, 1993], matching of electrodes visible in MR images [Gevins *et al.*, 1991, 1994; Lagerlund *et al.*, 1993; Ives *et al.*, 1993], and the usage of specifically designed equipments to assist the landmark identification between the two systems [Singh *et al.*, 1997; Simpson *et al.*, 1995; Barnett *et al.*, 1993]. However, these techniques generally lack practicability either because of variability in order of millimetres existed in landmark digitization [Singh *et al.*, 1997] or because additional handware and software are required.

An alternative approach to reduce errors of digitization of individual fiducials is the matching of head reconstruction between EEG and MRI [Brinkmann *et al.*, 1998; Huppertz *et al.*, 1998; Wang *et al.*, 1994], where the scalp surface reconstruction for EEG is usually performed by digitization between 1000 and 2000 arbitrary points with

equipments such as a sensor pen digitizer. Although additional time (10-20 minutes) are usually required for such a digitization process, it is considered to be essential for reliable and accurate determination of the electrodes' positions to ascertain the mapping accuracy.

One typical example was proposed by Lamm *et al.* (2001), where a spline interpolation method [Perrin *et al.*, 1987] was used to reconstruct the scalp surface from the 3D digitized electrode positions. An iterative point matching method [Zhang, 1994] was then used to match the interpolated scalp surface with the MRI segmented surface. This approach simplified the surface matching in which no additional equipments or digitization of the head surface were required, leaving the measurement of the 3D electrodes' positions as the only procedure for each subject. The additional electrode coordinates was measured by a 3D photogrammetric head digitizer, which consisted of twelve calibrated Olympus Camedia C-400 cameras mounted in a dome imaging active targets (light emitting diode marking electrodes). Bauer *et al.* (2000) reported that this head digitizer achieved an average measurement accuracy of 0.2mm and average test-retest reliability of 0.18mm. A simulated co-registration following these procedures demonstrated a registration accuracy of about 0.6mm. This accuracy is subject to further increase if individual rather than standard electrodes are used for digitization since they can better accommodate individual head shapes.

2.3 State of the arts

2.3.1 3D anatomical cranial-cerebral correlation

While cortical projection points of the 10-20 standard positions were expressed in 2D Brodmann's coordinates [Brodmann, 1909, 1912] by Homan *et al.* (1987), Okamoto *et al.* (2004a) projected and expressed the 10-20 positions on 3D space, in order to enhance the comparison among different brain mapping methods. The work presented by Okamoto *et al.* (2004a) can be divided into three steps:

1. Projection of 10-20 points onto the cortical surface

   Using a structural MRI head data, manual projection was performed by searching for the minimum distance between cortical surface and 10-20 positions on scalp surface. The depth of these 10-20 points from the skin surface to cortical surface were also examined by Okamoto *et al.* (2004a) (Table 2 - 4);

2. Normalization of the projected points in standard brain atlas

   10-20 points and their corresponding cortical projection points of seventeen healthy adults were normalized by Okamoto *et al.* (2004a) to both the MRI and Talairach coordinate systems, followed by deriving their individual positions and uncertainties (Table 2 - 2);

3. Probabilistic anatomical description of the 10-20 points on cortical surface (Table 2 - 3).

Furthermore, Okamoto *et al.* (2004a) reported that the uncertainty of projection positions on cortical surface is mainly caused by the following three factors:

1. Measurement error on head surface

   A mean displacement of 1.4mm/2.5mm of the 10-20 points in three measurements on the same day/three different days [Towel *et al.*, 1993] were included in the position estimation by Okamoto *et al.* (2004a);

2. Variation of the gross structures of brain and scalp between subjects

   The normalization does not perfectly match each individual brain with the template brain. The positional discrepancy between each normalized brain and the template brain contributed to the overall uncertainty [Okamoto *et al.*, 2004a];

3. Variation of normalization methods

   Okamoto *et al.* (2004a) reported that different magnitudes of uncertainty can be resulted from different template atlases or even different programs used for normalization.

	MNI coordinates				Talairach coordinates			
	Position			SD	Position			SD
	x	*y*	*z*		*x*	*y*	*z*	
Fp1	−26.1	83.5	−0.1	5.9	−25.9	80.9	−4.0	7.5
Fp2	32.1	81.3	−0.2	5.3	31.8	78.8	−4.0	7.0
Fz	−0.1	53.3	70.6	10.3	−0.1	54.9	62.3	9.8
F3	−42.6	58.0	39.6	11.5	−42.2	58.0	33.6	11.0
F4	46.9	56.7	40.2	9.5	46.5	56.8	34.2	9.0
F7	−68.5	37.9	−5.9	6.6	−67.8	36.4	−6.8	7.0
F8	71.4	35.6	−7.8	6.2	70.7	34.2	−8.2	6.3
Cz	0.6	−12.7	101.4	7.5	0.6	−7.6	93.8	7.3
C3	−62.7	−12.7	69.8	9.7	−62.1	−9.1	64.7	9.3
C4	64.2	−15.2	69.9	8.8	63.6	−11.5	65.0	8.4
T3	−84.6	−20.7	−10.8	8.5	−83.7	−20.5	−8.1	8.0
T4	86.0	−25.5	−9.4	7.4	85.1	−25.1	−6.7	6.8
Pz	−0.2	−76.6	88.5	9.0	−0.2	−70.1	85.0	8.7
P3	−46.8	−88.2	58.8	10.5	−46.4	−82.8	58.3	10.0
P4	44.8	−87.9	59.6	9.5	44.4	−82.4	59.1	8.9
T5	−72.2	−72.4	0.6	9.3	−71.5	−70.1	4.2	10.7
T6	70.8	−75.6	4.1	8.7	70.1	−73.1	7.6	10.0
O1	−31.9	−112.6	17.3	12.6	−31.6	−108.3	21.4	12.6
O2	28.1	−112.9	19.3	13.5	27.8	−108.5	23.2	13.3
All				9.2				9.2

All values are in millimeters. SD stands for standard deviation.

(a)

	MNI coordinates				Talairach coordinates			
	Position			SD	Position			SD
	x	*y*	*z*		*x*	*y*	*z*	
Fp1	−21.5	70.2	−0.1	5.0	−21.3	68.0	−3.0	7.5
Fp2	28.4	69.1	−0.4	5.1	28.1	66.9	−3.6	6.8
Fz	0.6	40.9	53.9	9.8	0.6	42.1	47.5	9.5
F3	−35.5	49.4	32.4	9.2	−35.1	49.3	27.4	8.8
F4	40.2	47.6	32.1	8.6	39.8	47.6	27.2	8.2
F7	−54.8	33.9	−3.5	6.5	−54.2	32.7	−4.4	8.4
F8	56.6	30.8	−4.1	6.2	56.0	29.7	−4.7	8.0
Cz	0.8	−14.7	73.9	8.3	0.8	−10.8	68.6	8.1
C3	−52.2	−16.4	57.8	8.0	−51.7	−13.2	53.9	7.7
C4	54.1	−18.0	57.5	7.1	53.6	−14.8	53.7	6.8
T3	−70.2	−21.3	−10.7	6.5	−69.5	−21.1	−8.1	7.2
T4	71.9	−25.2	−8.2	6.2	71.2	−24.7	−5.8	6.4
Pz	0.2	−62.1	64.5	9.2	0.2	−57.2	62.3	8.8
P3	−39.5	−76.3	47.4	7.4	−39.1	−71.7	47.3	7.1
P4	36.8	−74.9	49.2	7.6	36.5	−70.3	48.8	7.2
T5	−61.5	−65.3	1.1	7.8	−60.9	−63.2	3.9	9.2
T6	59.3	−67.6	3.8	7.3	58.7	−65.4	6.4	8.0
O1	−26.8	−100.2	12.8	10.1	−26.6	−96.5	16.1	9.7
O2	24.1	−100.5	14.1	11.1	23.9	−96.7	17.3	10.5
All				7.9				8.2

All values are in millimeters. SD stands for standard deviation.

(b)

Table 2 - 2. *Positions and their positional uncertainties of the standard 10-20 points on (a) scalp surface and (b) cortical surface. (Reproduced from Okamoto et al. (2004a))*

	Labeling by Talairach Daemon				Manual labeling			
	Anatomy	%	BA	%	Anatomy	%	BA	%
Fp1	L FL Superior frontal G	97	10	100	L FL Superior frontal G	94	10	100
	L FL Medial frontal G	2			L FL Middle frontal G	6		
Fp2	R FL Superior frontal G	96	10	100	R FL Middle frontal G	65	10	100
	R FL Middle frontal G	4			R FL Superior frontal G	35		
F3	L FL Superior frontal G	56	10	47	L FL Middle frontal G	81	9	63
	L FL Middle frontal G	44	9	43	L FL Superior frontal G	19	10	31
F4	R FL Middle frontal G	60	10	49	R FL Middle frontal G	98	9	52
	R FL Superior frontal G	40	9	34	R FL Superior frontal G	2	46	25
F7	L FL Inferior frontal G	84	47	81	L FL Inferior frontal G	88	47	63
	L FL Middle frontal G	16	45	13	L FL Middle frontal G	13	45	19
F8	R FL Inferior frontal G	94	47	94	R FL Inferior frontal G	100	47	60
	R FL Middle frontal G	6	45	6			45	29
C3	L PL Postcentral G	77	3	60	L PL Postcentral G	65	3	35
	L FL Precentral G	15	6	15	L FL Precentral G	35	4	29
C4	R PL Postcentral G	89	3	60	R PL Postcentral G	87	3	44
	R FL Precentral G	11	1	17	R FL Precentral G	9	1	24
T3	L TL Middle temporal G	94	21	94	L TL Middle temporal G	88	21	88
	L TL Superior temporal G	6	22	6	L TL Superior temporal G	12	22	12
T4	R TL Middle temporal G	96	21	95	R TL Middle temporal G	85	21	85
	R TL Superior temporal G	4	22	5	R TL Superior temporal G	15	22	15
P3	L PL Superior parietal L	59	7	71	L PL Angular G	89	39	89
	L PL Precuneus	24	19	24	L PL Supramarginal G	10	40	10
P4	R PL Superior parietal L	75	7	82	R PL Angular G	63	39	63
	R PL Precuneus	18	19	18	R OL Superior occipital G	22	19	22
T5	L TL Middle temporal G	61	37	51	L TL Middle temporal G	54	21	49
	L OL Inferior temporal G[a]	15	21	22	L TL Inferior temporal G	31	37	37
T6	R TL Middle temporal G	66	37	54	R TL Middle temporal G	52	21	40
	R OL Middle occipital G	19	39	24	R TL Inferior temporal G	25	37	37
O1	L OL Middle occipital G	69	19	48	L OL Middle occipital G	100	19	53
	L OL Cuneus	13	18	48			18	47
O2	R OL Middle occipital G	68	18	61	R OL Middle occipital G	88	18	71
	R OL Cuneus	20	19	33	R OL Superior occipital G	9	19	26

The two most likely two anatomical structures and Brodmann's area are shown. Note that anatomical structures and Brodmann's areas that are based on histochemical structures do not always coincide. Abbreviations: BA = Brodmann's area, L = left, R = right, FL = frontal lobe, TL = temporal lobe, PL = parietal lobe, OL = occipital lobe, G = gyrus, L = lobule.

[a] The Talairach Daemon program sometimes provides labels with mismatches between the lobe and gyrus.

Table 2 - 3. *Anatomical expression of each 10-20 points after cortical projection.*

(Reproduced from Okamoto et al. (2004a))

	Mean	SD		Mean	SD		Mean	SD
Fp1	13.5	2.2	Cz	25.8	4.0	P4	17.6	2.2
Fp2	12.6	1.9	C3	16.0	2.3	T5	12.7	1.9
Fz	19.7	2.4	C4	15.6	1.9	T6	13.0	1.4
F3	13.8	1.8	T3	12.9	2.1	O1	13.8	1.5
F4	13.9	1.9	T4	13.1	2.1	O2	13.7	1.2
F7	14.5	2.4	Pz	25.9	4.2			
F8	16.7	3.1	P3	16.9	2.6	All	15.9	5.6

All values are in millimeters. SD stands for standard deviation. Cortical depth is defined as the distance between a given 10–20 standard position on the head surface and its cortical projection point. The row indicated as "all" shows the mean cortical depth for all the 10–20 standard positions in 17 subjects and their standard deviations.

Table 2 - 4. *Distance between scalp surface and cortical surface for each 10-20 points.*

(Reproduced from Okamoto et al. (2004a))

2.3.2 Automated cortical projection

Regarding the manual projection approach which is not practical, Okamoto and Dan (2005) presented three automated projection methods:

1. Convex-hull projection (with MRI head data)

   This method first created a convex hull on the cortical surface. The corresponding projection point on cortical surface of a 10-20 point is found by searching for a virtual point on the hull producing the minimum distance between the scalp surface and the cortical surface;
2. Ballon-inflation projection (with MRI head data)

   A group of points on the cortical surface which are the closest to each 10-20 point (to be projected) are first found, followed by deriving the centroid of these points. The corresponding cortical projection point is then found by the intersection between the line (formed from the 10-20 point on scalp surface and the computed centroid) and the cortical surface. Compared with the convex-hull method, the ballon-inflation method has a higher sensitivity to local cortical structures;
3. Neighnour-reference projection (without MRI head data)

   The location of a 10-20 point on scalp surface in MNI coordinate system is first determined based on the closest four 10-20 standard points. The estimated 10-20 point is then projected into the cortical surface of a brain template by either convex-hull or ballon-inflation method. The error is estimated based on the positional uncertainty in MNI space presented in Table 2 - 2.

Okamoto and Dan (2005) reported that the convex-hull and ballon-inflation methods achieved a spatial resolution of a few millimetres and the neighbour-referenece method achieved a resolution of about 1cm in the absence of structural images. However, there are some limitations regarding these automated projection methods. First, the 10-20 points to be projected need to be first transformed into the corresponding points in the same coordinate system as the MRI head data; second, the nineteen standard 10-20 points with defined cranio-cerebral correlations appear not sufficient as reference points for the projection; third, the data of the reference points presented in Section 2.3.1 is

based on only seventeen subjects and therefore is unlikely to be reliable to represent the population.

Regarding the neighnour-reference method, Singh *et al.* (2005) further validated this registration method without MRI data by registering additional twelve subjects to the original MRI data containing seventeen subjects' MRI scans. Results indicated positional uncertainties for this virtual projection method are between 4.7mm and 7.0mm. This level of uncertainties has demonstrated the capability of the virtual registration method in multi-subject analysis when MRI dataset are not available. Singh *et al.* (2005) further stressed the importance of taking the positional uncertainties into account when mapping the 10-20 points onto the brain atlas, i.e. Talairach or MNI coordinate system.

2.3.3 Virtual 10-20 measurement on MR images

To comply with the developed automated projection methods (Section 2.3.2), Jurcak *et al.* (2005) developed a virtual measurement method to replace the conventional manual measurement procedure. The validation of the vrtual measurement method was carried out by normalizing the 10-20 points to the MNI atlas and comparing the normalized positions with those obtained from manual measurement (Section 2.3.1), where the positional discrepancies are presented in Table 2 - 5.

	Physical marking				Virtual computation			
	Position			SD	Position			SD
	x	*y*	*z*		*x*	*y*	*z*	
Fpz	2.9	87.1	0.2	5.3	1.2	86.5	−3.9	5.2
Fp1	−26.1	83.5	−0.1	5.9	−27.9	82.9	−4.9	5.8
Fp2	32.1	81.3	−0.2	5.3	31.2	82.0	−4.3	5.3
Fz	−0.1	53.3	70.6	10.3	1.3	57.0	66.3	6.1
F3	−42.6	58.0	39.6	11.5	−43.4	59.2	36.6	6.9
F4	46.9	56.7	40.2	9.5	46.2	59.5	36.7	5.7
F7	−68.5	37.9	−5.9	6.6	−68.8	37.1	−8.7	5.4
F8	71.4	35.6	−7.8	6.2	71.4	35.3	−8.5	4.7
Cz	0.6	−12.7	101.4	7.5	0.7	−11.4	99.7	8.2
C3	−62.7	−12.7	69.8	9.7	−66.0	−15.5	65.2	7.9
C4	64.2	−15.2	69.9	8.8	67.3	−16.1	65.1	7.1
T3	−84.6	−20.7	−10.8	8.5	−83.5	−23.8	−13.0	5.5
T4	86.0	−25.5	−9.4	7.4	84.6	−25.2	−11.9	4.5
Pz	−0.2	−76.6	88.5	9.0	−0.7	−82.4	82.4	9.9
P3	−46.8	−88.2	58.8	10.5	−43.2	−94.2	51.3	10.7
P4	44.8	−87.9	59.6	9.5	41.3	−94.4	51.1	9.9
T5	−72.2	−72.4	0.6	9.3	−70.1	−77.6	4.0	8.2
T6	70.8	−75.6	4.1	8.7	69.0	−78.5	4.1	7.6
O1	−31.9	−112.6	17.3	12.6	−28.8	−113.6	15.3	12.1
O2	28.1	−112.9	19.3	13.5	25.8	−114.1	14.9	11.5
Oz	−1.9	−118.2	18.8	12.4	−1.5	−117.8	15.8	12.9
All				9.0				7.7

Location of international 10–20 standard point positions on the head surface in MNI coordinates determined by physical marking and virtual computation. All values are in millimeters. SD stands for standard deviation.

Table 2 - 5. *Comparison of 10-20 positions between manual measurement and virtual measurement in MNI coordinate system. (Reproduced from Jurcak et al. (2005))*

Although the virtual measurement may not produce the 10-20 positions as accurate as these measured manually, the virtual method can be considered as an alternative to manual measurement if a large amount of data is involved where 10-20 positions need to be determined.

2.3.4 Virtual registration of NIRS to MNI space

In order to eliminate the need of structural MRI data and the manual measurement of 10-20 points using tools such as a 3D digitizer, Tsuzuki *et al.* (2007) proposed a virtual registration method, provided with the locations and deformations of the optode holder. Optode holders were virtually placed onto the head surface by running a simulation using the virtual holder deformation algorithm, which produced average discrepancies of less than 2cm as compared with the actual holder deformation, for all studies with real heads and most studies with spherical phantoms. However, Tsuzuki *et al.* (2007) reported that the stability of spatial estimations decreased from the most stable frontopolar region, to temporal and parietal regions, and to the least stable occipital region. Tsuzuki *et al.* (2007) further reported a mean discrepancy of about 1cm between the virtual registration method and the registration method using a 3D digitizer. The spatial errors were not equally distributed, where occipital region displayed larger positional discrepancy between these two methods.

Despite the absence of MRI scans and 3D digitizer, there are some limitations regarding this virtual registration method:

1. Many parameters are required from user input;
2. Small adjustments are required for each virtual holder registration;
3. The virtual holder deformation is sensitive to the type as well as the method of the holder used;
4. The distance between optodes needs to be exactly the same.

Chapter 3 Digital Photogrammetry

3.1 Introduction

Digital camera technology has rapidly evolved in the last decade. Meanwhile the advance of digital camera technology towards webcams has made the image resolutions of current webcams comparable to those found in high end cameras. Please note here that high end cameras still have many more physical pixels than webcams which mainly adopt interpolation. The evolving software enhancement techniques on webcam images have also compensated, to some extent, their disadvantages with relatively low cost imaging components. The latest webcams delivering HD imagery has even made their image quality apparently as good as that of those cameras which have well established their use in photogrammetric measurement, such as Nikon D100. However, the photogrammetric qualities of these evolving webcams have not been fully assessed from a limited amount of information available regarding the metric use of webcams. The limited reported research in this area leaves the possibility of adopting low cost webcams in metric use uncertain.

On the other hand, as described in Section 2.1.3.3 and 2.2.3, close range photogrammetric coordination of both optical sensing pads and a subject's head during optical imaging studies would support the use of optical imaging techniques for a better localisation of brain activation area and meanwhile solve problems due to movement artefacts. The optical topography system is highly portable and flexible allowing optical imaging studies to be carried out from patient's bedside to intensive care unit of premature neonates. Therefore if a photogrammetric system is to be built to support this emerging functional imaging technique, highly portable and flexible cameras providing sufficiently accurate 3D coordination would be required.

Regarding the above factors, this research focuses on exploring the potential of these highly portable webcams through investigations of their opportunities for close range measurements in optical imaging studies. Since this research is directed towards imaging heads, imaging distances will be similar to those found in video conferencing or casual use of webcams so that this application is likely to be at the optimal range for webcam photogrammetry. This chapter starts with the photogrammetric principles

(Section 3.2), followed by comparisons between conventional high cost cameras and webcams from different perspectives and the associated calibration methods adopted for webcams (Section 3.3). Photogrammetry with targets (Section 3.4) discusses target properties and lighting suited to clinical use and the fundamentals of automated target image measurement suited to webcam photogrammetry. The well-established 3D correspondence solution with its optimisation (Section 3.5) in favour of a fast coordination application are then described, followed by the considerations and underlying principles towards the development of a fast multi-webcam photogrammetric coordination system (Section 3.6).

3.2 Photogrammetric principles

Photogrammetry aims at obtaining precise and accurate 3D object information merely from their image(s). Close range photogrammetry is used to distinguish its measurement applications from aerial photogrammetry in that normally a measurement distance within 100 meters can be treated as close range. Among different approaches of close range photogrammetry, the method where images are obtained from camera positions all around the object and camera axes are highly convergent, pointing towards the middle of the object (convergent network of images) yields the most accurate results when imaging features all around an object (head) and is commonly applied to capture targeted objects (target points around the head and optical sensing pad of optical topography) [Atkinson, 2001]. This section discusses the fundamentals of the convergent network geometry in close range photogrammetry.

3.2.1 Collinearity for an ideal camera

Figure 3 - 1 shows the central perspective projection geometry which forms the basics of photogrammetric measurements [Atkinson, 1996]:

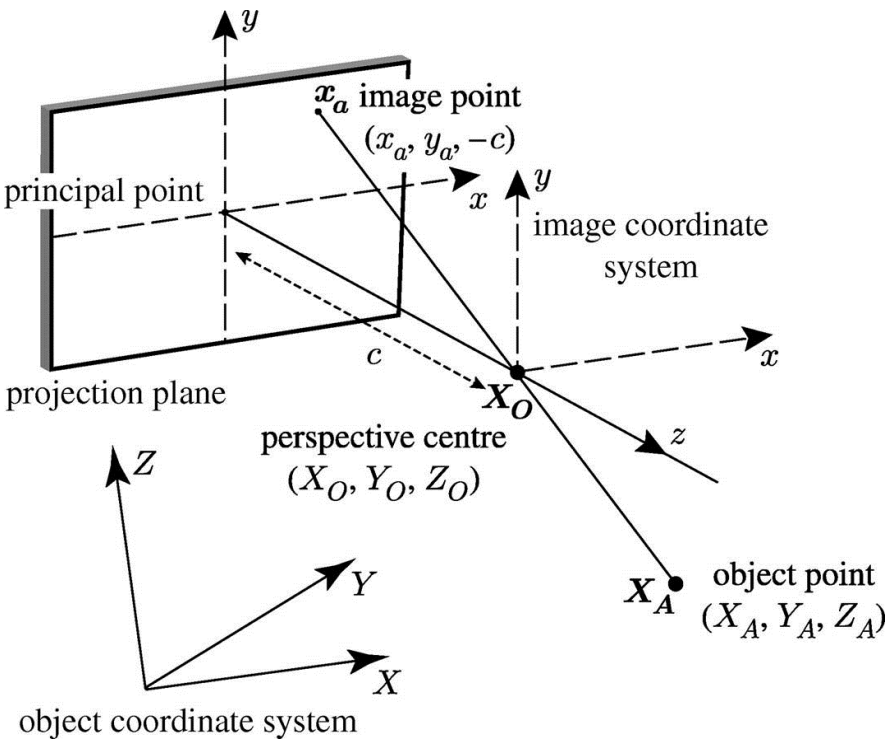


Figure 3 - 1. *Central perspective projection. (reproduced from Atkinson (2001))*

In central perspective projection model, a straight line is always assumed between a point $\mathrm{X_A}$ in object space and its associated projected point $\mathrm{x_a}$ on image passing through the perspective centre $\mathrm{X_O}$ of the imaging system. In the imaging system, the perpendicular distance between the perspective centre and the image plane is principal distance c. In an ideal camera system, the optical axis formed between the perspective centre and principal point will coincide with the centre of the image plane. However this is rare in practical case. To establish a geometric relationship between an arbitrarily located object space coordinate system $\{\mathrm{X,Y,Z}\}$ and the image coordinate system $\{\mathrm{x,y,z}\}$ with origin situated in the perspective centre, their originas can be both express as the perspective centre $\mathrm{X_O}$, e.g. the object point $\mathrm{X_A}$ with its object space coordinates $(\mathrm{X_A,Y_A,Z_A})$ can be treated as the addition of the vector from the origin of object coordinate system to perspective centre $\mathrm{X_O}$ with its coordinates $(\mathrm{X_O,Y_O,Z_O})$ and the vector from $\mathrm{X_O}$ to $\mathrm{X_A}$. While the image plane is parallel to the x and y axis of the image coordinate system, the image point $\mathrm{x_a}$ can be expressed as $(\mathrm{x_a,y_a,-c})$ with a distance of c back from the image coordinate system.

However, the image coordinate system also needs rotation to fully align its three axises with those of object coordinate system. Therefore under the assumption of an ideal camera, the relationship between an object point and its image point can be mathematically expressed by the collinearity equations [Atkinson, 2001]:

$$x_a = \frac{-c\left[r_{11}\left(X_O - X_A\right) + r_{12}\left(Y_O - Y_A\right) + r_{13}\left(Z_O - Z_A\right)\right]}{\left[r_{31}\left(X_O - X_A\right) + r_{32}\left(Y_O - Y_A\right) + r_{33}\left(Z_O - Z_A\right)\right]}$$

$$y_a = \frac{-c\left[r_{21}\left(X_O - X_A\right) + r_{22}\left(Y_O - Y_A\right) + r_{23}\left(Z_O - Z_A\right)\right]}{\left[r_{31}\left(X_O - X_A\right) + r_{32}\left(Y_O - Y_A\right) + r_{33}\left(Z_O - Z_A\right)\right]} \qquad 3 - 1$$

where r_{ij} are the elements of the rotation matrix.

3.2.2 Departures from collinearity

When the central perspective projection is used in photogrammetry as the mathematical standard for image formation in a camera, significant discrepancies leading to systematic errors will occur. These errors can be caused by a number of factors such as physical lens distortions, unflatness of the imaging sensor and orthogonality of the optical axes. Departures from the central perspective projection due to imperfections of imaging components in real camera design can be modeled as systematic errors in the collinearity condition. The procedure of seeking how much the geometry of image formation in a real camera differs from the geometry of a central perspective projection is known as camera calibration [Atkinson, 2001].

Camera calibration is used to estimate a number of parameters which describe the internal imaging geometry of a camera. The most fundamental parameter is the principal distance which changes with focus. Principal distance must be known approximately and subsequently estimated [Fryer, 1988; Clarke and Fryer, 1998]. Suitable starting values of the principal distance might come from the manufacturer's data [Fryer, 1988]. Fraser and Al-Ajlouni (2006) utilised a series of magnifications from a zoom lens expressed as a function of principal distance and subsequently wrote them to the image EXIF header in order to model radian distortion. Fryer (1988) found discrepancies up to 1.5mm in a metric camera of stated principal distance 99.60mm and up to 2.5mm for lenses in non-metric cameras addressing the need to subsequently estimate the principal distance from the starting values. While the computer vision community often estimate principal distance on a frame by frame basis, there is an understanding in the photogrammetric community that reliable metric results can only be achieved with a calibration process considering a number of views within which physical variations in the imaging system are kept to a minimum. Camera calibration accounting for the physical variations as well as understanding the optical/electronic and compression based effects on image geometry become very important for non-metric cameras especially for webcams so that they can be utilised effectively and reliably for photogrammetry [Mcnamee *et al.*, 2001].

Real lenses have lens distortions accounting for the departures from collinearity (for an ideal camera with perfect lens), where the degree of lens distortion relies on the individual lens design. In general the less the lens distortion, the higher the cost the lens has. Brown (1976) stated that 'since radial and tangential distortion could be effectively modelled, there was no impediment to using any well-regarded commercial lens of suitable principal distance, aperture and angular field…for it is image quality throughout the format that now becomes the overriding factor in the ultimate determination of metric potential'. The comment from Brown has lead to a rigirous exploration of the potential of non-metric cameras with focusable lens in terms of photogrammetric measurements, where these cameras have been proved to be capable of providing comparable results to metric cameras if their lens distortions can be effectively modelled [Clarke and Fryer, 1998]. Although it is possible but costly to produce a 'distortion free' lens, for this research work a lens (camera system) with constant geometric distortions is considered to be sufficient since the distortions can then be modelled from the camera calibration process.

3.2.2.1 Radial lens distortion

Variations in angular magnification with angle of incidence are usually expressed as radial lens distortion [Atkinson, 2001]. Radial distortion is usually expressed as a polynomial function of the radial distance from the point of symmetry, which usually coincides with the principal point:

$$\delta r = K_1 r^3 + K_2 r^5 + K_3 r^7 \qquad 3 - 2$$

where

δr is the radial displacement of an image point;

$r^2 = (x - x_o)^2 + (y - y_o)^2$;

(x, y) are the coordinates of the image point;

(x_o, y_o) are the coordinates of the point of symmetry, commonly the principal point;

K_1, K_2 and K_3 are coefficients who values depend on the camera focus setting.

The distortion δr can be resolved into two components:

$$\delta r_x = \delta r\left(x - x_o\right)/r$$

$$\delta r_y = \delta r\left(y - y_o\right)/r \qquad 3 - 3$$

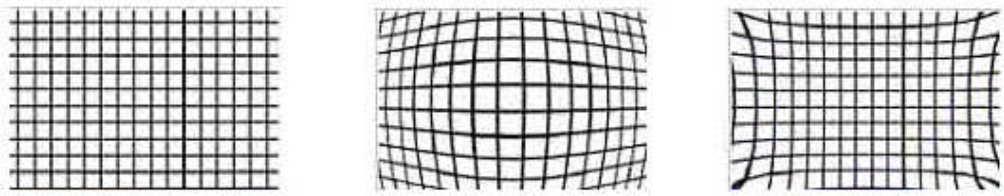


Figure 3 - 2. *Radial distortion.*

left: Original grid; middle: Barrel distortion; right: Pincushion distortion.

While radial distortion for modern film cameras is normally 1-2 µm, for an inexpensive CCD camera lens the maximum radial distortion can be as much as 300µm [Atkinson, 2001]. Since this book was published there has been a major consumer orientated development in low cost webcam technology. For example, the analysis carried out by Page *et al.* (2008) demonstrated a very small distortion (<0.5 pixels within area of image centre and up to 3.5 pixels towards image corner) using a single Logitech QuickCam Pro 4000 webcam with a resolution of 640×480 representing the new low cost end of the market. These results well agree with that found in this research (Section 4.4.2.2.3). However, a distortion model is necessary to correct this error if accurate measurement is demanded [Page *et al.*, 2008].

3.2.2.2 Tangential distortion

Tangential distortion is caused by the misalignment between the lens components during the manufacturing process. Such displacement can be modelled in both x and y directions on an image place [Atkinson, 2001]:

$$\delta x = P_1\left[r^2 + 2\left(x - x_o\right)^2\right] + 2P_2\left(x - x_o\right)\left(y - y_0\right)$$

$$\delta y = P_2\left[r^2 + 2\left(y - y_o\right)^2\right] + 2P_1\left(x - x_o\right)\left(y - y_0\right) \qquad 3 - 4$$

where

P_1 and P_2 are coefficients who values depend on the camera focal setting;

$r^2 = \left(x - x_o\right)^2 + \left(y - y_o\right)^2$;

(x, y) are the coordinates of the image point;

(x_o, y_o) are the coordinates of the point of symmetry, commonly the principal point;

$(x - x_o, y - y_o)$ represents the coordinates of image relative to the principal point.

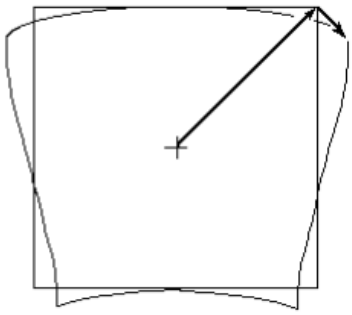


Figure 3 - 3. *Tangential distortion.*

Fryer *et al.* (1994) stressed the importance of lens distortion correction in photogrammetric applications. While Curry *et al.* (1986), Gulch (1986), and Beyer (1993) gave detailed methodologies to calibrate video cameras, Brown (1971) has a significant contribution of calibration of lens distortions under a specific focus setting. The underlying principle of these calibration techniques is always based on a straight line from an object point through the lens perspective centre to its associated point on the image plane.

3.2.2.3 Image plane flatness correction

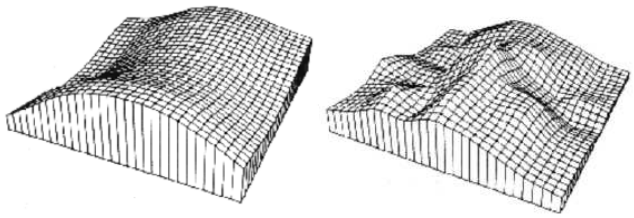


Figure 3 - 4. *Ideal surface model (left) and actual surface model (right).*

Since the actual image plane is not flat, as shown in Figure 3 - 4, there will be a difference between the actual and the ideal image coordinates, as mentioned earlier in Section 3.2.2, which needs to be corrected. The actual image location relative to the ideal one can be derived by [Atkinson, 2001]:

$$x = (x - x_o) + \left(\frac{x - x_o}{r}\right)\delta r + \delta x + a_1(y - y_o) + a_2(x - x_o)$$

$$y = (y - y_o) + \left(\frac{y - y_o}{r}\right)\delta r + \delta y \qquad 3 - 5$$

where

$(x - x_o, y - y_o)$ represents the coordinates of image relative to the principal point;

r represents radial displacement from the principal point, i.e. $r^2 = (x - x_o)^2 + (y - y_o)^2$;

δr represents the radial displacement of an image point;

δx and δy are defined in equations 3 - 4;

a_1 and a_2 represent orthogonality and affinity of image coordinate system shown in Figure 3 - 5.

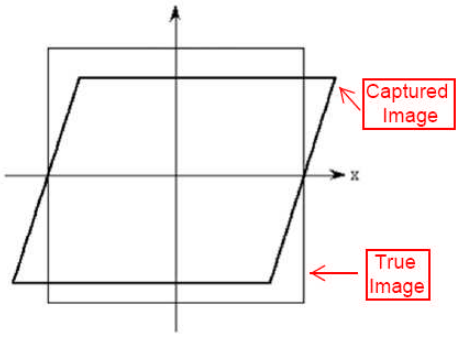


Figure 3 - 5. *Affinity and Orthogonality.*

3.2.2.4 Camera calibration parameter summary

The model used in this thesis is defined by the common set of parameters, where the first eight parameters are the 'physical' model originally formulated by Brown (1971), and the last two represent the extended model to correct the image plane flatness [Beyer, 1993]. Remondino and Fraser (2006) reported that this model can be treated as the optimal calibration set for digital cameras, based on numerical investigations of different sets of calibration parameters such as those carried out by Abraham and Hau (1997). However, the model needs to be refined during the calibration process so that only significant terms are included.

1. PP_x - X image coordinate of the principal point;
2. PP_y - Y image coordinate of the principal point;
3. PD - Principal distance of the camera;
4. K_1 - 3rd power term of radial lens distortion;
5. K_2 - 5th power term of radial lens distortion;
6. K_3 - 7th power term of radial lens distortion;
7. P_1 – 1st term of tangential (decentring) lens distortion;
8. P_2 – 2nd term of tangential (decentring) lens distortion;
9. a_1 - Orthogonality of the image coordinate system;
10. a_2 - Affinity of the image coordinate system.

3.2.3 Photogrammetric bundle adjustment

One of the main tasks of photogrammetry is to compute the precise 3D positions on an object given a set of images. Recently with the increasing use of digital cameras, features like points and lines are observed in the images and their 3D coordinates are determined in a given coordinate system under the perspective projection model, followed by running the bundle adjustment [Brown, 1976]. The bundle adjustment is a global minimization of the re-projection error based on the collinearity equation (Section 3.2.1). It was developed in the 50's and since then extended [Granshaw, 1980; Triggs *et al.*, 2000] to model possible sensor and lens systematic errors. A key advantage of the bundle adjustment is its flexibility in allowing the user control of the imaging geometry whilst accurately deriving interior and exterior orientation parameters. A strong convergent network design is particularly important for calibration of cameras with relatively poor imagery such as webcams.

Photogrammetry aims to get the best possible accuracy with a certain system and image network, which is required by many applications such as industrial measurements, deformation or movement analysis. Pre-calibrated cameras (cameras with laboratory determined calibration parameters) are often used, not only to avoid deficiencies since each camera may contribute only a single image to the network (Section 3.2.3.1), but also to speed up the measurement process. In order to use

consumer digital cameras for such applications, such as the off-the-shelf webcams used in this research, calibration must be carried out and the stability of the estimated parameters ensured. Where more images are taken per camera, a self-calibrating bundle adjustment can effectively recover all the camera parameters under particular network conditions [Remondino and Fraser, 2006]. In most photogrammetric applications, the uncertainty of the measurements is taken into account and a final statistical analysis is performed. In close-range photogrammetry several commercial packages can automatically recover the orientation and calibration parameters of a network of images, using retro-reflective or colour coded targets [VMS, iWitnessTM] or exterior orientation devices [V-StarTM, DPA-ProTM, AustralisTM]. For this research camera calibration was processed in VMS (Geometric Software).

3.2.3.1 Image networks

A set of cameras viewing an object is termed a network, the geometry of which is a key consideration for accurate measurement. Camera placement must be carried out with a careful design in order to cover the area of interest with a minimum number of cameras. The network geometry can be negatively affected if the relationship between an object point and its associated image point cannot be visually established. Automatic sensor planning in the computer vision community can be found from [Tarabanis *et al.,* 1995; Triggs *et al.,* 1995; Yi *et al.*, 1995], while motion planning [Latombe, 1991], and image-based modelling [Fleishman *et al.,* 2000] have also received a lot attention. The common goal of these tecniques is to derive features of a target object, usually static with known 3D coordinates, with a minimum number of imaging sensors. Mason (1995) proposed that the construction of a quality metric should be seen as an important step for automatic 3D photogrammetric measurements, where a quality metric includes the target geometry explicitly to evaluate various camera configurations. For example, Mason and Gruen (1995) developed a work called CONSENS which consisted of multiple cameras used for optical triangulation. The method describes an ideal configuration of four camera stations. Some research works in terms of uncertainty analysis for camera placement have also been reported. For example, a scalar function of the covariance matric was used by Olague *et al.* (1998) as the uncertainty measure. An ellipsoid fitted to the projected error pyramids was proposed by Wu (1998) in order to

estimate the 3D volume undertainty. These works of uncertainty analysis stressed that limited image resolution is the primary factor leading to 3D coordination uncertainty. However, occlusion has also been frequently reported as another source of error in featured-based multiple camera measurements.

Various studies in close-range photogrammetry [Fraser, 1997; Clarke *et al.*, 1994; Gruen and Beyer, 2001; El-Hakim *et al.*, 2003] have suggested the following factors which affect the accuracy of measurement:

1. Base-to-depth ratio

   Normally the higher the ratio, the higher measurement accuracy the network can achieve;

2. Convergent image network

   A convergent network increases the overlapping area between images, as compared with parallel images. Therefore object occlusion can be minimised;

3. Number of images of a target

   Measurement accuracy of point normally increases significantly with the number of images containing this point. However, such increase usually becomes less significant when the number exceeds four;

4. Number of measured points per image

   For a relatively weak imaging network containing targets which are not well distributed within the field of view, the measurement accuracy normally increases with the number of measured points in each image;

5. Image resolution

   The accuracy of 3D coordination is closely related to the image resolution, especially for natural featured targets.

Factors which affect camera calibrations are:

1. Imaging network

   Self-calibration is reliable only provided with a strong network geometry, i.e. each camera contributes multiple images in the network but not just one and the imaging network contains multiple convergent images;

2. Object point array

   A large number of target points which are spatially well distributed through image formats are normally desired. Moreover, a 3D point array is normally much more preferred than a flat 2D array since the 3D distribution is able to give a maximum variation of image scales on image plane and therefore likely to correctly recover the principal distance;

3. Image rotations

   To prevent the shifting of errors from interior (principal point) to exterior (perspective centre) orientation parameters, at least two images per camera in the network should be rotated by 90 degrees;

4. Datum definition

   The datum of a photogrammetric bundle adjustment can be defined by the point arrays with know coordinates (externally constrained). However, any errors from the determination of these coordinates will propagate into datum definition and therefore the imaging network.

Network design plays an important role to increase the accuracy and reliability of photogrammetric measurement. This involves deciding how many cameras are needed, where the cameras should be located to have a good imaging geometry. The network configuration defines the imaging geometry and determines the calibration quality. Detailed discussions of the network design and configuration in this research will be given in Section 4.5.2.

3.2.3.2 Self-calibrating bundle adjustment

The collinearity equations, described in Equation 3-1, can be simplified as:

$$x - x_0 = -c \cdot f_x(x, y)$$

$$y - y_0 = -c \cdot f_y(x, y) \qquad 3 - 6$$

where $(x - x_0, y - y_0)$ represents the coordinates of image relative to the principal point.

The above collinearity equations extended to include any possible systematic errors are called a self-calibrating bundle adjustment [Gruen and Beyer, 2001], where systematic errors are usually described by corrections of the image coordinates expressed as a function of additional parameters. The additional parameters used in this research have been described in Section 3.2.2.4.

The extended collinearity equations can be expressed as:

$$x - x_0 = -c \cdot f_x(x, y) + \Delta x$$

$$y - y_0 = -c \cdot f_y(x, y) + \Delta y \qquad 3 - 7$$

A self-calibrating bundle adjustment is a statistically rigorous process to estimate all the interior, exterior orientations as well as target coordinates, as compared with only exterior orientation and target coordinates estimated by the conventional bundle adjustment. The self-calibration bundle adjustment usually demands a strong image network. Equations 3-7 are the observation equations estimating the unknown parameters and can be described as:

$$l = f(x) \qquad 3 - 8$$

which relates the image observations l to the parameters x.

Since the number of observations is much more than the number of parameters to be estimated, these equations are usually non-linear. Initial approximated values to unknown paramters are therefore necessary in order to linearise these equations [Luhmann *et al.*, 2006]. After the first order Taylor expansion, Equation 3-8 can be expressed as:

$$l - e = Ax \qquad 3\text{-}9$$

where:

l represents the observations/measurements;

e is the residual between the measured and computed values;

A is a $n \times u$ design matrix (where n represents the number of observations and u represents the number of unknowns).

The estimation of the vector of unknowns $\hat{x}$ is usually performed iteratively with minimum variance estimation of least squares [Luhmann *et al.*, 2006], leading to a system of normal equations:

$$\hat{x} = \left(A^T PA\right)^{-1}\left(A^T Pl\right) \qquad 3\text{-}10$$

where P is the weight matrix of the observations. For most practical applications P has the form $\sigma^2 I$ where σ^2 is the variance of image coordinate measurements. The non-linear bundle problem is therefore solved as a sequence of linear problems: at each iteration, a correction vector $\hat{x}$ is first estimated and added, and the process is repeated until the solution is converged. The internal adjustment accuracy is given by computing the residuals v of the observations and the variance factor $\hat{\sigma}_0$ as shown in Equation 3-11 and Equation 3-12:

$$v = A\hat{x} - l \qquad 3\text{-}11$$

$$\hat{\sigma}_0 = \sqrt{\frac{v^T Pv}{n-u}} \qquad 3\text{-}12$$

Moreover, let

$$Q_{xx} = \left(A^T PA \right)^{-1} \qquad 3 - 13$$

where 3-13 is the inverse of the normal equation matrix. The symmetric covariance matrix K_{xx} can be derived by:

$$K_{xx} = \hat{\sigma}_0^{\ 2} Q_{xx} \qquad 3 - 14$$

Precision measures are computed from the covariance matrix which takes every change in the network configuration and every model variation into account [Gruen, 1978]. The function of the diagonal elements of K_{xx} is:

$$\hat{\sigma}_{kk} = \hat{\sigma}_0 \sqrt{q_{kk}} \qquad 3 - 15$$

which represents the standard deviation of each adjustment unknown while elements in position ij ($i \neq j$) represent the covariances between the unknown x_i and x_j . A precision evaluation of the estimated 3D coordinates can then be derived from the traces of the corresponding covariance matrices. The Q_{xx} matrix also gives information about the correlations among the unknown parameters from the correlation coefficient ρ_{ij} [Luhmann *et al.*, 2006]:

$$\rho_{ij} = \frac{q_{ij}}{\sqrt{q_{ii} \cdot q_{jj}}} \qquad 3 - 16$$

If the absolute values of the correlation coefficient ρ_{ij} is close to 1, there is high correlations between x_i and x_j . And if highly correlated parameters exist, one of the two must be removed from the bundle adjustment.

3.2.3.3 Further considerations on the additional parameters

In a self-calibrating bundle adjustment, the additional parameters can be introduced as:

1. Camera-invariant

One set of additional parameters are used for all images, independent of camera focus settings and number of cameras used;

2. Frame-invariant

One set of additional parameters are used for each image. This method is usually used by a network with different cameras or a network with individual images acquired with different focus settings, in applications such as robotics, machine vision inspections;

Case 2 is usually considered [Tecklenburg et al., 2001] as it is not realistic to assume that the camera parameters are the same for all cameras, especially for consumer digital cameras where such variations across cameras might arise from the manufacturing process. This approach also allows the derived lens distortion parameters suited to the particular focus setting of each individual image to be used for multi-camera 3D coordination. Therefore this approach is adopted in this research. However, non-determinable parameters (over-parameterisation) can lead to network adjustment singularities caused by parameter correlations. Regarding this, a 3D object point array with control points in place is used in this research to avoid over-parameterisation [Remondino and Fraser, 2006]. Tests on additional parameters are always necessary if the imaging network is not strong [Gruen, 1976; Gruen, 1981]. Inclusion of insignificant additional parameters is likely to increase the possibility of error shifting between parameters resulting in instability of the entire bundle adjustment estimations. Therefore the camera model needs to be refined during the calibration process to include only significant parameters (Section 3.2.2.4).

Usually the determinability and the significance are tested on the additional parameters:

1. Correlation among the parameters (Equation 3-16)

This is the most common approach. The additional parameters usually have high correlations among themselves or with the camera parameters. Since the principal point offset is correlated to projection centre $\left(X_0, Y_0, Z_0\right)$ while lens distortion is correlated to the principal distance and rotation elements, the system errors might not be well reflected from these additional parameters if the errors are transferred to those exterior orientation elements. Therefore statistics analysis is necessary to find out any highly correlated parameters. Generally any correlation higher than 0.95 should be eliminated [Remondino and Börlin, 2004];

2. Statistical testss

Normally the Student's test is applied, with the null-hypothesis that "the additional parameter x is not significant" compared to the alternative hypothesis "the additional parameter x is significant":

$$t = \frac{\hat{g}_i}{\hat{\sigma}_i} = \frac{\hat{g}_i}{\hat{\sigma}_0 \sqrt{q_{ii}}} \qquad 3 - 17$$

where $\hat{g}_i$ is the estimated value of parameter i; $\hat{\sigma}_i$ is the standard deviation of the parameter i.

Student's test is a one dimensional test valid only if the tested parameters are independent. Gruen (1981) presented a stepwise procedure to check the determinability and significance of the additional parameters during different stages of the least squares adjustment. Covariance matrix is checked to determine if any additional parameter(s) are undeterminable so that subsequent elimination can be performed on such parameters.

3.3 Consumer grade cameras and their calibration

Based on Brown's report in 1976 (Section 3.2.2), non-metric cameras have received significant attention in terms of their photogrammetric applications in the last three decades. The evolution of digital cameras has also provided these cameras with higher image resolution but lower price. However, since these cameras are not primarily design for metric use [Fryer *et al.*, 2007], their stability in terms of photogrammetric measurements might not be as high as those metric cameras due to features such as autofocus [Rieke-Zapp and Peipe, 2006].

Camera stability has been extensively investigated in the last two decades. For example, Physical behaviour of the principal point location was studied and different calibration parameters for Kodak DCS420 and DCS460 digital cameras were compared with each other [Shortis *et al.*, 1998]. These investigations led to a conclusion that systematic error introduced by principal point offset can be easily caused by the physical movement of the camera. This result has been well supported by Sanz-Ablanedo *et al.* (2010) where non-negligible systematic variations of the internal geometry were found when cameras are rotated. Similar studies can also be found such as from Shortis *et al*. (2001) where the stability of principal point location was determined, Mills *et al.* (2003) where the principal point location, principal distance and distortion parameters were analysed. Results from studies by Shortis *et al.* (2006) with different digital SLR cameras demonstrated that a fixed lens is able to provide a much more rigid internal imaging geometry and therefore much better results in terms of the accuracy, precision and stability of photogrammetric measurements, as compared with zoom lens. The geometric stability of seven Nikon Coolpix 5400 cameras was studied by Wackrow *et al.* (2007) over a time interval of one year, where results demonstrated that this camera model is able to achieve 3D coordination accuracy of 1.4mm within a measurement distance of 1.5m.

In particular, the geometric stability of six Pentax Optio A40 cameras was investigated by Sanz-Ablanedo *et al.* (2010). Results demonstrated that the variations of internal imaging geometry of these cameras during or between usages are in the similar order of

magnitude compared with cameras of medium or low accuracy in photogrammetric applications, such as the 3µm variation of principal distance in a D7 Rollei metric camera with a 7mm lens [Peipe and Stephani, 2003] and in a Kodak DCS 460 with a 24mm lens [Läbe and Förstner, 2004]. Similar to the results presented by Shortis *et al.* (2006), Sanz-Ablanedo *et al.* (2010) reported a major systematic change of ~3.6% of the principal distance after zooming was applied. Regarding these findings, Sanz-Ablanedo *et al.* (2010) stressed the necessity of re-calibration if there is any change of the zoom lens. However, such necessity is dependent on the accuracy requirement of a particular photogrammetric measurement application.

For close range photogrammetry, a "camera" is any imaging device that generates a central perspective and has a consistent internal geometry [Atkinson, 2001]. The internal geometry of a camera is defined by the camera calibration, which is only consistent if the camera is at the same focus and zoom settings. The camera is effectively a different imaging device defined by a new calibration if its focus or zoom changes. Therefore for continuous coordination purpose, cameras are required to be either manual focus or fixed focus and the zoom settings should be turned off to ensure the camera parameters from the calibration process are suited to the cameras subsequently used for 3D coordination.

3.3.1 Considerations of off-the-shelf low cost webcams

Off-the-shelf camera systems incorporate low cost components, shown for example by a comparison of internal components from a Kodak DCS420 SLR camera and a Logitech C500 webcam (Figure 3 - 6). For the Logitech C500 webcam which was used in this research, the internal components except its logic board comprise of a case made from two half sphere-shaped pieces of plastic with four holes with screws for connection. Three triangle-shaped plastic rods are used to fix the logic board in position. The image sensor is soldered onto the C500 webcam logic board and its lens and colour filter are fixed in position in front of the sensor by two screws.

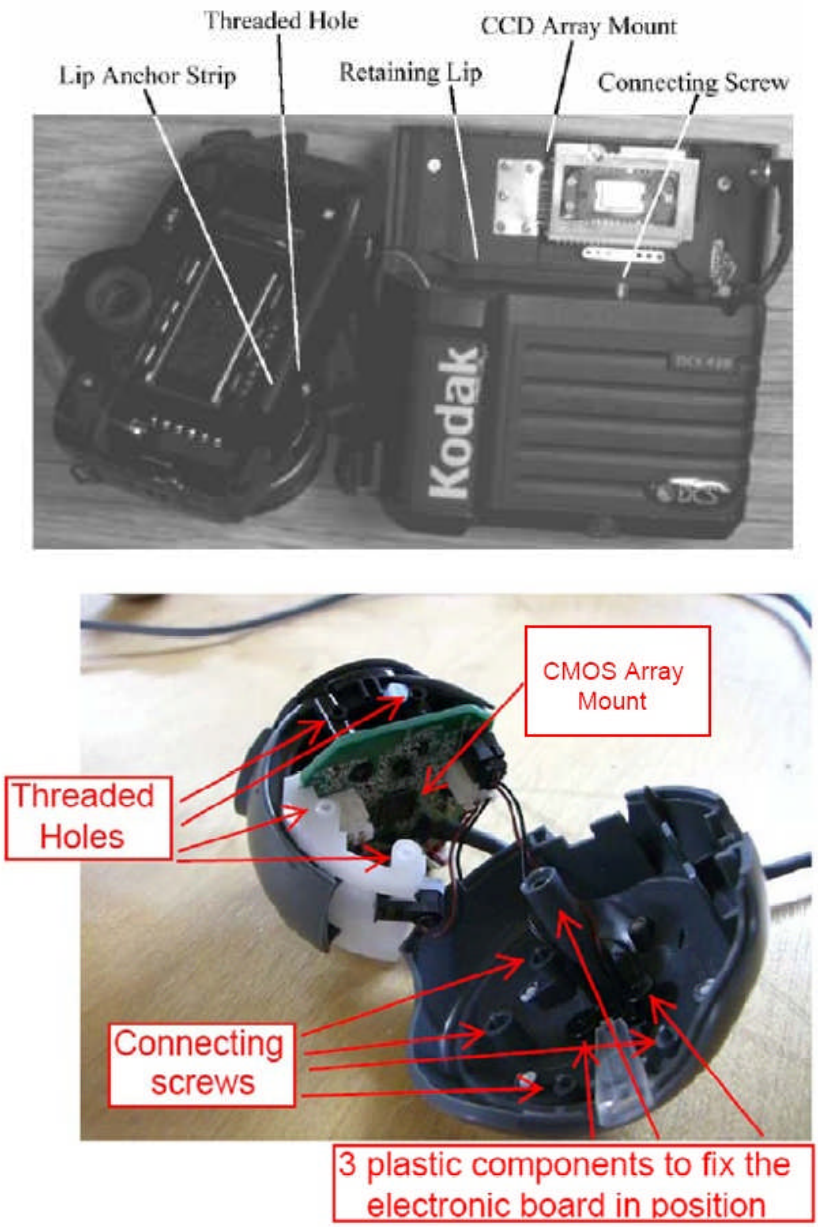


Figure 3 - 6. *Internal structure of Kodak DCS420 (upper, reproduced from Shortis and Beyer (1997)) and Logitech C500 webcam (lower).*

The simple construction of low cost webcams, such as the C500, could be an advantage as the sensor and lens are physically connected when compared with other digital cameras such as Kodak DCS420 which use a combination of springs and hinges. This might mean that the internal geometry of the webcam sensor is good whilst the exterior orientation of the webcam is less stable due to the physical stability of its housing. If this is the case, maximising its exterior orientation stability is critical in order to deliver 3D coordination with acceptable level of accuracy. To do so no handling or rolling of the camera should be allowed by mounting the camera on a tripod or similar support [Shortis and Beyer, 1997]. This concept has been followed in our hardware

design of the multi-webcam coordination system where we have re-engineered the mechanical mounting.

Shortis and Beyer (1997) also stressed that the more important issue for continuous use of cameras is that the calibration is consistent over the short to medium term. Following this concept, the testing and analysis of the developed system (Chapter 4) will emphasise the significance of variations of calibration parameters (with respect to 3D coordination accuracy) over periods of days and months, rather than individual exposures, since such quantification of the level of variation would define guidelines for the frequency with which online systems must be re-calibrated [Shortis and Beyer, 1997].

The discovery of digital sensor technologies has allowed the microscopes to function as quantitative measurement tools for applications from life sciences [Koenig, 2001] to engineering studies [Jordan *et al.*, 1998]. Pollak and Hutter (1998) reported their method for mounting a digital webcam on an optical microscope, effectively providing a low-cost version of digital microscopy. Recently, webcam-based microscopes, where webcams are directly turned into digital microscopes, have been developed and a systematic explanation and quantitative assessment of the imaging performances of webcams have been provided by Parikesit *et al.* (2010). In particular, Parikesit *et al.* (2010) compared the imaging performance of a US$30 Prolink webcam (sensor: CMOS; 1280×1024 pixels; interface: USB 2.0; maximum frame rate: 30 fps; lens focal length: 2.8mm), which has similar specifications with the Logitech C500 webcams used in this research, with a US$100 QX5 computer microscope (640×480 pixels) which has been previously employed in a scientific investigation [Chen and Ugaz, 2006]. Parikesit *et al.* (2010) concluded that the reproduction of the inverted lens microscope from the webcam provided the best performance, where the spatial resolution (6.9μm) is twice better than the QX5 microscope (12.4μm, at 200× magnification). This spatial resolution also allowed the inverted-lens microscope to resolve translucent objects with a translucent background, for a spatial periodicity of up to 10μm, which could be relevant for life sciences investigations on unstained biological cells. Also highlighted in this work are two webcam limitations. First, there is no diffuser in webcams resulting in non-uniform illumination intensity across the whole field of view; second, optical

aberrations (both chromatic and radial distortions) have been found in the webcam because of the lens quality, which is inferior to the lens used in the QX5 microscope which did not display such aberrations. Where the first problem can be solved by using a well-positioned diffuser, the second problem can be solved using digital image processing such as [Kozubek and Matula, 2001; Yu, 2003].

Experiments with a Creative II webcam by McNamee *et al.* (2001) addressed the convenience of non-metric cameras in photogrammetry applications due to the evolution of digital cameras nowadays has brought higher performance but lower price for these cameras. However, similar to the report from Brown (1976), McNamee *et al.* (2001) also stated that accurate recovery of camera parameters is required in order to achieve accurate photogrammetric measurements. Page *et al.* (2008) addressed that errors existed in 2D coordinate measurements (Logitech Pro 4000 USB webcam) are essentially systematic caused by lens distortion. Lens distortion modelling due to the lens quality of an ordinary USB (TerraCAM Pro) webcam has also been address by Pervölz *et al.* (2004). A small random error of locating an object point on image was also reported, where these errors are very small and are of the order of 0.1 pixel when markers are used [Page *et al.*, 2008]. In particular, Page *et al.* (2008) suggested it is possible to have coordination accuracy of 0.3mm with a measurement distance of about 300mm, provided that all sources of error have been considered. This result is well supported by the result in this research with the same (Logitech Pro 4000 USB) webcam (Section 4.2.3.4).

Within computer vision, most webcam research is based upon a single webcam with a simple calibration method. 2D/3D coordination precision and accuracy, as well as stability and reliability of a system of webcams in terms of continuous 3D measurement have not been fully investigated and reported to the author's best knowledge within the computer vision domain.

Sources of error regarding image measurement precision are not easy to identify from acquired images since there can be various causes such as noise generated from electric voltage in individual sensor pixel. Several authors [Dahler, 1987; Burner *et al.*, 1990; Beyer, 1993; Robson *et al.*, 1993] discussed limitations to target measurement

precision. Therefore in order to reliably deliver 3D coordination with acceptable accuracy and precision from webcams, critical factors including the imaging sensor, image compression, software image interpolation, chromatic aberrations and frame rate adopted in current webcams are necessary to be taken into account and discussed. In particular, the current techniques of image noise reduction for CMOS sensor used in high end digital cameras, where these techniques might be broadly applied to webcams in future, are presented as well.

3.3.1.1 Electronic sensor

The electronic sensors used in consumer grade digital cameras are either Charge Coupled Device (CCD) or Complementary Metal-Oxide Semiconductor (CMOS). While because of the semiconductor, the manufacturing process, energy consumption and cost for CMOS are all lower than for CCD sensors of equivalent resolution, CCD is better than CMOS in terms of noise depression and colorization. There has been on-going research recently on CMOS conducted by Canon which has effectively reduced noises in CMOS. One typical example is Canon EOS-1D Mark IV [Canon, 2009]. Different approaches, both hardware and software, have been simultaneously implemented by Canon in order to increase the signal-to-noise ratio. First, image processing functions were implemented on electric voltage data (analog) to reduce the noise prior to the analog-to-digital conversion. This method has an advantage that the noise existed in analog data should be more preditable and therefore easier to correct since it has not been propagated through the conversion process. Second, Canon completely separated the digital and analog circuitries in their manufacturing process so eliminate the possibility of coupling of the current spikes from digital circuitry back to the analog one, where both fixed and random pattern of noise can be generated by this effect. Third, Canon largely increased the light gathering area of each individual pixel to compensate the conventional disadvantage of CMOS sensor with relatively smaller light gathering area as compared with CCD sensor. The signal-to-noise ratio under such change can be largely increased. Finally, Canon increased the light convergence capability by increasing both the size and the number of microlens used to direct light to the light sensitive area within each pixel. Though such change of microlens would greatly increase the image quality, the cost behind it is believed to be high.

Compared to Canon's digital SLR cameras, webcams are capable of containing electronics which can support image compression, sharpening and resizing. However in the absence of manufactures data it is only possible to guess at embedded capabilities from the resultant images.

3.3.1.2 Image compression

Logitech officially stated that the compression format for C500 webcams is M-JPEG/YUY2, where these compression techniques are summarised as follows:

M-JPEG

M-JPEG stands for Motion JPEG and is used by many digital cameras which are able to capture videos. Each video frame is individually compressed as a JPEG image by discrete cosine transform (DCT). Such a lossy form of transformation from the time doman to the frequency doman would therefore loose some image details due to the quantization of DCT transformation. Furthermore, image information with low frequency tends to be filtered out with increasing compression. One typical example of such effect can be found from image areas with smooth intensity changes, where these areas usually appear 'blocky' after such compression is applied. However, the advantage of this compression scheme is its high speed of processing and low hardware requirements, which has a potential for the photogrammetric measurement based on each video frame.

YUY2 (Y'UV422)

YUY2 is developed based on the higher sensitivity to luminance than to colour from the visual perception of human. Y' represents the brightness; U and V represent the colour. The difference between Y and Y' is that gamma compression is applied to change Y from the brightness perceptually to Y' which represents brightness by is an electronic voltage. YUY2 implements chroma subsampling where the colour (U and V) resolution is only half of that of brightness (Y). 4:2:2 has become the most common ratio for Y, U and V under such compression scheme.

RGB24

RGB24 is a device-dependent colour model and it is mainly for image representation and display. The number '24' implies that there are a total of 8 bits of information for each colour component in each pixel. Since there are three colour components, there are a total of 16,777,216 ($2^{8\times3}$) combinations of R, G and B values for each image.

Take video streaming from a digital webcam as an example, suppose YUY2 is used for video compression (encoding). Such compressed video data needs to be decoded to RGB model for displaying of video frame on the monitor. Such a decoding process is performed by the electronic board of the webcam itself. Computer CPU then writes bytes into the frame buffer inside the video memory to represent the RGB value of each pixel. These RGB signal values are finally converted to pixel intensity values on computer monitor through gamma correction performed by the electronic video generator. Equations 3-18 and 3-19 are the common formulae for encoding/decoding between YUV and RGB [Keith Jack, 2007]:

RGB to YUV:

$$Y = (0.257\times R)+(0.504\times G)+(0.098\times B)+16$$

$$C_r = V = (0.439\times R)-(0.368\times G)-(0.071\times B)+128$$

$$C_b = U = -(0.148\times R)-(0.291\times G)+(0.439\times B)+128 \qquad 3\text{ - }18$$

YUV to RGB:

$$B = 1.164(Y-16)+2.018(U-128)$$

$$G = 1.164(Y-16)-0.813(V-128)-0.391(U-128)$$

$$R = 1.164(Y-16)+1.596(V-128) \qquad 3\text{ - }19$$

Different RGB colour models are usually used for CMOS and CCD image sensors. The camera driver for Logitech QuickCam Pro 4000 webcams is PWC (Philips USB webcam) while the driver for Logitech C500 webcams is Logitech's UVC (USB Video Class). According to Logitech official documentation, the UVC in C500 webcams support Bayer

mode (both colour processing and format conversion are disabled resulting raw output) with some widescreen resolutions not supported.

The above compression techniques used in webcams are closely related to the issues discovered in this research, where attaching multiple cameras to a system would result in automated changes in compression (Section 4.4.3.3 and 4.5.5.3).

3.3.1.3 Image interpolation

Image interpolation happens when an image is resized or remapped from one pixel to another. Such interpolation is usually performaed to estimate the intensity value of a pixel (unknown) from intensity values of its surrounding pixels (known). Image interpolation generally can be divided into non-adaptive and adaptive interpolations.

For non-adaptive interpolation, all pixels are equally important and such interpolation method is independent of edge or texture. Among non-adaptive algorithms, nearest neightbour interpolation is the simplest since the estimation of a pixel is purely based on its closest pixel. Bilinear interpolation uses information from a total of four surrounding pixels to estimate the interpolated pixel, where a weighted average is usually used to derive the estimated value since the distance between any surrounding pixel and the interpolated pixel is the same for all surrounding pixels. Similarly bicubic interpolation uses a total of sixteen surrounding pixels, together with different weight based on relative distance between the surrounding pixel and the interpolated one. Among these three algorithms, bicubic method is able to produce the sharpest image because it takes the most information available into its estimation process.

Adaptive interpolations are usually found in licensed softwares such as PhotoZoom Pro. The underlying principle of these interpolations is minimising artefact details (such as sharp edges) while maximising artefact-free details when an image is enlarged.

3.3.1.4 Chromatic aberration

Camera lenses consist of several dispersive glasses which do not refract all constituent colours of incident light at equal angles [Sidney, 2002]. Chromatic aberrations describe such departures from perfect imaging due to dispersion in glass. While longitudinal chromatic aberration (axial colour) describes the change of the lens's point of focus for different wavelengths (left, Figure 3 - 7), transverse chromatic aberration (lateral colour) describes the change in a lens's magnification for different wavelengths (right, Figure 3 - 7) [Woodlief and Whiley, 1973].

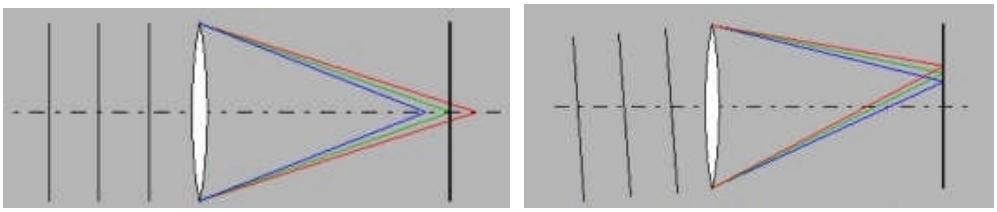


Figure 3 - 7. *Longitudinal (left) and transverse (right) chromatic aberrations.*

Table 3 - 1 compares the two chromatic aberrations:

Longitudinal chromatic aberration	**Transverse chromatic aberration**
Only observed in places with high contrast	
Causes fringes all around objects	Only affects tangential details
Occur at any position in image	Absent in image centre and progressively worsens towards image corners
Yield fringes of a single colour	Yield two differently coloured fringes at either side of the tangential structure

Table 3 - 1. *Features of longitudinal and transverse chromatic aberration.*

Fringes due to axial colour are particularly visible in objects which are just out of focus, e.g. green in front of the object and purple behind it (Figure 3 - 8). For lateral colour, colour components with different wavelengths are focused onto different positions along the image plane (Figure 3 - 8). Both axial and lateral colours degrade the overall resolution of a lens [Bruce, 1978].

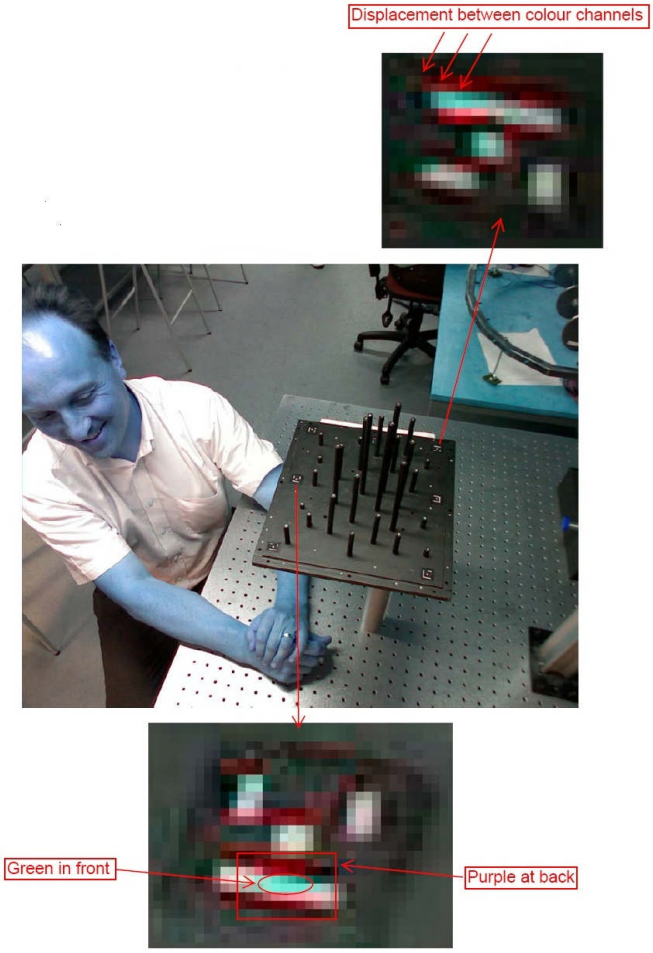


Figure 3 - 8. *Chromatic aberrations observed in webcam (Logitech C500) images.*

The shape of a blurred circle is defined by the shape and number of a lens' diaphragm blades [Zeiss, 2010]. When targets are not exactly focused on the image plane, these image spots are no longer circles but appear as shapes with as many sides as blades. The more blades, the better since they will give a closer approximation of a circle. While 35mm SLR lenses usually have six blades, seven blades are equipped in most Nikkor SLR lenses and nine blades can be found in Lenses for Nikkor telephotos. However, for consumer grade digital cameras such as webcams, low cost lenses with only a few blades are used, where a lens with four blades for Logitech C500 webcams can be observed from blurred circles (Figure 3 - 8) appearing as squares.

Nowadays a monochrome off-the-shelf digital camera rarely exists and therefore new calibration issues arise with colour cameras in terms of higher-accuracy photogrammetric measurement. While it is difficult to reduce the blur effect caused by axial colour, a degree of misregistration of the colour channels caused by lateral colour can be corrected from post-processing steps [Cronk *et al.*, 2006].

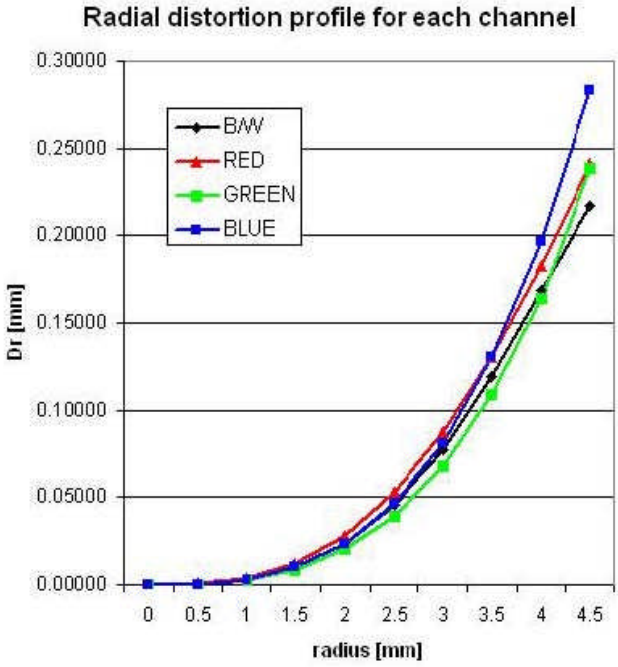


Figure 3 - 9. *Radial distortion of 3 colour channels and their associated B/W image. [Remondino and Fraser, 2006]*

The effect of chromatic aberration was investigated by Remondino and Fraser (2006), with a SONY DSC F828 digital camera (8 Mega pixels and a focal length of 10mm). The distortion profiles (Figure 3 - 9) demonstrates that while green colour has the smallest discrepancy as compared with the B?W image, the discrepancy from blue colour can be as large as ten pixels towards the sensor edges.

3.3.1.5 Frame rate

Although Logitech states on its official website that the frame rates for both Pro 4000 and C500 webcams are up to 30 FPS (frames per second). In its official documentation, however, Logitech states that for one C500 webcam, the maximum frame rate with maximum hardware (video) resolution (1280×1024) in M-JPEG format is 15 FPS, while the maximum frame rate with maximum hardware resolution in YUY2 format is only 7.5 FPS. Considering the extremely high cost needed for high resolution M-JPEG compression, whereas for low cost webcams like the Logitech C500 model, it is likely that the maximum hardware resolution of 1280×1024 uses YUY2 compression. It is also worth to note that due to the manufacturing cost, it is very unlikely for these webcams to have a very fast sensor to quickly update video frames.

3.3.2 Calibration methods

Appropriate methods such as targets are required to recover camera parameters. In this research a calibration frame with evenly distributed 3D object point array (Figure 3 - 10) is used to comply with the calibration guidelines outlined in Section 3.2.3.1.

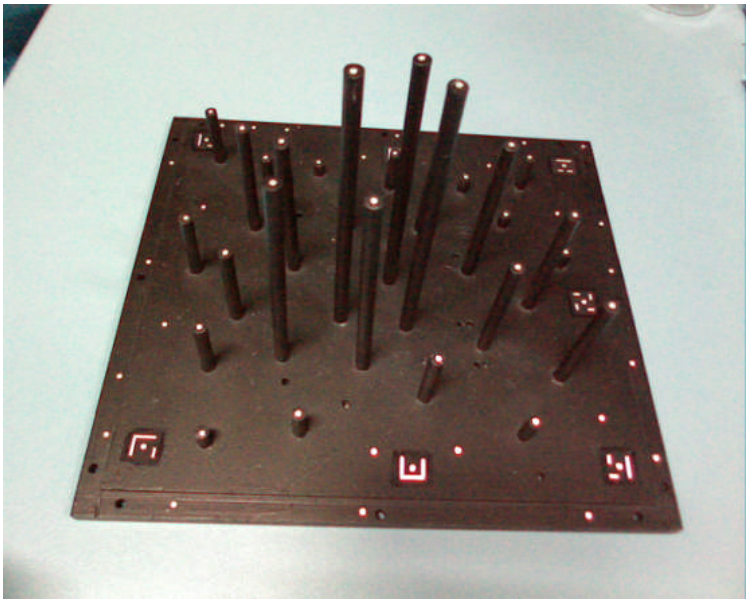


Figure 3 - 10. *Target test field used for camera calibration.*

Self-calibrating bundle adjustment (Section 3.2.3.2) is used to simultaneously estimate the locations and orientations of each image as well as a common set of camera calibration parameters. Robson *et al.* (1993) reported that the principal point offsets associated with their uncertainties can be determined easily and used as the initial approximate values for the subsequent bundle adjustment process. An efficient method to recover the principal point offset and principal distance is to establish a strong imaging network from a multiple camera convergent geometry around a calibration object such as that shown in Figure 3 - 10. Clarke *et al.* (1994) reported that in a single convergent network around the test field, unreliable or inaccurate estimates of some or all of the interior orientation parameters are likely to occur due to the fact that each camera contributes only one image in the network. This issue can be solved by multiple views of the test field produced by each camera. In this research 10 sets of views of the test field was achieved by rotating the test field each time a set of photos was taken by a series of cameras, i.e. a 4-camera setup produced 4×10=40 views.

3.4 Photogrammetry with targets

Although there have been a number of advances in the efficiency and effectiveness of image-based metrology systems in the last three decades, the most important factor has been the automatic detection, recognition, identification and measurement of artificial targets used to signalise points of interest [Shortis *et al.*, 2003]. Without the ability to recognise and measure targets automatically, the efficiency and response times of metrology systems are severely limited by the necessity of manual measurement of target images.

An efficient strategy for taget image detection is simply making targets either the brightest or darkest objects in the image [Shortis *et al.*, 1994]. Brown (1984) reported that good results were obtained with a combination of retro-reflective targets and lighting control in applications of industrial metrology. Shortis *et al.* (1994) stressed that since background information can be effectively minimised by under-exposure since it is normally out of interest in terms of photogrammetric target measurements. However, retro-reflective targets with controlled lighting conditions become impractical for most clinical studies, where either there exist various light sources such as those found in patient's bedside (left, Figure 3 - 11) or lighting needs to be minimised such as those found in optical topography studies (right, Figure 3 - 11 and also Figure 3 - 17).

Figure 3 - 11. *Typical clinical environment. Left: Various light sources around patient's bedside; Right: Dim to no light required for clinical optical topography studies.*

Shortis *et al.* (1994) reported that due to the fact that the photogrammetric target measurements usually demand subpixel positioning accuracy, the natural features on the object to be measured as well as the limitation of pixel resolution of imaging sensor has made precise measurement of natural features rather difficult. The positioning accuracy can be further decreased if there is a slight difference of feature positions between the fields of view of different cameras in the network, or even worse when there is no feature being identified [Clarke, 1994]. For example, there are no identifiable features on the optical sensing pad which is made of black coloured plastic (left, Figure 3 - 12), while light coloured targets form a relatively high contrast with the sensing pad (right, Figure 3 - 12) making each target an individual distinguishable feature.

Figure 3 - 12. *Features on an optical sensing pad without targets (left) and with targets (right).*

3.4.1 Target requirements for digital photogrammetry

Accuracy is the comparison of a measurement against a standard. It indicates how close a measurement is to the standard value. On the other hand, precision (or uncertainty) indicates the repeatability of measurements [Atkinson, 2001]. A small precision value suggests a small volume containing the actual point location. Photogrammetric target coordinate precision values are least squares estimates computed from the imaging geometry of the photogrammetric network in combination with the uncertainty of the target image measurements. Such information is internal to the network and hence the computed precisions do not reflect the accuracy of measurements [Robson *et al.*, 1993]. When precision is high but accuracy is low, it is highly possible that the fault lies with un-modelled characteristics of the camera or an incorrect definition of object scale. Therefore accuracy and precision are equally

important to access the capabilities of camera(s) in terms of photogrammetric measurements.

Since the target coordination uncertainty (in 3D space) is directly propagated from the uncertainty of target image positioning (in 2D image plane), the primary requirement for accurate measurement is the accurate target image measurement. The fact that a target image usually occupies between 25 and 100 pixels within an imaging sensor of a low pixel resolution has made such requirement difficult to fulfil. The precisions of the estimated target coordinates rely on many factors including the geometry of intersection (narrow intersection angles to targets should be avoided with angles closer to 90 degrees being preferred), the principal distance of the camera, the camera-to-target range and the number of images of each target (lines of sight). However, a well-designed target intersection should realise precisions which are predictable based on the following formula [Wang *et al.*, 2008]:

$$st = \frac{R}{PD} \times \frac{si}{\sqrt{n}} \qquad 3 - 20$$

where

st = predicted precision of the target coordinates;

PD = principal distance of the camera;

R = camera to target distance;

si = precision of the target image measurements;

n = average number of images for each target.

The subpixel precision of target image measurement has been extensively investigated in the last three decades. Clarke (1994) reported a precision level between 0.1 and 0.01 pixel by using bundle adjustment; Trinder (1989) and West and Clarke (1990) studied the precision of centroiding method while Förstner and Gülch (1987) and Gruen and Baltsavias (1988) studied the least squares method. All these studies led to a conclusion of a precision level between 0.2 and 0.005 pixels in terms of target image measurement, where such precision can be affected by size, shape and intensity distribution of the target (Figure 3 - 13). Brown (1984) recommended that the background should have at least five times radius than that of target itself in order to achieve a high target measurement precision. Brown (1984) also addressed that target

image diameters should be large enough in the images to optimise the target image measurement, where practical target image diameters can be computed using:

$$d = D \times \frac{PD}{R} \qquad 3 - 21$$

where

d = target diameter in the image;

D = target diameter in the object space;

PD = principal distance of the camera;

R = camera to target distance.

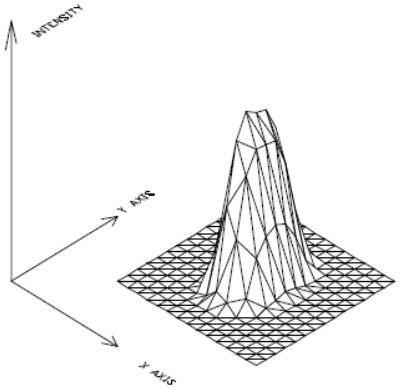


Figure 3 - 13. *3D view of a typical target image. [Clarke, 1994]*

3.4.2 Types of Target

Targets can be active or passive. Active targets are usually those with projected light spots, where passive targets are usually manually attached to the object to be measured [Clarke, 1994]. Retro-reflective targets have been widely used in photogrammetric applications for decades due to its high returned light intensity in favour of target identification. The advantage of retro-reflective targets includes high target contrast and much less influence of intensity variation from background in terms of target positioning. However, retro-reflective targets do have a limitation of axial illumination, where Clarke (1994) recommended a light radiation within 0.5 degree from camera axis and an angle within 45 degree between target and camera in order to achieve sufficient and consistent light return. Tests conducted by Clarke *et al.* (1995) confirmed this recommendation in which the returned light intensity can be largely decreased once the angles exceed the range stated above. However, target coverage

of the upper sphere of human scalp, plus that lighting conditions in many clinical studies (Figure 3 - 11), have rejected retro-reflective targets as an appropriate choice for targets used in this research.

Huang and Trinder (1993) discussed the advantage of spherical balls as being imni-directional, i.e. the target is a circle independent of viewing directions. Such advantage is confirmed by Aw and Koo (1993) in their photogrammetric measurement of a satellite dish. Moreover, the use of coloured targets can offer the advantage of unique identification by their spectral characteristics [Clarke, 1994].

Considering the target coverage on human scalp for clinical studies such as optical topography, planar circular targets can be attached onto locations which can readily be seen by multiple cameras overhead while spherical targets can be attached to places where occlusions can easily occur such as places inside hair or the horizontal level to ears provided that cameras are located above the head. Figure 3 - 14 shows a combination of planar and spherical targets used in optical topography studies in this research.

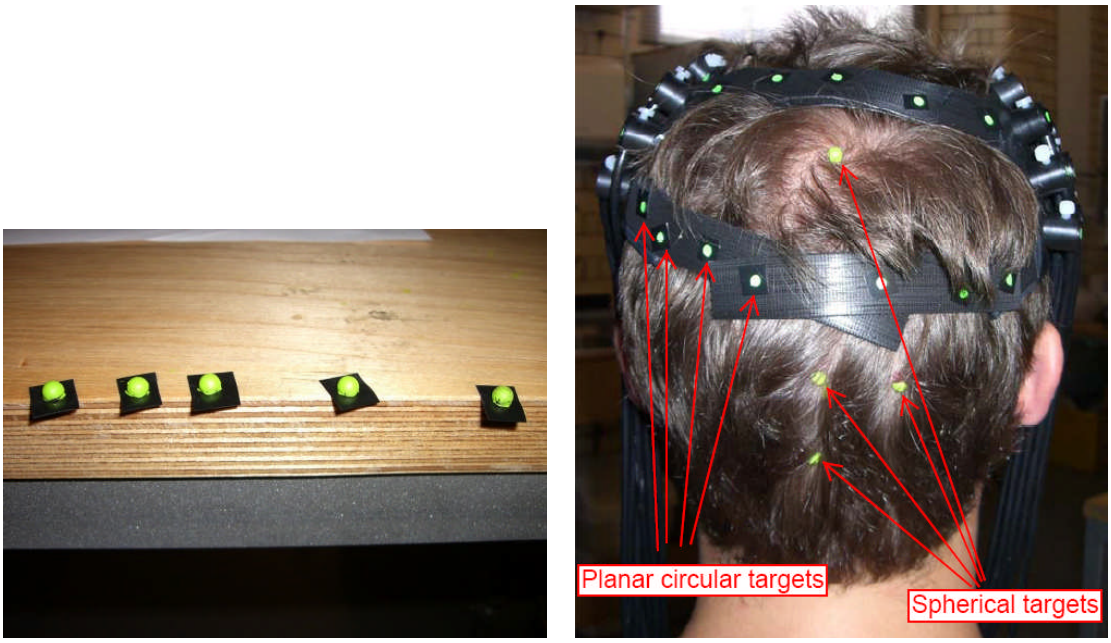


Figure 3 - 14. *Spherical targets (left) and a mixture of planar and spherical targets attached on human head (right).*

3.4.3 Target thresholding techniques

In order to extract target information from their background, thresholding is necessary not only to make targets more identifiable but also remove unnecessary image noises. For digital cameras especially those with relatively low sensor resolution, thresholding is usually considered as the first step to measure target images since in many cases target images are practically immeasurable due to a large amount of image noises resulted from background illumination, analog-to-digital quantization errors [Beyer, 1993], chroma subsampling of Bayer filtering process [Shortis *et al.*, 1993] etc.

Conventional thresholding techniques used in photogrammetry include both manual jusgement and automatic thresholding within a window containing a target. There are a number of window based thresholding techniques developed over the last two decades. The simplest approach is proposed by Snow *et al.* (1993) to add an arbitrary constant to the maximum intensity identified at the window edge. Wong and Ho (1986) achieved an efficient thresholding with passive targets in ambient lighting by setting the threshold value based on the average of the minimum and mean intensities of pixels contained in the window. A more complex but accurate method is based on the statistical analysis of intensity distribution for all the edge pixels of the window containing the target. Take the thresholding method used in VMS [VMS, 1999] as an example. Assuming the background noise in the image is random distributed, the mean and the standard deviation of the edge pixel intensities can be computed. Please note that in order to prevent the window's edge from containing target image pixels, if the mean computed is larger than a preset value, the window will automatically shift slightly until the mean computed falls within the tolerance. The threshold value is then set based on the maximum expected intensity value, e.g. 3×standard values+ the mean value. However, these statistical methods require either iterations of computation to derive the threshold or adjustment of window's position to prevent target intrusion into window's edges.

Each of the above techniques has its own characteristics, advantages and disadvantages. When considering a proper thresholding method suited to webcam-based fast coordination under complex scenes without the assumption of

appropriate lighting in this research, disadvantages of the above techniques such as the complexity of computation and high possiblity of a target image intrusion into the window edge make these techniques difficult to apply.

3.4.4 Practical target image measurement

In practice using prior knowledge of both the targets and the structure being measured can assist in the choice of detection algorithms [Chen and Clarke, 1992]. Conventionally in photogrammetric 3D measurement there are many equal sized targets and the background intensity changes slowly providing good contrast between targets and background (Figure 3 - 15).

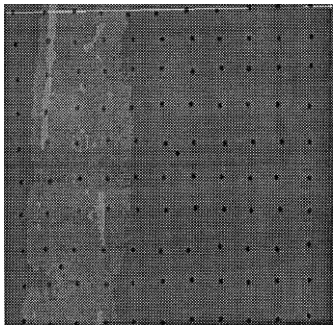


Figure 3 - 15. *Targets and background usually found in photogrammetric 3D measurement application (reproduced from Chen and Clarke (1992)).*

However, in this research two problems have been raised. First, in order to include the object within the views of all cameras so that a strong convergent geometry can be formed, the targets to be detected normally occupy only a small portion of each image from each camera. This has resulted in a scene containing a lot of background clutters where under-exposure is very easy to occur under this particular lighting requirement. Second, the sensitivity provided by low cost sensor embedded in these off-the-shelf webcams is much lower than those found in high-end digital cameras such as Nikon D100 SLR camera. These problems can be seen in Figure 3 - 16 and Figure 3 - 17, where most pixels are located in the dark area (left of the histogram) and some pixels located in the booming area (right of the histogram) indicating over-exposure. Background noise can also be seen in the large number of pixels situated in the low-intensity area. In contrast, little useful information is present in the middle of the histogram. Extremely uneven distribution of light intensities among image pixels delivered from these low cost webcams has demonstrated a very poor image quality.

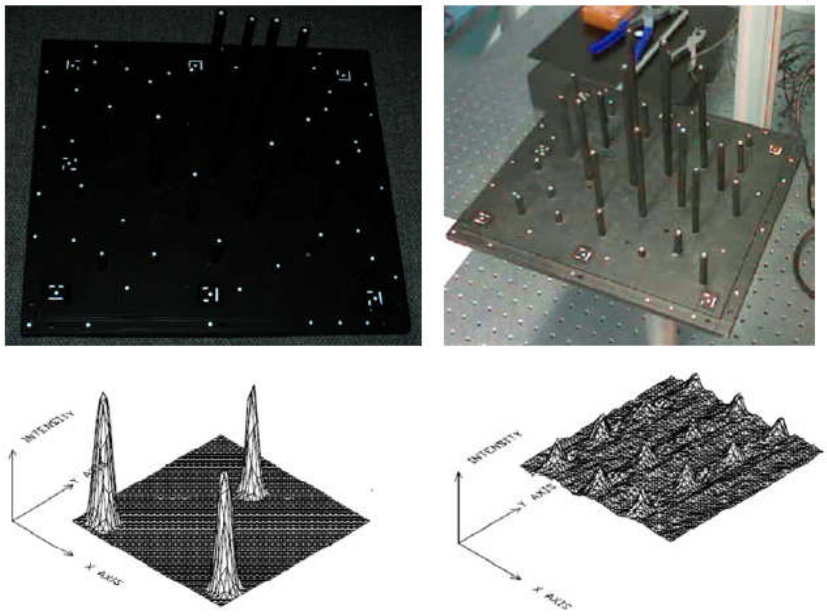


Figure 3 - 16. *Comparison of target image quality of Nikon D100 camera (left) and Logitech C500 webcam (right).*

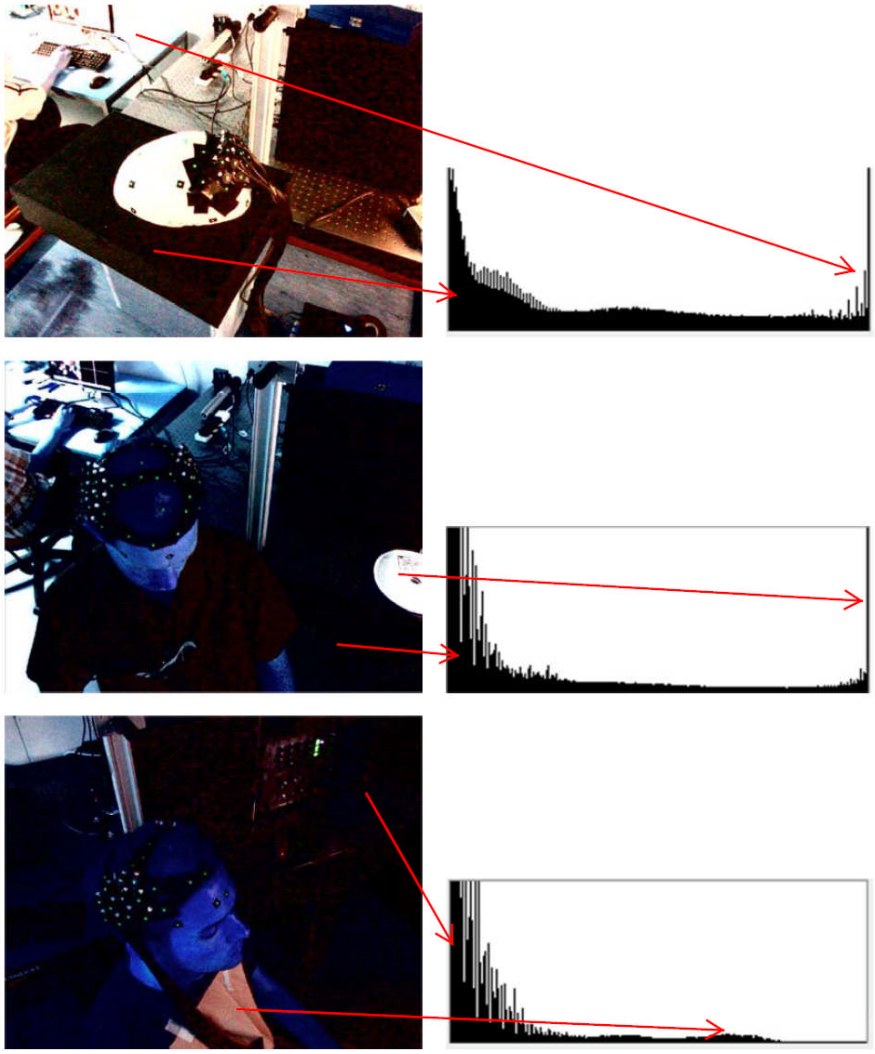


Figure 3 - 17. *Cluttered background in volunteer studies using off-the-shelf webcams.*

In this application low cost cameras are used with minimal lighting control to match clinical room lighting conditions. Retro-targets have been avoided to keep the equipment simple as no axial illumination is built into the cameras. As a result the combination of natural targets and the camera's inbuilt image compression poor image quality is the main obstacle for target detection.

Established photogrammetric targeting methods use thresholds to separate brighter targets from a dark background prior to subpixel target location computation using either a centroid or ellipse fit [Robson *et al.*, 1998]. However such detection methods are not only inappropriate when isolating target perimeters from complex background images such as those that might be found in this research, but also too much information from the images needs to be processed, i.e. all pixels with different intensity values within an image needs to be considered and processed for target recognition, tracing and feature extraction prior to target location computation. This is impractical for this application where 3D target coordinates need to be computed per frame and therefore the target recognition and location process here need to be as computationally simple and meanwhile effective as possible to allow enough time to be dedicated to the subsequent target correspondence matching and 3D space intersection method which derive 3D target coordination. Therefore target segmentation aims at rapidly extracting only useful information for target location calculation from sequences of images by discarding all unnecessary image information at this early stage in favour of a fast coordination application.

3.4.4.1 Target recognition - RGB to Grey scale Conversion

When a colour image comprising Red, Green and Blue channels is extracted from video stream (YUY2 to RGB24 conversion applies), every light element is represented from value 0 to 255 based on its light intensity and there are a total of 256 distinct colours for each light element. Therefore, a total of $256\times256\times256=16777216=2^{24}$ colours can be represented by these 3 light elements. Since every pixel is composed of a combination of the 3 light elements, every colour image is stored as 3×2D arrays, with each 2D array represents one light element, and the value of each 2D array represents the light intensity of each light element.

Consider the complexity and processing time for all 3 arrays, a simpler but accurate method is required for fast processing. Therefore the target measurement part of the system utilized grey scale images, in which the images captured from the webcams were transformed to grey scale first. The digital colour image is represented as RGB values, with a particular combination of proportions of red, green and blue colours. When a colour image is transformed into a grey scale image, normally the RGB values are transformed into the YIQ model, with Y representing luma information, I and Q representing chrominance information [Buchsbaum, 1975]. The transformation between YIQ and RGB is defined as:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad 3 - 22$$

Since the Y component and I, Q components are processed separately in YIQ, the Y value is the grey value in an 8 bit grey scale image (0 to 255), as indicated by Figure 3 - 18. The bigger the grey value, the higher the light intensity of the pixel will be. However, here we should note that some camera systems have 10, 12 or even 16 bit grey scales where the number of bits depends on the signal to noise ratio and dynamic range capabilities of camera's sensor.

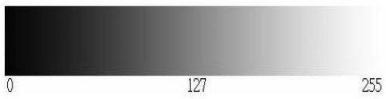


Figure 3 - 18. *The distribution of grey scale values. (Reproduced from Efford (2000))*

3.4.4.2 Target recognition - Segmentation

Given the imaging environment in this research, the acquired image is always accompanied by the complicated background which should be removed. For example in the experiment with a polystyrene head (Figure 3 - 19) carried out in the laboratory environment, of the background was not masked out with a sheet of background material. In this way the background complication is similar to the clinical environment. It must be emphasized that lighting control, for example the mains frequency related flicker of some lighting components, and other unstable factors caused by the dynamic capturing condition will increase the difficulty of background removal.

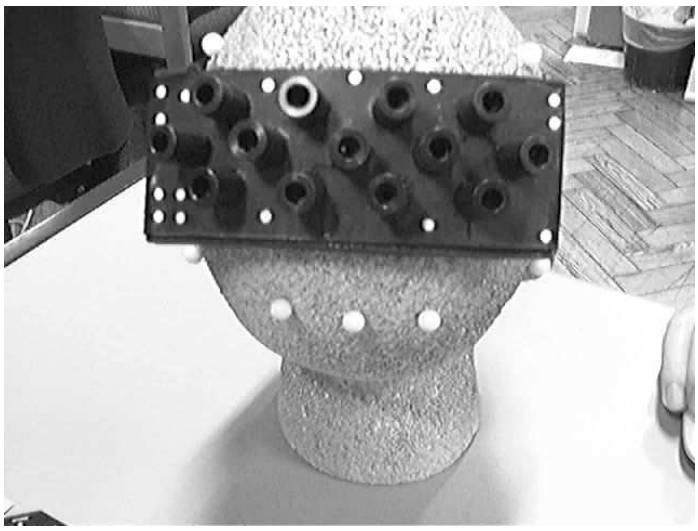
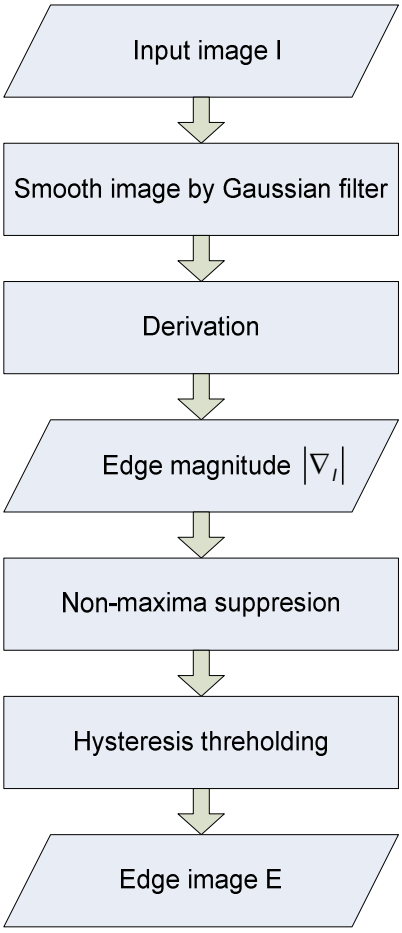


Figure 3 - 19. *Sample image (640×480) acquired by Logitech Pro 4000 in laboratory.*

Many systems use image thresholding to isolate targets from the background, however in the cluttered case, compounded by areas of high brightness such techniques do not perform well. In this research, reliability, effectiveness and speed are the most important criteria for target image segmentation from the background. An important function of the segmentation process is to act as the first operator to remove most noise in the images, which can speed up subsequent processes, for example by reducing ambiguities in the target image correspondence process.

One way of detecting features that could be target images is to use an edge detector. The Canny edge detector has been widely used since it was first introduced by Canny (1986). For these edge detection methods with fixed thresholds, the intensity variation across edge pixels can easily leave those edge pixels with relatively low intensity filtered causing edge disconnection. This drawback is effectively compensated by the Canny detector due to its fundamental difference from the conventional edge detection algorithms in that its hysteresis hypothesis is capable of reconnecting those disconnected edge pixels after the initial edge detection process. Luo and Duraiswami (2008) investigated the Canny detection optimised by the Open Computer Vision library [OpenCV] and reported that the processing speed of the operator making full use of the multiple cores and threads is much faster than any other implementation on a modern computer. The Canny edge detection process can be classified into four steps: smoothing, gradient derivation, non-maxima suppression, and hysteresis thresholding, as shown in Flowchart 3 - 1.



Flowchart 3 - 1. *Canny edge detector.*

1. Smoothing

Discrete representation and quantization errors (Section 3.3.1.2) can introduce false or noisy pixels in an image. A Gaussian filter mask can be used to remove noisy pixels. The Gaussian filter mask with a 3×3 kernel is defined by

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \qquad 3 - 23$$

2. Gradient derivation

A 2D convolution is performed on the image to determine the first derivative of each pixel gradient. Regions with sharp intensity changes are represented by higher gradient values. Sobel or Prewitte operators can be used to calculate the gradient of an image. The magnitude and direction of the gradient at each pixel of an image is calculated in this step:

$$|\nabla I| = \sqrt{(\nabla_x I)^2 + (\nabla_y I)^2}$$

$$\angle \nabla I = \tan^{-1}\left(\frac{\nabla_y I}{\nabla_x I}\right) \qquad 3 - 24$$

3. Non-maxima suppression

The wide ridges are then trimmed around local maximum in gradient magnitude down to edges of single pixel width. In step 2 (gradient derivation), we have already computed the gradient magnitude and direction, then stored them in images g and θ, respectively. We obtain a thinning output from g and θ, where the thinning is g_s. The algorithm proceeds as follows:

For all (x,y):

Step 1. Approximate *θ(x,y)* by $\hat{\theta} \in \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$;

Step 2. if *g(x,y)<g* at neighbour in direction $\hat{\theta}$ or $\hat{\theta} + \pi$, then *g(x,y)=0*.

The gradient magnitude of each pixel (computed from step 2) is then compared with its neiboughhood pixels along the derived gradient direction (step 2). Only the pixels with higher gradient magnitude remain and those pixels with lower magnitude are subsequently supressed. In this way the search process is able to derive the ridge pixels representing potential edges.

4. Hysteresis thresholding

It is this step making the Canny detector achieve a better detection result as compared with those edge detection methods using fixed thresholds. For the ridge pixels (potential edges) derived from non-maximum suppression, a two-thresholding method is applied. Any ridge pixels with intensity higher than the higher threshold are immediately accepted as the permanent edge pixels and any ridge pixels with intensity lower than the lower threshold are immediately accepted as non-edge pixels. At this stage those ridge pixels with intensity between the higher and lower threshold are accepted as the potential edge pixels and therefore further judgement needs to be made to categorise these pixels. If any neighbourhood pixel(s) of a potential pixel are already defined as permanent edge pixels, this potential pixel is finally accepted as permanent edge pixel.

Otherwise it is defined as a non-edge pixel. This step gives the broken edge pixels a range, defined as the difference between the lower and higher thresholds, to be reconnected back to its neighbourhood edge pixels. This connected-component analysis is particularly important for edge detection of images with relatively poor quality, such as those delivered by low cost webcams, where edge discontinuity often occurs. With the two-threshold technique, the true edges which have been weakened are able to be re-defined from the potential edge group back to the definite edge group.

Recently Luo and Duraiswami (2008) have implemented the Canny algorithm within the CUDA framework [CUDA, 2007] making use of the parallel computing from the graphic processing unit (GPU). This implementation has been compared with the assembly optimised CPU implementation of Canny in OpenCV library. Luo and Duraiswami (2008) reported that provided with an image within 1024×992 size, CUDA demonstrated a speedup of processing with image size. However, such speedup becomes less apparent when the image size is over 1024×992. This is mainly because the edge number within image of this size usually causes a large number of hysteresis thresholding operations exceeding the parallel pipelines available on the architecture of the graphics processing unit. Serial processing is therefore resulted lowering the overall processing speed. Luo and Duraiswami (2008) concluded his study by stating that "*the CUDA function performed better for low edge count images in terms of both absolute runtimes and relative speedup with its CPU counterpart*".

Considering that the image size used in this research is larger than that of 1024×992 with a fair amount of edge counts per image, plus the fact that to allow slight target movements in the fast tracking application, completely fresh image data is required to compute the 3D target coordinates per camera frame. While Luo and Duraiswami (2008) reported a substantially high cost is needed for data loading from the GPU main memory, the loading of fresh data from multi-camera to GPU per frame would be impractical to implement and therefore CUDA was not implemented in this research. However, with the increasing number of multiprocessors in GPU nowadays, CUDA implementation for speedup can be an interesting future research direction.

3.4.4.3 Target recognition - Morphological operations

Normally an image after edge detection processing contains some disconnected edges. In the worst case, the object(s) of interest cannot be recognized since the shape(s) are greatly affected by this disconnection. This could produce a large variation across image frames in terms of the number of targets being detected for fast tracking. Considering the target detection consistency across frame sequences as a critical factor to achieve fast and meanwhile reliable photogrammetric target measurements, morphological operations are necessary to be introduced into the edge image. The functions of morphological operation include image smoothing, noise removal and connecting disconnected edges.

The most common methodology of morphological operations contains dilation and erosion, which is called mathematical morphology in digital image processing [Gonzalez and Woods, 1992]. Dilation followed by erosion is called closing, whereas erosion followed by dilation is called opening. Normally closing is applied in order to connect disconnected edges. Suppose an image A has been processed to a binary or edge image so that the intensity values are either 1 or 0. Here we define a matrix B with elements 1, the closing of image A by B is as follows:

$$A \oplus B = \left\{ c \in E^n \middle| c = a + b \quad for \quad some \quad a \in A \quad and \quad b \in B \right\}$$

$$A \otimes B = \left\{ x \in E^n \middle| x + b \in A \quad for \quad every \quad b \in B \right\} \qquad 3 - 25$$

where A and B contain $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$ respectively and they are both n dimensional (E^n). Figure 3 - 20 and Figure 3 - 21 show examples of dilation and erosion.

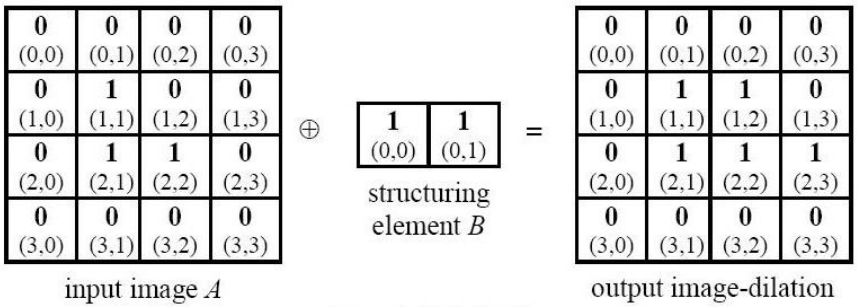


Figure 3 - 20. *Dilation.*

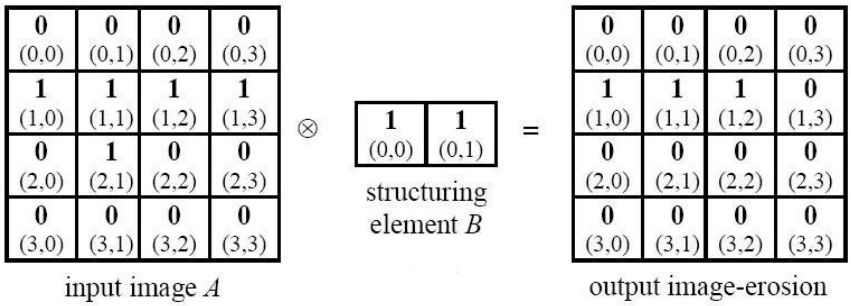


Figure 3 - 21. *Erosion.*

3.4.4.4 Target recognition - Edge tracing

After segmentation is applied, only the object's boundary pixels which form a certain degree of contrast with its surrounding image pixels are assigned 1 (intensity value) and all the other pixels (including those inside the target boundary) are set to 0. The next step is edge tracing, which aims at finding true targets from all edged objects by applying more target selection criteria and meanwhile storing information for target location calculation. Edge tracing consists of two steps: (i) contour tracing of target edges on binary images output from image segmentation, and (ii) extraction of structure parameters from RGB images, which include area, perimeter and circularity, in order to validate targets.

It is necessary to trace the contours of all potential objects which appear in the binary Image. Various methods have been proposed such as the chain code method [Pavlidis, 1982]. A scan-line filling method is chosen in this research since it is more efficient for the larger sections of images that must be recognised as not being targets [Shortis *et al.*, 1994], which is particularly the case in this research.

1. Trace the outline of any object (object edge) based on the binary image determined through the target segmentation (Section 3.4.4.2 and 3.4.4.3), as opposed to established photogrammetric targeting methods where the threshold is usually determined at this stage interactively from a visual segmentation [Shortis *et al.*, 1994]. Since the aim of tracing here is only to check if any connected components can form a complete shape (potential targets), compared with the conventional methods, not only no human intervention such as visual segmentation is needed, a simpler and quicker search is achieved by the binary image containing only edge pixels;
2. When the intensity value of a pixel is found to be 1 tracing is performed by an eight-directional search so that all the neiboughood pixels of this recognised image pixel (with 1 as the intensity values) can be included. The entire image is searched in this way from left to right and top to bottom. The search process is completed when all the candidate target image pixels are found;
3. A candidate target image is said to be found if the eight-way search is completed, i.e. the starting pixel has been visited twice. These image pixels are then temporarily removed from the search process with their image coordinates stored in computer memory. At this point the perimeter length and maximum x and maximum y dimensions are computed, where these data are stored in a structure and passed to the next part of the program.

When the tracing process finishes a list of co-ordinates representing target images are then used in the next step to calculate target image centroids.

3.4.4.5 Target recognition - Target structure parameters

Even after image segmentation and edge tracing it is likely that there are still some objects falsely recognised as true (legitimate) targets, so that it is necessary to distinguish between targets and non-targets. Typical features used contain perimeter length, size and circularity. The perimeter length of the subject can be calculated using the traced contour X,Y coordinates. The area can be calculated by counting all of the pixels inside and on the perimeter of the subject. A shape factor is used to express the differences between circular and non-circular subjects. The definition is given in Equation 3-26 [VMS, 1999].

$$C = \frac{A}{\pi \left(\frac{L}{2}\right)^2} \qquad 3 - 26$$

where A is the object area, L is the longest distance across the object. The equation gives the ratio of the subject area to the area of the circle circumscribing the subject. The nearer to a circle the object is, the closer to 1 the ratio is.

When all the three conditions are met with some predetermined bounds it is very likely that this object is a true target and therefore its coordinates are stored, otherwise the object is rejected. This process is repeated for the whole image until all target objects recognised from edge tracing have been tested.

3.4.4.6 Target location

Once target images have been identified, it is then necessary to refine the precision with which they are located in the image. However, target positioning accuracy can easily biased by a number of factors such as low target contrast, distance between the camera and the target, target occlusion resulted from narrow lines of sight [Clarke *et al.*, 1993].
A survey of subpixel techniques using grey scale images indicates that high location accuracies have been reported by a number of authors for objects such as edges, Gaussian blobs, patterns, etc. [West and Clarke, 1990]. Wong and Ho (1986) and Trinder (1989) demonstrated that the target centre used in photogrammetry is no exceptions. The subpixel accuracy with which a target can be located using these methods has been reported as high as 0.01 pixel, though accuracy can be heavily limited by noise [Deng, 1987]. In this research the target recognition was performed on binary images and the identified true targets and their associated edge coordinates (from Section 3.4.4.1 to 3.4.4.5) can then be used for the target location calculation performed on grey scale images.

Many methods have been used for subpixel location of targets. Examples include interpolation, correlation, centroiding, differential, and shape fitting. In this research the centroiding method is selected because the computation will give consistent results on small circular shaped targets even when viewed from different angles [Chen and Clarke,

1992]. Shortis *et al.* (1994) conducted simulations to study the effects of quantization, threshold variations, target size on target location, sensor saturation and DC offset on target location accuracy of six common target location methods (average of perimeter, binary centroid, grey scale centroid, squared grey scale centroid, ellipse fitting and Gaussian distribution fitting), where squared centroid performed most accurately in 4 out of 5 simulations. Shortis *et al.* (1994) further pointed out that centroiding performance can be negatively affected after thresholding is applied to remove noices, where this effect is smaller for the squared centroid method.

Combining all the above factors, the centroiding adopted in this research is squared centroiding on grey scale images:

$$\begin{cases} x_{centroid} = \dfrac{\sum_{i=1}^{n} i^2 \cdot x_i}{\sum_{i=1}^{n} i^2} \\ y_{centroid} = \dfrac{\sum_{i=1}^{n} i^2 \cdot y_i}{\sum_{i=1}^{n} i^2} \end{cases} \qquad 3 - 27$$

where

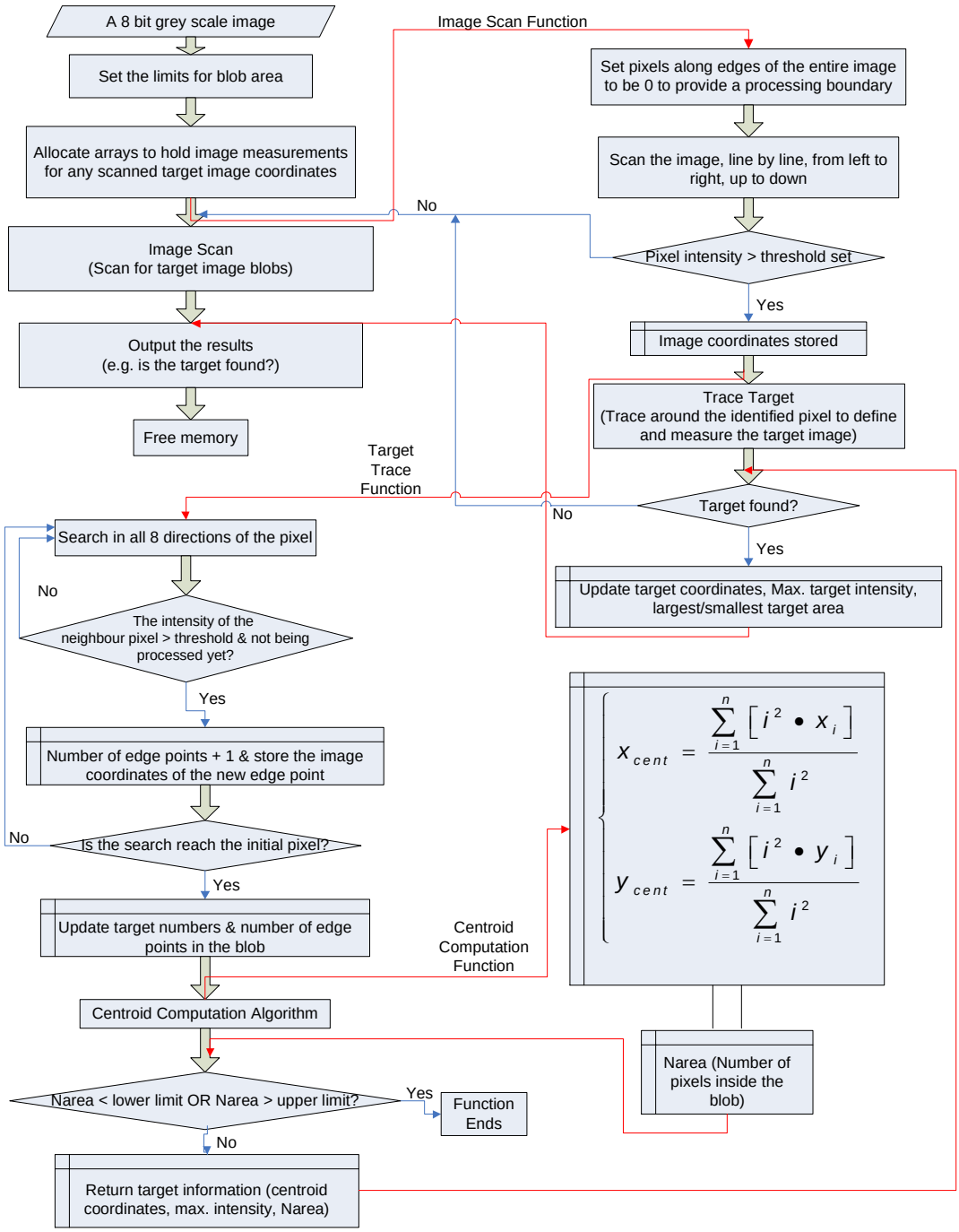
n is the total number of pixels contained in the target blob;

i is the grey scale intensity value of the pixel;

x_i, y_i are the x-coordinate and y-coordinate of the pixel.

3.4.4.7 Summary flowchart

The procedures described in Section 3.4.4.4 (edge tracing), Section 3.4.4.5 (extraction of target structure parameters) and Section 3.4.4.6 (target location), adapted in this research are summarised in Flowchart 3 - 2:



Flowchart 3 - 2. *Target image detection consisting of three parts:*

1. Image scan; 2. Target tracing; 3. Centroid computation.

Target image centroids of each image are then stored for subsequent 3D correspondence (Section 3.5) to compute target 3D coordinates.

3.5 Correspondence and epipolar geometry

Having measured target image locations, or more generally image feature locations, finding the correct correspondence is critical since 3D data completely relies on the correct match of the corresponding points between multiple images. Some research uses block matching to find correspondence without camera calibration [Li *et al.*, 1996]. However, the block matching requires a large amount of computation which is impractical to apply to fast 3D coordination. In order to increase the efficiency of finding correspondence, the relative geometry among cameras can be used to minimize the search space. In this section, fundamental knowledge of epipolar geometry will first be given, followed by the overview of correspondence matching techniques and finally the 3D space intersection matching with optimisation methods which has been applied in this research.

When both cameras see a point X in 3D space, the three points, X, C and C' forms a plane, which is called the epipolar plane (Figure 3 - 22). The line between C and C' are called the baseline. The intersected points between the baseline and the left and right image, e and e', are called the epipoles, where e is the projection of C' observed by the first camera and e' is the projection of C observed by the second camera. The two lines, l and l', which intersect the epipolar plane with the left and right images respectively, are called epipolar lines. A new epipolar line will form whenever X moves and all these epipolar lines will pass e on the left image and e' on the right image.

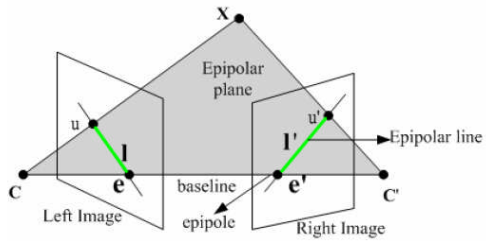


Figure 3 - 22. *Epipolar geometry [Sonka, 1998].*

Let u and u' be the projected points of X on the left image and the right image respectively. Then the ray CX contains all possible 3D locations which can be

projected onto the point u on the left image. The projected line of the ray CX on the right image becomes the corresponding epipolar line l' with respect to l. That means, the corresponding point u' on the right image of a point u from the left image must lie on epipolar line l'. This geometric relationship, which is called epipolar constraint, reduces the original 2D correspondence search to the 1D line search, which largely increases the search efficiency.

3.5.1 Overview of correspondence techniques

The automation of the correspondence is important in digital photogrammetry and computer vision. The matching of target images from one image plane with the corresponding target images in other image planes is an important step to make photogrammetric system automatic. A good review of target matching can be found in the thesis of Baltsavias, 1991. Since this research only concerns the correspondence of close range ordered images, correspondence dealing with wide baseline and unordered images will not be discussed and can be found in [Roth, 2004]. In UCL, a 3D target matching method (Section 3.5.2) has been previously developed based on a 3D intersection, as compared with the method using 2D epipolar line search. A space tree search process is adopted as opposed to the common random search matching process. The matching reliability is improved by utilising multiple camera views and appropriate multiple constraints based on groups of target images.

3.5.2 3D correspondence solution

The introduction of correspondence establishment by considering the intersection between camera rays in object space can be attributable to Chen *et al.* (1993, 1995) and Clarke *et al.* (1995). The underlying principle is based on the collinearity, where the camera ray formed from the image point through the perspective centre of the optical system to the object point is a straight line. In an ideal imaging system a corresponding point on the image plane from the right camera should produce a camera ray which intersects with the camera ray from the left camera (Figure 3 - 23). However, a combination of factors including image measurement uncertainty, principal point offset uncertainty, lens distortion uncertainty cause these camra rays not able to intersect with each other. However, the useful information is that the corresponding points should

produce rays which intersect with a minimum perpendicular distance (Figure 3 - 23), as compared with rays formed by other points. Chen *et al.* (1993) and Sabel *et al.* (1993) proposed algebraic and vector based methods respectively to compute the minimum distance. Since it is possible for a non-corresponding image point from the right image to form a camera ray intersecting the camera ray from the left camera with the same minimum distance, solution is necessary to minimise such ambiguities and this is usually solved by increasing the number of cameras in the imaging network. The correspondence approach based on 3D intersection has been proven to work reliably for more than twenty camera views [Chen *et al.*, 1993].

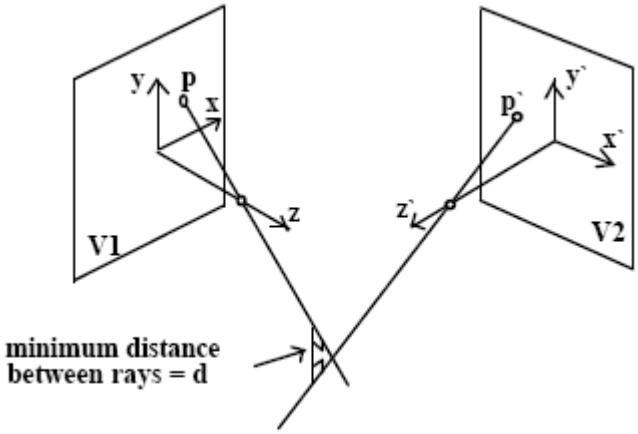


Figure 3 - 23. *3D intersection.*

The geometry of epipolar line has been given prior to Section 3.5.1. In practice, the 3D coordinates of a target could be either poorly estimated or unknown. The epipolar line of the target image in another image plane can be determined by the following method:

For each target image within the image plane V_1 four linear equations can be defined. The first two can be obtained from the collinearity equations whilst the second two can be obtained by projecting the target object (X,Y,Z) to the other image plane. Combining the equations and eliminating the common variables X, Y, and Z, the 2D epipolar line equation can be obtained. Figure 3 - 24 illustrates an epipolar line from another image plane which has been projected onto the current image plane. Two image planes will not always provide a unique correspondence solution and therefore more camera views are usually necessary to disambiguate targets. There are some further points to note regarding this technique:

(i) Accurate estimation of camera orientation parameters is necessary as their poor estimation will produce an error in image location and cause correspondence problems;

(ii) A straight line is assumed for the epipolar line search method. However, in general case the epipolar line can be distorted by systematic errors in the imaging system such as lens distortion and image plane unflatness;

(iii) A tolerance band will often be used. The size is related to the precision of the estimated camera orientation and the target image measurement method.

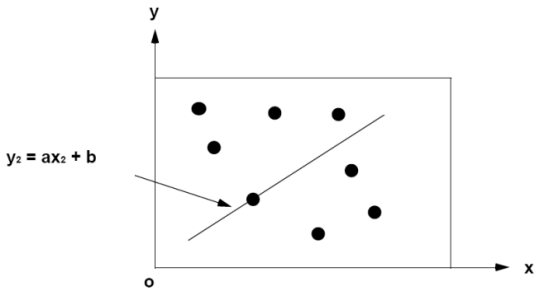


Figure 3 - 24. *An epipolar line from another viewpoint projected onto the current viewpoint.*

Collinearity equations and a functional model which describes distortion of the imaging system are included for target matching in 3D space. A tolerance value can also be closely related to errors in 3D coordinate estimation. If additional parameters are taken into account the collinearity equations can be expressed as [Methley, 1986]:

$$\frac{(X_a - X_0)}{m_{11}\left((x-x_0)+x_{lens}\right)+m_{21}\left((y-y_0)+y_{lens}\right)+m_{31}(-c)} = \frac{(Z_a - Z_0)}{m_{13}\left((x-x_0)+x_{lens}\right)+m_{23}\left((y-y_0)+y_{lens}\right)+m_{33}(-c)}$$

$$\frac{(Y_a - Y_0)}{m_{12}\left((x-x_0)+x_{lens}\right)+m_{22}\left((y-y_0)+y_{lens}\right)+m_{32}(-c)} = \frac{(Z_a - Z_0)}{m_{13}\left((x-x_0)+x_{lens}\right)+m_{23}\left((y-y_0)+y_{lens}\right)+m_{33}(-c)} \qquad 3 - 28$$

The above equations satisfy a 3D line equation in vector form:

$$\frac{X_a - X_0}{p} = \frac{Y_a - Y_0}{q} = \frac{Z_a - Z_0}{r} \qquad 3 - 29$$

where

$$p = m_{11}\left((x-x_0)+x_{lens}\right)+m_{21}\left((y-y_0)+x_{lens}\right)+m_{31}(-c)$$

$$q = m_{12}\left((x-x_0)+x_{lens}\right)+m_{22}\left((y-y_0)+x_{lens}\right)+m_{32}(-c)$$

$$r = m_{13}\left((x-x_0)+x_{lens}\right)+m_{23}\left((y-y_0)+x_{lens}\right)+m_{33}(-c)$$

Equation 3-29 describes a line in 3D space. If a target lies on the surface of a 3D object, the 3D line equation describes a ray projected from this target through the perspective centre of the lens. If all of the targets from another image are projected through its own perspective centre, provided with appropriate exterior orientation parameters, the ray from the corresponding target will pass closest to the single projected 3D line (Figure 3 - 23), with the minimum distance between projected rays (equation 3-30):

$$D = \frac{\begin{vmatrix} X_{02}-X_{01} & Y_{02}-Y_{01} & Z_{02}-Z_{01} \\ p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 \end{vmatrix}}{\sqrt{\begin{vmatrix} p_1 & q_1 \\ p_2 & q_2 \end{vmatrix}^2 + \begin{vmatrix} q_1 & r_1 \\ q_2 & r_2 \end{vmatrix}^2 + \begin{vmatrix} r_1 & p_1 \\ r_2 & p_2 \end{vmatrix}^2}} \qquad 3 - 30$$

where p_1, q_1, r_1 and p_2, q_2, r_2 are vectors of equation 3-29 for two viewpoints.

Similarly, target image candidates corresponding to other targets can be matched in this way. The 3D space matching method is illustrated in Figure 3 - 25 for the two projected rays of a matching target using two camera viewpoints as an example.

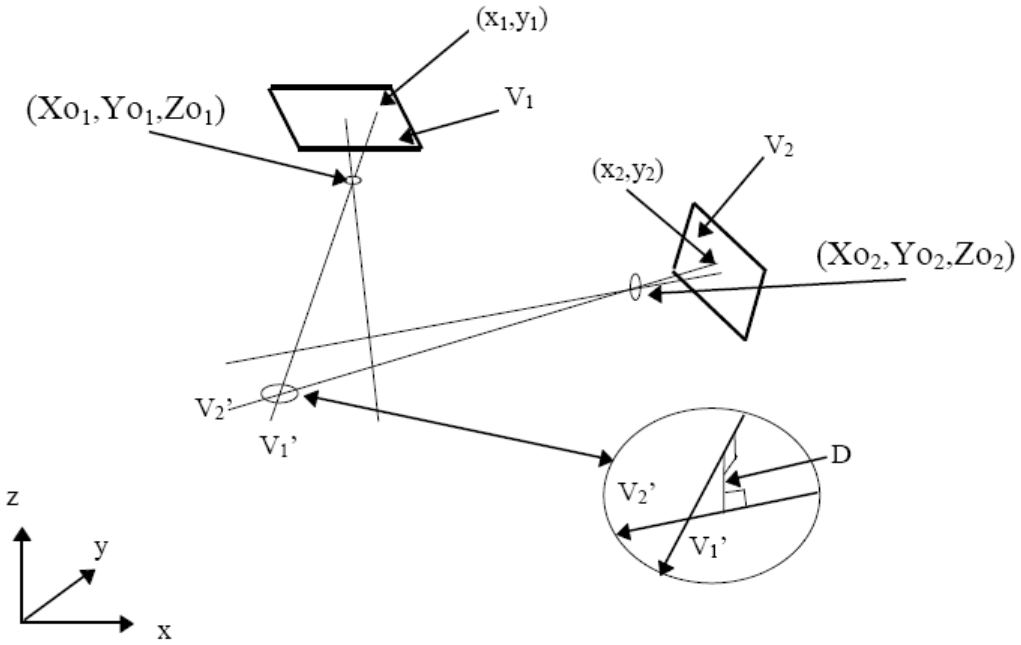


Figure 3 - 25. *Minimum distance searching between two projected rays.*

There are two basic differences between the 3D space method and the conventional 2D epipolar line method:

1. The epipolar line method is a 2D projection of a space ray, which passes close to a point in another view, while the 3D space method is the space intersection of rays through image points from two or more views;
2. The threshold values for correspondence matching are different. The epipolar method uses the target image coordinate residuals whilst the 3D space method uses the computed RMS standard deviations of the target 3D coordinates.

3.5.3 Correspondence optimisation

Although the matching reliability is improved with the 3D space method, the number of target image matching can be significantly increased with the number of projected rays and number of target images existed within camera views. Therefore it is necessary to optimise the 3D space method in terms of its efficiency in multi-camera as well as multi-target photogrammetric measurements.

Hierarchical approaches has been a popular approach to accelerate the matching process [Zhang, 1994; Turk and Levoy, 1994; Neugebauer, 1997]. The underlying principle of these methods is to reduce the number of points in the matching process. Methods using sub-sample of image data have also been proposed by many authors, such as considering only points in smooth surface areas [Chen and Medioni, 1992], only points with high intensity gradients [Weik, 1997] and selecting points based on the surface normal distribution [Rusinkiewicz and Levoy, 2001]. However, these methods are highly dependent on the content of point clouds selected and therefore sensitive to factors such as target contrast and image noise.

An effective solution to accelerate the matching process is by considering the search structure, where the complete search space can be refined to a small group to establish the target matching group. This optimisation method was first developed by Bentley (1975) as a k-dimentional binary search tree (K-D tree). The complete search space is represented by the tree root while its mutually exclusive subsets are represented by the

leaves extended from the root. The space partitioning method for 3D point matching (Figure 3 - 26) first introduced by Chetverikov (1991) was since then extended by Gruen and Akca (2005) to image space. The underlying principle of this space partitioning method is to first perform hierarchically adaptive nearest neighbourhood search, followed by a matching process where only those relevant subsets representing the potentialmatching group were matched with each other simultaneously.

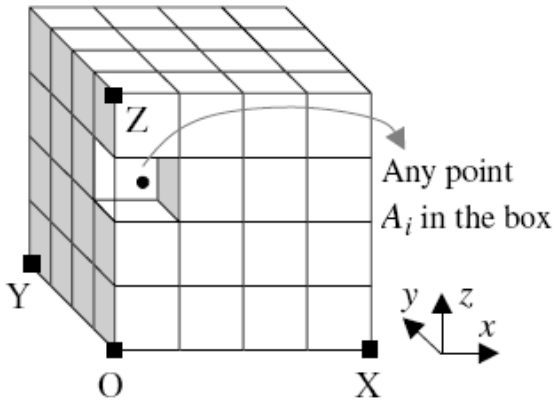


Figure 3 - 26. *3D boxing.*

To find a potential corresponding point in image space, back projection can be used to project the 3D object point initially estimated from at least two camera views back to image planes of other camera views (Figure 3 - 27). The corresponding point from each camera view can then be found by searching for the closest target image point around the projected point within a (usually subpixel) tolerance value. This approach has been successfully used to solve the correspondence problems in many close range photogrammetric measurements [Fraser, 1997]. However, the assumption of this approach is the existence of all camera views including an object point, i.e. a matching group may not be reliably established if target occlusion occurs for one or more camera views. Further optimisation methods are therefore needed to include target occlusion in the matching process.

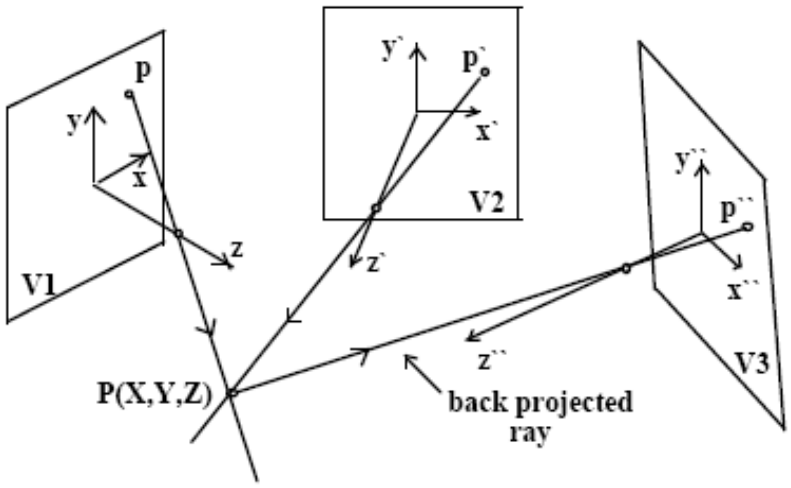


Figure 3 - 27. *Back projection.*

In order to reliably establish the correspondence for all image points in all camera views, each point in one camera view is necessary to be compared with all points from all the other camera views. Regarding this technique, Chen *et al.* (1995) proposed an optimisation method utilising a tree search structure so that each comparison and judgement can be made for a group of points rather than individual points.

3.5.3.1 Optimisation of the 3D intersection matching method

Section 3.5.2 describes the 3D intersection matching method developed in UCL using object space constraints, which is equivalent to the epipolar line method performed in image space. Additional parameters such as lens distortion are included for the minimum distance computation and a tolerance value directly related to the measurement process can be selected. The advantage of this method is its capability of solving correspondences between large numbers of viewpoints which are not required to overlap [Chen *et al.*, 1993]. However, as previously mentioned, this method must be optimised so that the matching process can be completed within a reasonable time in favour of photogrammetric automation.

3.5.3.1.1 Matching tree search process

In order to match the targets imaged in multiple camera views all possibilities are required to be considered and tested. Consider an example where there are a number of possible combinations (Figure 3 - 28):

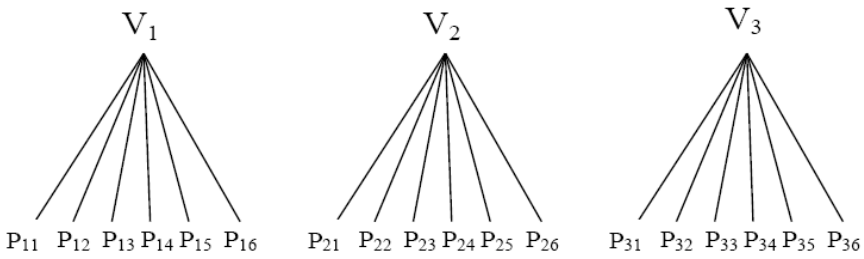


Figure 3 - 28. *An example of combinations.*

A simple search method is described as follows:

1. choose a target image from V_1;
2. choose any target image from V_2;
3. choose any target image from V_3;
4. using the three target images selected, calculate the distance between any two target images out of the three combinations. The target image group where all the measured distances between rays satisfy the distance tolerance is temporarily accepted as a corresponding target image group. The group is later accepted permanently as a corresponding target image group if none of its target images are candidates in any other group;
5. choose another target image from V_1 and repeat the previous three steps (2-4).

For 3 viewpoints and 6 targets (Figure 3 - 28), there are 216 (6 x 6 x 6) combinations of target image groups. For each group two target images need to be chosen (out of three) for distance computation, so there are 3 such computations (e.g. ab, bc, ac). Therefore, a total of 648 (3 x 6 x 6 x 6) distance computations are needed. Generally, assuming the number of targets is the same in each viewpoint and there are m viewpoints and n target images in each viewpoint, then the total number of distance computations taking account of all possibilities is:

$$N = C_m^2 * \left(C_n^1\right)^m = \left(\frac{1}{2}\right) * m * \left(m-1\right) * n^m \qquad 3 - 31$$

However, in practice there are usually more than 50 targets and more than 5 views making this method rapidly become computationally impractical. Regarding this this search method is required to be optimised by narrowing the scope of search and one way to achieve this is to only search for the group being matched. For example, the

previous method (Figure 3 - 28) can be rearranged into a matching tree (Figure 3 - 29) where each level represents a viewpoint and the nodes are the target images within that viewpoint. The search will continue only if a node corresponds with the node from its upper level, otherwise the node and all nodes below it will be removed from the search tree. This method, called a “matching tree search process”, was first applied in pattern recognition by Grimson (1990), where a matching judgement is made at each node of the tree.

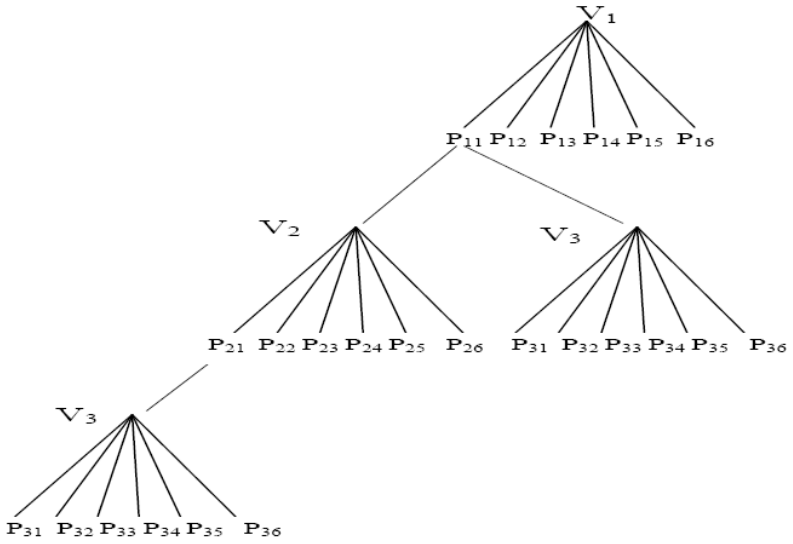


Figure 3 - 29. *A tree of searching structure with levels and nodes.*

Finding correct target matches using this approach can be treated as a search within multiple dimensional spaces. Each viewpoint (level) represents a coordinate axis and the target images in each viewpoint (nodes) are expressed by discrete coordinate labels on each axis. The place of any two axes represents a possible match between two target images and testing all nodes ensures that all possible matches are found. However, the problem with this approach is that the matching space can be too large in practice. For example, if there are m viewpoints and each viewpoint has n target images, then there are n^m points in discrete 3D space and $C_m^2 \times n^2$ points in discrete 2D planes. Therefore alternative methods are necessary to minimise the number of searches without missing any valid targets. In the following section, one of the artificial intelligence and model-based recognition methods which have been developed and optimised is discussed in favour of target matching.

3.5.3.1.2 Global consistency width-first search

To illustrate this search method, 3 viewpoints and 6 target images in each viewpoint are selected as an example (Figure 3 - 30). A target image from V_1 is held fixed. A judgement is made between this target image and one of target images from V_2 (as root). If the matching candidate is not satisfied another target image from V_2 is selected. No judgement involving V_3 will be made until a match has been found between V_1 and V_2. For example, the matching pair $\{P_{1i}, P_{3k}\}$ between V_1 and V_3 is searched only after a matching pair $\{P_{1i}, P_{2j}\}$ has been found. This search process, called the width-first tree backtracking, will terminate at root level if there is no matched target image from V_2 with the fixed target image from V_1. That is, the search will progress to the leaf node only after the root has been satisfied. This method is considered to be efficient since the computation time can be substantially saved by eliminating unnecessary searches. On reaching leaf pairs $\{P_{1i}, P_{2j}\}$ and $\{P_{1i}, P_{3k}\}$ may have been found to be matching pairs. However an additional judgement is needed for $\{P_{2j}, P_{3k}\}$ to ensure the global consistency. The advantage of the width-first search method is that the software implementation requires only a single loop process to realise the search. This is critical in terms of applications demanding fast photogrammetric 3D measurements as an example in this research. Optimisation is achieved by this method in which a possible matching group can be formed when the search reaches any target image in V_3 and it is consistent with the node of V_2.

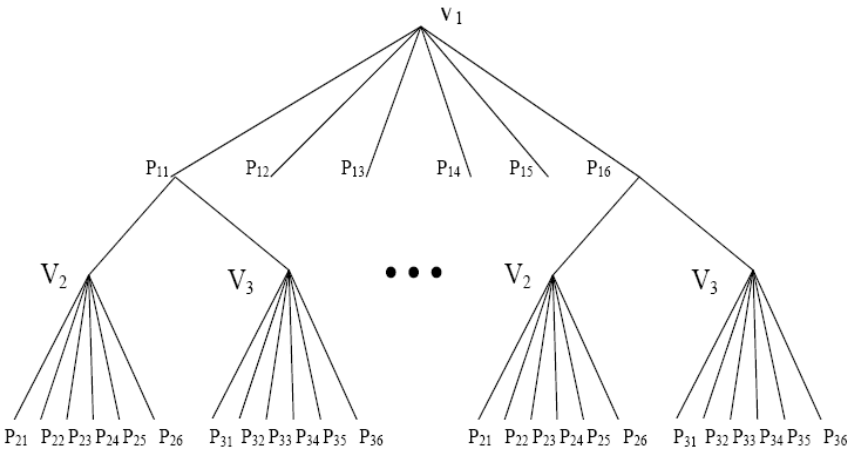


Figure 3 - 30. *Illustration of width-first tree search.*

3.5.3.1.3 Introduction of pseudo-target images to account for occlusion

The 3D space matching method assumes that all targets are imaged in every viewpoint. However, if some targets are occluded, a condition with no leaf of the tree producing a consistent matching group will then occur. Therefore the constrained search method might terminate before reaching a leaf and will finish prematurely. A solution to this problem is to extend the search method to allow for occlusions by introducing an imaginary or pseudo target image.

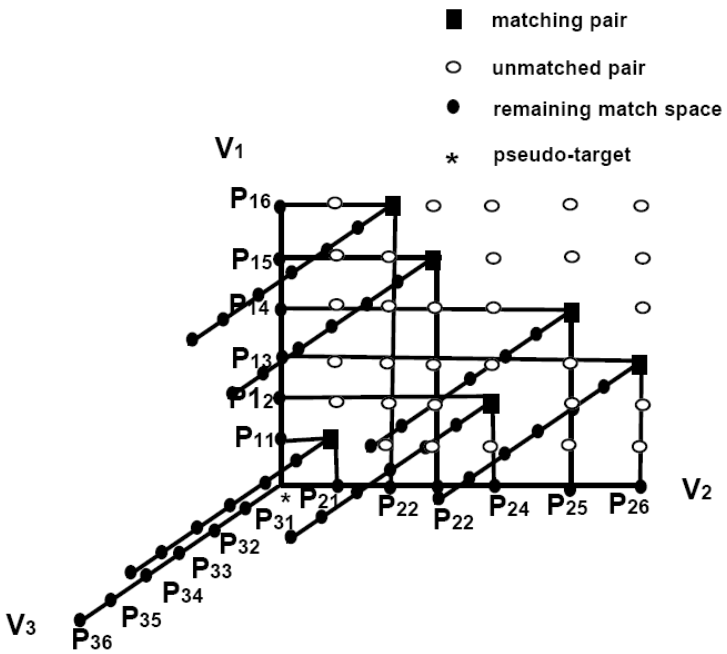


Figure 3 - 31. *Addition of pseudo-target image to account for occlusion.*

At each node of the matching tree, an extra branch corresponding to the pseudo-target image, denoted by a "∗", is added (Figure 3 - 31). An extra target is needed in the correspondence space on all viewpoint axes but this addition is made without the need for any extra matching judgements since the pseudo-target is excluded from the final object space match computation. To add this to the matching scheme, relationships between the group of matching target images and the pseudo-target image must be defined. Since the pseudo-target image is excluded from the match computation, it will not affect the match algorithm itself and therefore any search constraint linking the matched target image to the pseudo-target image will be consistent. With the addition of pseudo-targets it is possible for the constrained search method to overcome problems caused by target occlusions, which might frequently happen in multi-camera photogrammetric measurements especially for those cameras with relatively poor image quality such as webcams used in this research.

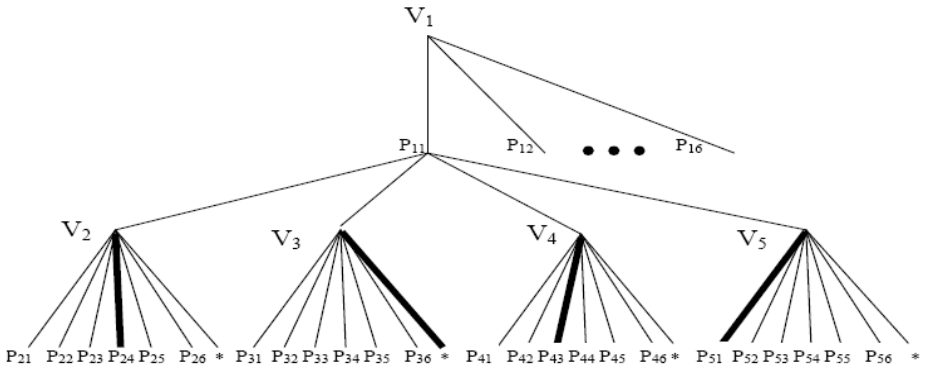


Figure 3 - 32. *Addition of pseudo-target in width-first tree.*

As an example, given five viewpoints and a target which appears in every viewpoint except viewpoint 3, the correct match group should be $M = \{P_{11}, P_{24}, P_{43}, P_{51}\}$. With the introduction of pseudo-target image, one valid matching group through the width-first search tree is shown in Figure 3 - 32, which passes all the viewpoints by including both real target images and a pseudo-target image. A hypothetical match group is then obtained: $P = \{P_{11}, P_{24}, *, P_{43}, P_{51}\}$. To allow fewer target images than the total number of viewpoints to be used as a correct matching group, a viewpoint threshold is defined by m_t where $1 < m_t < m$. By this definition, a matching group of target images accounting for occlusion can be obtained provided that the target is imaged in m_t or more viewpoints.

3.5.3.1.4. Hypothesis-testing and heuristic methods

When searching through the matching tree, a possible matching group is produced each time a leaf is reached. However, not all matching groups are correct. For example, it cannot be guaranteed that the group $P = \{P_{1i}, P_{2j}, P_{3k}\}$ is a matching group and any extra groups, for example, $\{P_{1i}, P_{3k}\}$ or $\{P_{2j}, P_{3k}\}$ must be checked. Therefore if m viewpoints are used (m-1)(m-2)/2 extra checks are necessary. In this way the extra computation required increases with the number of viewpoints according to a square relation. Regarding this a more efficient method should be used to verify the matching group.

A hypothesis-testing method can be used to optimise the search. When the search reaches a leaf, the group of target images found is considered as a hypothetical matching group. Using this group of target images, a least squares technique can be used to estimate the 3D coordinates of the corresponding target, followed by a cross checking among all the target images in the group to ensure they are the actual target images (image plane projections) of the computed target. This check is carried out by using the spatial target position to verify the correctness of the match. A description of the hypothesis-testing method is as follows:

- Hypothesise the existence of a matching group

Using the width-first search method (Section 3.5.3.1.2), a group of target images, called a hypothetical matching group, consisting of target images from the top level (root) to bottom level (leaf) can be obtained.

- Refine the spatial position of the hypothetical matching group

Using the hypothetical matching group, i.e. $P = \{P_{1i1}, P_{2i2}, ..., P_{\min}\}$ where $(P_{1i1} \subseteq V_1, P_{2i2} \subseteq V2, ..., P_{\min} \subseteq V_m)$, the 3D coordinates (X_a, Y_a, Z_a) of the corresponding target can be estimated by 3D space intersection method using least squares estimation (Section 3.5.2).

- Test the hypothetical matching group

Given the estimated 3D target coordinates, the testing process can be specified. Let the set of pairings $P=\{P_1,P_2,P_3,...,P_m\}$ $(P_1 \subseteq V_1, P_2 \subseteq V_2, P_3 \subseteq V_3, ..., P_m \subseteq V_m)$ denote the hypothetical matching group. For each target image in each viewpoint P_i, the ray P_iO_i passing through the camera perspective centre O_i and the ray TO_i between the estimated space target and camera perspective centre O_i are computed. The minimum spatial distance between the two rays P_iO_i and TO_i is then computed followed by testing to check if it is within pre-selected error tolerance. If all target images in the group satisfy the test, the matching set is taken as being globally correct and therefore accepted as a final correct matching group. If any of the target images do not satisfy the test, the matching group will be discarded to ensure global consistency in the search. Such an approach is particularly appropriate when the matching group consists of many viewpoints (a multi-camera photogrammetric system).

- Iterative 3D target matching

The epipolar method to match the corresponding target images generally requires the camera orientation parameters to be accurately defined. However, these parameters are difficult to obtain with suitable precisions prior to solving the correspondence. This problem can be overcome by iteratively combining the 3D matching method with a bundle adjustment. In this way the bundle adjustment can be used not only to iteratively calculate the 3D coordinates of matched targets but also to iteratively refine camera orientation parameters. First the initial camera exterior orientation parameters are only estimated. These initially estimated parameters together with an initial 3D tolerance are then used to match a few target images for the first iteration, followed by a bundle adjustment which is able to refine these parameters and update the tolerance value. These updated parameters from the bundle adjustment are then used for the next matching process. The network becomes gradually strengthened by repeating the above process until all target correspondences are found. The 3D tolerance band of the space distance between two rays can be obtained and updated from the RMS target coordinate standard deviation derived from the bundle adjustment. Therefore the tolerance is directly related to the strength of the network solution.

However, the matching result can be significantly affected if the target matching process is unreliable, i.e. non-corresponding points are accepted as the final matching group. Regarding this target ambiguities and occlusion should be minimised. The bundle adjustment gives a lot of flexibity of introducing additional targets into the network. Only targets appearing on all camera views are initially introduced to the network to ascertain the stability of the network adjustment. As the network strength increases, more targets (targets only appearing on a subset of views) can be introduced. The occlusion problem can be effectively overcome by increasing the number of camera views in the network. As the network strength increases, the refined tolerance becomes more capable of rejecting those targets where ambiguities arise. This method has been found to be valid in the general case, and appears particularly suited to accurate industrial measurement where highly complex objects requiring multiple views must be measured reliably [Chen *et al.*, 1993].

3.5.4 Summary

In this section a method of 3D target matching using an optimised constrained search, originally developed at City University, but subsequently refined and re-coded at UCL is discussed. A tree search method is used to minimise the computation time. Provided that camera orientation parameters are not accurately known, they can be iteratively updated using an integrated bundle adjustment technique. The use of multiple viewpoint constraints, together with the gradual introduction of targets, is used to improve the robustness of the matching process. Pseudo-target images are defined in the search process to overcome occlusion problems and make the matching strategy flexible. Hypothesis-testing and heuristic processes are used to make the matching computation more efficient. The benefits of the method when compared to the conventional epipolar line method, provided that camera calibration parameters and viewpoint orientations are known are:

1. The method is extended to deal with multiple viewpoints in a flexible manner;
2. Any ambiguities caused by occlusions can be resolved;
3. The method is optimised to work efficiently with a large number of targets and many viewpoints;
4. The estimation of any additional parameters can easily be included because of the close connection with the bundle adjustment.

Additional benefits of the iterative use of the optimised 3D matching method are:

1. High accuracy is not required for the initial estimation of camera parameters. The initially estimated values can be iteratively refined from the bundle adjustments;
2. Tolerance values for target matching are continuously updated from the bundle adjustment;
3. Targets can be added to or removed from the network to make the matching process more robust and controllable;
4. Target images can be arranged in an arbitrary order to build up the matching tree.

While this bundle adjustment based method is particularly suited to accurate measurement (1mm coordination accuracy for clinical use, Section 2.2.3) where highly complex objects, such as the spherical shape of human head partially covered with

optical topographic sensing pad are to be measured. This application requires multiple (webcam) views which must be reliably measured since consistent coordination is needed for clinical studies where each study lasts for at least five minutes.

A calibration object with coded targets particularly suited to webcam imagery has not been built during the time of this research making a fully automated iterative bundle adjustment refinement procedure difficult to implement, provided with the relatively poor webcam image quality.

A webcam calibration based on bundle adjustment is performed (using VMS (Geometric Software)) independently prior to optical topography coordination applications so that target ambiguities and occlusions can be avoided through manual inspections and corrections and a strong (webcam) network configuration can be ascertained, in order to accurately define the camera parameters for each webcam in the system.

Once up and running with such a parameter model, the epipolar search method and 3D space intersection (Section 3.5.2) with its optimisation (Section 3.5.3.1.1-3.5.3.1.4) can be used to match the corresponding target images in optical topographic studies. To achieve this, the background assumption is that the camera exterior orientations are fixed during these studies and that the camera interior orientations are stable enough to hold the defined parameter set. The elimination of the bundle adjustment process not only increases the efficiency of the coordination performance, i.e. increase the frame refresh rate, but also allows a fully automated multi-webcam photogrammetric coordination system to operate without any control or reference targets in favour of clinical studies.

3.6 Multi-camera tracking system

Since this research aims at developing a tracking (coordination) system built with multiple low cost cameras in order to support optical topography, reported research involving motion tracking and target coordination from multi-camera systems (in both photogrammetry and computer vision communities) would be precious in terms of the considerations and precautions for system development. This section gives an overview of multi-camera tracking techniques, with special focus on the state of art of human shape and motion tracking. The scope of discussion then focuses on the multi-camera image based 3D measurement (mainly from the photogrammetry community), where measurement automation is addressed. Considerations of the development of a fast low cost multi-camera tracking system in this research are finally discussed, where the platform architecture and working principles underneath such a system are particularly analysed (See Appendix 1, 2 and 3).

3.6.1 Overview

Multi-camera tracking techniques have rapidly evolved during last two decades. The experience that the researchers obtained and reported has provided useful insights of building a system with multiple cameras. Generally the more cameras involved, the better the tracking results in terms of shape recovery and tracking accuracy. However, the cost of building such a system could significantly increase with the number of cameras used.

One research system which employs many synchronous cameras is the Virtualized Reality system developed at Carnegie Mellon University [Narayanan *et al.,* 1998; Kanade *et al.,* 1997; Rander *et al.,* 1997; Kanade *et al.,* 1998; Saito *et al.,* 1999]. In that system more than 30 cameras are mounted in a room and are synchronized to record what is happening in the work volume. Target information such as shape, motion, etc. is then estimated through post-analysis of the recorded data from cameras.

One issue specifically reported is the significant amount of effort and cost required to synchronize cameras [Kanade *et al.,* 1998]. A tracking system that does address

asynchronous input is the head tracker developed at the University of North Carolina [Welch *et al.,* 1996, 1999, 2001; Azuma *et al.,* 1995; Azuma *et al.,* 1994; Gottschalk *et al.,* 1993]. In this system the sensors (photodiodes) are mounted on the target and observe outward infrared light-emitting diodes (beacons) mounted on the ceiling. During the tracking only one beacon is seen at any time, and therefore only partial information about the target position is received. There are some commercial systems developed for motion capture [Motion Analysis Cooperation; Vicon Motion Systems], where Vicon provides an impressive level of automation in its use in UCL Department of Computer Science and NASA. These systems require active or passive markers being mounted on the target to provide fast and robust 2D feature extraction, with a small increase of the system's intrusiveness. Many (10s or 20s of) cameras are usually employed to obtain a large working volume and to reduce occlusion. However, cameras are still required to be synchronised, and not much effort has been reported in addressing asynchronous and usage of a mixture of different cameras. These systems also provide a limited tracking support of multiple points because point features are simple and it is difficult to distinguish one from another.

3.6.2 Tracking of human

3.6.2.1 Static 3D shape modeling

The 3D modeling of human subjects using image data has been deeply studied in the last decade. Nowadays requirements for 3D human models have increased for applications such as medicine and virtual environments, most notably in the television and film industries. For such applications, a human model is usually composed of the 3D body shape as well as the body movement. Laser scanning has been recognised as a common technique to capture the static 3D body shape and texture. For example, Cyberware's [Cyberware] new Head & Face Colour 3D Scanner (Model PX, Figure 3 - 33) provides a maximum scanning coverage of human head. Hundreads of thousands of measurements can be taken by Cyberware's Whole Body Colour 3D Scanner (Model WBX containing four scanners, Figure 3 - 33) under 17 seconds to generate an accurate 3D data set. According to Cyberware, the Head & Face Colour 3D Scanner can be useful in medical applications, such as plastic surgery. Similarly, it is useful for

designing helmets closely fitited with head. However, the Head & Face Colour 3D Scanner costs $67,000 while the Whole Body Colour 3D Scanner costs $200,000. Such systems also provide a very limited portability in terms of flexible diagnosis in clinics.

Figure 3 - 33. *Left: Cyberware's new Head & Face Colour 3D Scanner; Right: Cyberware's new Whole Body Colour 3D Scanner.*

3.6.2.2 Motion tracking

The main idea for applications such as video-surveillance, identification, authentication and monitoring of human activities is to detect and track moving targets. Motion capture techniques with sensor-based hardware, e.g. AscensionTM, Motion AnalysisTM and ViconTM are used to acquire the movement information. Regions of moving targets should be separated from the static environment and the methods should cope with occlusions, changes in illuminations and different types of motions. To identify and separate the moving targets, different approaches have been proposed such as background subtraction [McKenna *et al.*, 2000; Rosales and Sclaroff, 2000], combination of motion, skin colour and face detection [Gavrila and Davis, 1996] and spatio-temporal learning [Dimitrijevic *et al.,* 2006].

Most approaches require data to be acquired from a stationary camera because the camera movement cannot be distinguished from the moving targets. For a stationary camera, the most simple but quite efficient approach is the subtraction between two consecutive frames, where the generated image has much larger values for the moving components. More robust approaches are based on background learning and

subtraction [Haritaoglu *et al.*, 1998; Stauffer and Grimson, 1999; McKenna *et al.*, 2000; Kim *et al.*, 2004]. For example, moving people in video sequences are identified by distinguishing between moving foreground and static background from RGB information in the images [Kim *et al., 2004*]. As the fixed camera is imaging the same static scene, each frame is analysed to create a background model which is then used in the subtraction process. The detected moving areas (in the form of white pixels) are used to extract the foreground.

A lot of work has also been devoted to the segmentation of useful features from the background solely based on the natural appearance of the target and how to represent the segmented features. For example, the extracted features can range from edges [Hogg, 1984] to blobs that are regions of coherent colour [Darrell *et al.,* 1994; Iwai *et al.,* 1999; Hilton, 1999]. One example is the P finder system developed by Wren *et al.* (1997 and 2000), in which the human body is tracked using statistical models to segment the image into blobs. The advantages of using these natural and complex features include the reduction of the system intrusiveness and a higher ability to distinguish multiple features because they are described by more parameters.

Considering this thesis is aimed at target coordination located on optical sensing pads and the scalp, it would be impractical to implement methods such as frame subtraction to only detect the moving targets while leaving most of the moving components (e.g. face, upper body) useless for data acquisition. Plus factors such as electronic compression and CMOS sensor used in low cost cameras cause a fair amount of random image noises (differ from frame to frame). The noise level in the resulting image (from subtraction) can be higher than in either of the two original images because of the random distribution of the noise within each image [Sprawls, 2000]. Approaches using background learning always require a time consuming training phase in which the video sequence is analysed and a reference background image is established. Since the portability, flexibility and robustness are considered as the most critical factors regarding the system development (to comply with the properties of optical imaging techniques, Section 2.1) in this research, techniques based on background learning become impractical to implement for clinical studies, which are always carried out at different places, e.g. patient's bedsides or clinical testing rooms. Moreover, since

segmentation techniques involves features which are more complex than target points, the segmentation and extraction steps take significant processing and usually do not provide the frame rate for a fast tracking application. Meanwhile because these features are usually coarse and imprecise, they are less suitable for applications demanding a higher accuracy on positioned data.

Combining the above factors, a process including noise removal, target image detection and subsequent 3D target coordination performed on per frame basis appears to be a more practical and robust solution for the webcam-based coordination suited to optical imaging studies, provided with an acceptable time delay between frames. While expensive hardware is required for capturing human 3D static shape to a high accuracy, the current motion capture techniques mainly focus on the motion tracking while the precision and accuracy are not the major concern. There is also a strong need to fill the gap between these two apparently independent techniques, i.e. to obtain 3D locations with an acceptable accuracy level while allowing a certain degree of body movement during the capturing process, preferred with a low cost designed system.

3.6.3 Examples of image based 3D measurements

3D target coordination is a process from the data acquisition to the output of 3D target coordinates. The requirements specified by many applications such as medical photogrammetry are the high geometric accuracy, the automation, the low cost and the flexibility of the technique. Among various 3D object measurement techniques, image based measurement using correspondence matching to derive 3D object information has been widely applied in various applications such as geometric surface measurement of architectural objects [Van den Heuvel, 1999; El-Hakim, 2002].

Image based techniques usually acquire 3D measurements from multi-stations often with low cost sensors, using the perspective camera model, making the image based measurement very portable. However, in most of the cases human interactions are still required for accurate results. Accurate 3D coordination from image measurements can be very difficult if uncalibrated cameras are used since the wrong parameter recovery could lead to result inaccuracy. Various commercial packages [CanomaTM,

ImageModelerTM, iWitnessTM, PhotoGenesisTM, PhotoModelerTM, ShapeCaptureTM], based on manual or semi-automated measurements, are available to calibrate cameras and coordinate 3D object points from a multi-image network after the bundle adjustment process. The research based on multiple image based measurement can be divided into three types:

- Automatic 3D measurement from uncalibrated images

Many researches have been conducted in order to fully automate the entire 3D measurement process, from camera calibration to computation of 3D coordinates of points of interest, though without much success. Nister (2004) proposed a fully automated measurement process, where an uncalibrated camera was first used to take a sequence of images under very small baselines, followed by automatically extracting interest points and sequentially matching them across views. Camera parameters are recovered in the next step and finally 3D coordinates of the matched points are computed. This automatic procedure relies on a large portion of verlapping between images. The sequence is usually initialised from the first two images and subsequently iterated through the bundle adjustment. A self-calibrating bundle adjustment is often used to recover the camera interior orientation parameters. The 3D surface model is then automatically generated by dense depth maps.

For complex objects, a detailed 3D model is normally obtained with more matching procedures [Scharstein and Szeliski, 2002]. Features first identified from the scene and then automatically matched are the key towards the automation of image-based measurements. Therefore factors causing variation of image measurements such as object occlusion and light variation could significantly affect the result. The network configurations leading to feature ambiguity have been identified by Hartley (2000) and Kahl *et al.* (2001).

- Semi-automatic 3D measurement from calibrated images

Semi-automated approaches are much more common for complex geometric objects. Liebowitz *et al.* (1999) presented a method of obtaining 3D object information from a small number of images. Following manual measurements of points of interest, the

method constraints the network using the geometric configuration and the camera parameters. Images can be calibrated without any assumption of object shapes [El-Hakim, 2002] or using known object shapes [Lee and Nevatia, 2003]. For calibration without information from object shapes, El-Hakim (2002) measured some seed points manually and then fitted a quadratic or cylindrical surface in multiple images, followed by using the photogrammetric bundle adjustment to achieve a higher geometric accuracy in terms of parameter recovery. For calibration based on object shapes, Lee and Nevatia (2003) adapted an interactive approach where it was the user who provided the shape information such as height and width.

- Automatic 3D measurement from calibrated images

For automatic 3D measurement techniques, camera calibration is usually performed as an independent process prior to the 3D measurement. The automation is highly dependent on the automation of feature detection, the automated establishment of the correspondence across different camera views as well as a strong imaging geometry. One example is the geometrically constrained matching approach proposed by Gruen *et al.* (2001) to obtain 3D information of dense points of a complex object, following a semi-automated calibration process with a Minolta Dynax 500si analogue SLR camera.

To conclude, there is no single 3D solution which is able to work across different scenes and meanwhile is fully automated satisfying requirements of every application [El-Hakim, 2001]. This is because an automated method usually requires good features in multiple images with a well convergent network configuration [Fitzgibbon and Zisserman, 1998; Pollefeys *et al.*, 1999; Nister, 2004], which is difficult to satisfy in practice due to limitations such as occlusions, illumination changes and background noises. Automated coordination often ends up with data containing too many targets that are wrongly recognised due to such limitations. User interaction is still needed in most applications, where good results have been reported [Debevec *et al.*, 1996; Gruen, 2000; El-Hakim, 2002; Gerth *et al.*, 2005].

In terms of the webcam-based application developed in this research, accurate recovery of camera orientation parameters accounting for the highly possible

manufacturing variations across webcams is critical in order to have an accurate coordination performance from webcam imagery. Moreover, considering the optical imaging studies where human subjects move unconsciously, rapid 3D coordination without human interaction would be highly desired. Therefore camera calibration for the multi-webcam system in this research should be performed separately prior to the target tracking process (Section 3.5.4).

3.6.4 Automated fast 3D measurements with webcams

El Hakim (1986) defined real-time as: *"a system without interruptions, or appreciable time lags, between acquiring the image and the final results that are the three dimensional co-ordinates",* where *automated* could also be used instead of *real-time*. A fast 3D measuring process implies that its major components, feature extraction, feature matching and 3D computation needs to be completely fast enough without loosing efficiency. Such measurements require at least two camera views to intersect features in object space. Ariyawansa and Clarke (1997) intensively investigated the matching problem for stereo imagery and reported that a stereo pair of cameras is likely to be not sufficient in most circumstances and therefore extension to more camera views in the imaging geometry is always desirable. Chen *et al.* (1995) and Faugeras and Mourrain (1995) studied imaging geometries consisting of multiple cameras and derived general solutions for m cameras. Ariyawansa and Clarke (1997) concluded that "*Although it is true that four cameras are adequate for most measurement tasks, future tasks will most likely benefit from additional views solving the problem of many occluded objects in complex manufacturing situations with background clutter*".

3D measurements with consumer grade cameras have become an interesting topic in the photogrammetric research community [Remondino and Fraser, 2006]. Among these cameras, webcams has become the most popular tool available on marketplace for video conferencing nowadays. Its inbuilt USB interface allows video frames being captured and transmitted to the host computer in nearly real time, and then further transmitted through internet connection to the other computer. Inheriting this advantage of direct image data accessibility of webcams, if a suitable software platform can be established in the host computer to directly access and process the image data

captured from webcam hardware, the first component to achieve fast 3D measurement, i.e. fast image data processing, is very likely to be attainable, in contrast to the conventional photogrammetric post data processing after inputting static images captured by digital cameras.

As previously mentioned (Section 3.3.1), the simple construction of hardware components (simple connection between lens and sensor in webcam) might offer webcams an advantage with a relatively stable internal geometry as compared with the much more complex digital SLR cameras with interchangeable lens. Combining with the direct image data accessibility, a potential to shift the role of webcams from video conferencing (or casual use) to low cost but fast close range photogrammetric measurements can be foreseen, noting that provided with an acceptable camera stability, defects from low cost manufacturing process such as a much larger lens distortion (of webcams as compared with high cost digital SLR cameras) can be recovered from the well-established iterative photogrammetric bundle adjustment process (Section 3.5.3.1.4).

To reliably set up and control a webcam based coordination system, the internal platform architectures has to be thoroughly investigated first. Appendix 1, 2 and 3 detail various software development kits (SDKs) towards direct user control over webcams adopted in this research and in particular, issues such as how the application software communicates with webcams (hardware) through the host computer hardware.

3.7 Chapter summary

This chapter has reviewed the photogrammetric and imaging principles necessary to develop a low cost webcam target based measurement system which could be deployed along with optical topography under clinical conditions.

Photogrammetric principles start with the collinearity equations and factors causing departures from collinearity in reality, followed by fundamentals of photogrammetric self-calibrating bundle adjustment. Accuracy and precision of target measurements have been addressed for photogrammetric target based measurement. Target thresholding techniques have also been reviewed, where difficulties of practical target image measurement with webcams in this research have been explored and discussed, followed by associated methods of target recognition and target location adopted in this research with respect to webcam imagery.

Following the 2D target measurements, 3D target measurements require correspondence matching and 3D location computation. Regarding this, principles of the epipolar geometry, the 3D space intersection method to compute the 3D coordinates, together with its correspondence matching optimisations have been reviewed.

In particular, different aspects of off-the-shelf webcams have been addressed, from its simple hardware construction to its internal components such as sensors and optics, as well as the image processing techniques used by these devices and underlying principles of software platforms in order to build such a system (Appendix 1-3).

Chapter 4

Design of a Low Cost Photogrammetric System

4.1 Introduction

This thesis investigates the photogrammetry with low cost webcams where compromises in camera design, particularly the image quality and stability must be mitigated to achieve accurate measurement. This chapter starts by considering each of the key features of these webcams in terms of ideal metric performance, followed by the hardware and software development towards an automated system of accurate 3D measurements with these low cost webcams.

4.2 Accuracy and Precision of Webcams

4.2.1 Introduction

One of the goals in this research is to investigate the suitability of current webcam technology for an accurate 3D coordination of targets located on a medical device (optical topography sensor). Both the accuracy and the precision of target measurement delivered by these low cost webcams are critical in determining if they can be used for optical topography studies. In this study, we are interested in what level of accuracy the webcams can intersect targets and the result will be used as a reference as the accuracy of surface measurement.

Two typical consumer level low cost webcams were selected for this study which commenced in 2009. The first is the Logitech QuickCam Pro 4000 series webcam which is representative of webcams from 2007-2008. The second is the Logitech C500 series webcam which represent a next generation released in 2008-2009. Figure 4 - 1 and Table 4 - 1 compare the two camera models used in the thesis:

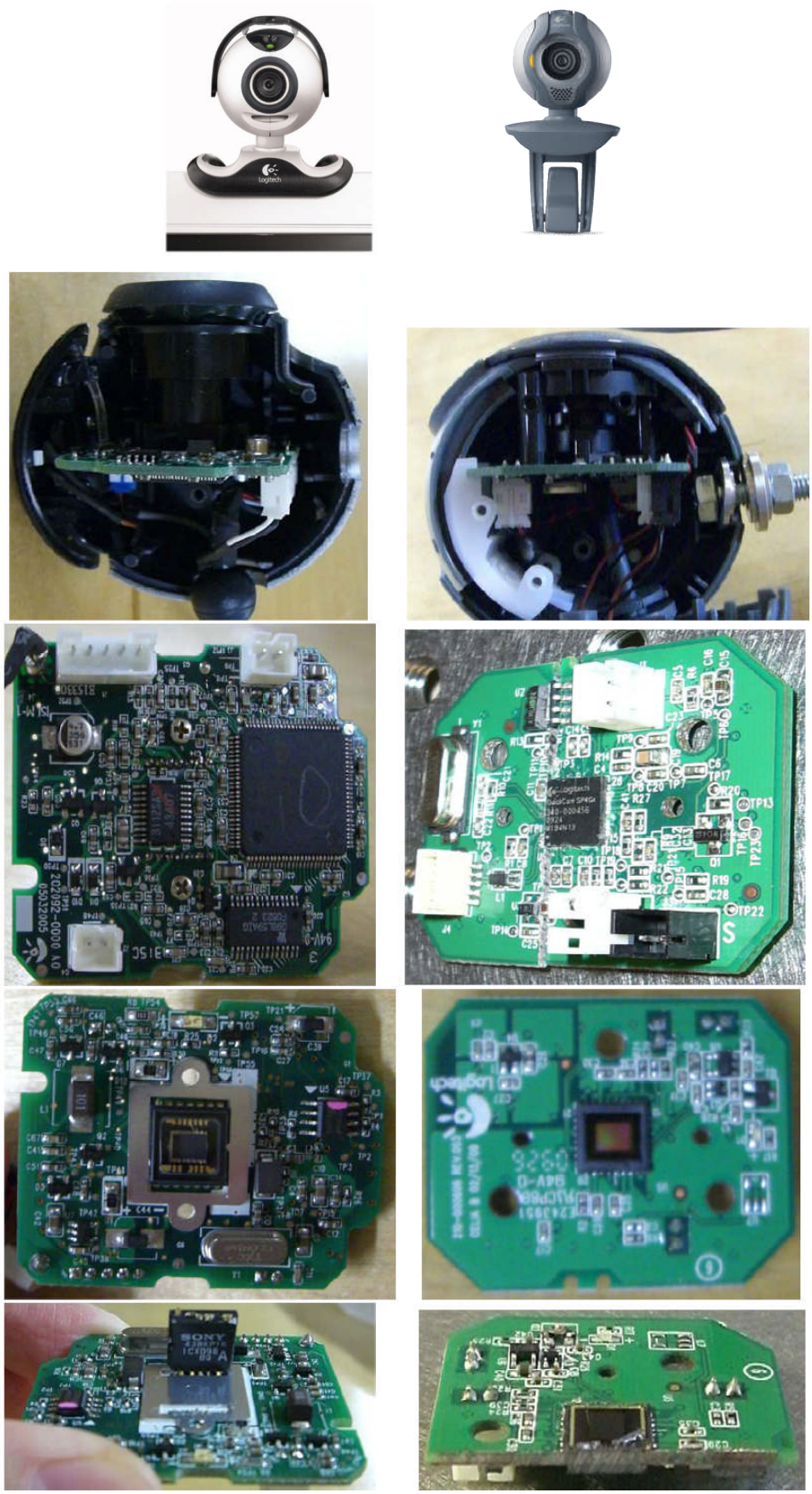


Figure 4 - 1. *Logitech Quickcam Pro 4000 (left) and C500 (right) webcams.*

Image sensor	1: VGA CCD (*SONY ICX098BQ diagonal 4.5mm (type ¼)); 2: CMOS (*Model: unknown) *number of effective pixels: 1: 659×494; 2: unknown *total number of pixels: 1: 692×504; 2: unknown
Lens	Glass lens 1: (4.5mm f/2.2) 2: unknown
Still image-capture resolution	1: 1.3 megapixels; 2: 5 megapixels
Still image-capture format	JPEG
Video-capture resolution	1: Up to 640 x 480 pixels; 2: Up to 1280x1024 pixels
Video-capture format	AVI
Compression	M-JPEG/YUY2
Frame rate	Up to 30 frames per second
Focus adjustment	1: Manual; 2: Fixed focus
Lighting	1: Ambient indoor incandescent, fluorescent (50 or 60 Hz flicker-free), daylight; 2: RightLight technology
Controls	Digital zoom
Built-in flash	No
Driver	1: PWC (Philips USB webcam driver); 2: Logitech UVC (USB Video Class)
Interface	1: USB 1.1; 2: USB 2.0
Power	Powered through USB

Table 4 - 1. *Logitech Quickcam Pro 4000 (1) and C500 (2) Webcam Specification.*

** 1: (Sony datasheet of CCD image sensor ICX098BQ, 2003);*

**2: Remain unknown after disassembly of webcam.*

In this study the focus on two webcam models was to investigate accuracy and precision delivered in order to find out if these systems are suitable for accurate photogrammetric measurement in optical topography studies. The C500 webcams, besides accuracy and precision validation, were investigated to verify if there are any significant changes in terms of photogrammetric target image quality, stability and

reliability compared to earlier generations. Images necessary for detecting relative motion between head and sensing pads during optical topography studies require nearly real time video stream collection. Work therefore focussed on this capture mode since it is highly reliant on camera hardware. Results are presented as a comparison between hardware images (extracted video frames) and software images (still images) for C500 webcams used in subsequent OT tests.

4.2.2 Method

4.2.2.1 Experimental Setup

As a first experiment to study accuracy and precision, webcams were arbitrarily located in a half circle shape. This imaging geometry provides convergent photogrammetric recovery of target coordinates and provides some tolerance to occlusion given small rotations of the subject. Imaging of a calibration object was carried out at a range of approximately 300 mm (Figure 4 - 2). In the test situation the webcams were located on tripod heads positioned around 250mm higher than the calibration object so that each webcam could see as many targets as possible. The manual focus of each webcam was adjusted to provide a sharp image of the targets on the target array.

Figure 4 - 2. *Sample image showing webcam locations with respect to the calibration object.*

4.2.2.2 Accuracy and precision validation

Accuracy validation of the selected webcams was performed by comparing target intersection results from webcams with a standard. For both Pro 4000 and C500 webcams the 3D target positions computed from a convergent image network consisting of 10 images made with a high resolution (6MP) Nikon D100 camera fitted with a fixed focus 28mm Nikon lens, supported with inter-target distances measured with calipers, were used as a standard result. These data were then compared against target coordination results from webcam images at two still image settings: lower resolution (pixel size: 8μm; image size: 640, 480) and: higher resolution (pixel size: 4μm; image size: 1280, 960 for the Pro 4000 model and 1280, 1024 for the newer C500 model). Whilst the number of pixels for each webcam was known, the actual pixel size was unknown at this point (Section 3.3.1.1). Since the camera is only determining directions, provided that the pixels are square only the ratio between the principal distance and the pixel dimensions need be estimated. Since the principal distance can be estimated from the self-calibrating bundle adjustment, an arbitrary pixel size can be chosen accounting for the selected pixel size(s) above.

4.2.2.3 Camera calibration

Webcams are designed for low cost imaging rather than accurate measurement. Each camera will depart from the ideal perspective projection (Section 3.2.2) and must be calibrated in order to correct for systematic errors which would otherwise propagate into the estimated target coordinates. Furthermore, whilst they are manufactured on a production line, small variations in the optics, sensor surface flatness and alignment in each webcam will give rise to small differences in imaging geometry from one sensor to the next. Each camera should therefore be calibrated to determine its own individual set of correction parameters (Section 3.2.2.4). To generate an image network suited to calibration, the calibration object was rotated to provide 10 convergent images for each camera.

Reliable camera calibration requires that a convergent network of images of a rigid object are taken from each camera, the assumption being that the internal imaging geometry of the camera remains constant. Since the camera orientation parameters from calibration will be used for 3D target coordination in clinic, the ideal case is to calibrate these cameras prior to each optical topography study to make sure the internal imaging geometry of each camera is constant between each calibration and target coordination. However, individual calibration represents a significant overhead for an optical topography system so an experiment comparing the suitability of one common calibration for all cameras in the network vs. individual calibrations (Section 4.3) was designed. Meanwhile, this early study also simulated the optical topography application by implementing target intersection right after the calibration data from camera calibration had been obtained. A total of four data sets were generated for target intersection for Pro 4000 webcams:

1. 640*480 with uniform calibration parameters for each camera;
2. 1280*960 with uniform calibration parameters for each camera;
3. 640*480 with individually calibrated parameters for each camera;
4. 1280*960 with individually calibrated parameters for each camera.

For the study of C500 webcams, the same calibration object was used and individual camera calibration was performed, where the setup and image generation for the network are similar to those for Pro 4000 webcams but not acquired at the same time. The C500 webcam has a fixed focus which is not adjustable as compared to the adjustable focus for the Pro 4000 webcam. Four calibrations were carried out using one C500 webcam but with the following four different imageries:

1. 640*480 software resolution (pixel size: 8μm);
2. 640*480 hardware resolution (pixel size: 8μm);
3. 1280*1024 software resolution (pixel size: 4μm);
4. 1280*1024 hardware resolution (pixel size: 4μm, the highest hardware resolution).

4.2.3 Results

4.2.3.1 Bundle adjustment data

Table 4 - 2 summarises the calibration variables from the Nikon D100 network:

Program Control Variables	
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	4.55
Minimum images for a network target	4
Rejection criterion for image errors	5.0μm
Input Summary	
Number of camera calibration sets	1
Number of target image observations	479
Total number of exposures	7
Number of exposures in the network	7
Total number of targets	94
Number of targets in the network	70
Results for the calibration solution	
Unit weight estimate (sigma zero)	0.95
RMS image residual (μm)	3.88
Number of observables in the network	909
Number of unknowns in the network	269
Number of redundancies in the network	640
Mean number of images per target	6.4
Target Precision Summary	
Mean precision of target coordinates (μm)	59.43
Relative precision for the network (1: X)	5000

Table 4 - 2. *Summary of camera calibration for Nikon D100 network.*

Table 4 - 3 summarises the variables for each of the four Pro 4000 webcams at two image resolutions (640×480 and 1280×960), where relative precision for the network represents the maximum dimension in the photogrammetric targets divided by the mean 3D target precision/uncertainty. In the table "L" denotes the low resolution setting and "H" the high setting.

Program Control Variables	**Camera 1/2/3/4 L: 640×480; H: 1280×960**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	**L**: 0.5/0.5/0.5/1.12; **H**: 0.73/0.63/0.5/0.45
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**Camera 1/2/3/4 L: 640×480; H: 1280×960**
Number of camera calibration sets	1
Number of target image observations	**L**: 127/128/127/78; **H**: 277/288/281/273
Total number of exposures	10
Number of exposures in the network	10
Total number of targets	33
Number of targets in the network	**L**: 13/13/13/9; **H**: 31/32/32/32
Results for the calibration solution	**Camera 1/2/3/4 L: 640×480; H: 1280×960**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	**L**: 1.12/0.98/1.03/0.71; **H**: 0.42/0.54/0.43/0.37
Number of observables in the network	**L**: 266/266/264/164; **H**: 562/588/566/538
Number of unknowns in the network	**L**: 111/111/111/99; **H**:165/168/168/168
Number of redundancies in the network	**L**: 155/155/153/65; **H**: 397/420/398/370
Mean number of images per target	**L**: 9.8/9.8/9.8/8.4; **H**: 8.9/9.0/8.7/8.2
Target Precision Summary	**Camera 1/2/3/4 L: 640×480; H: 1280×960**
Mean precision of target coordinates (μm)	**L**: 33.11/28.31/30.29/43.26; **H**: 12.76/16.10/13.55/14.29
Relative precision for the network (1:X)	**L**: 6000/7000/7000/4000; **H**: 17000/13000/16000/15000

Table 4 - 3. *Summary of individual camera calibration for the 4 Pro 4000 webcams.*

Table 4 - 4 summarises the variables from calibration with the four different settings described in Section 4.2.2.3 (denoted 1/2/3/4) for the C500 webcams:

Program Control Variables	**1/2/3/4**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	0.51/0.41/0.23/0.36
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**1/2/3/4**
Number of camera calibration sets	1
Number of target image observations	182/177/166/208
Total number of exposures	10
Number of exposures in the network	10
Total number of targets	33
Number of targets in the network	25/25/21/26
Results for the calibration solution	**1/2/3/4**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	1.04/1.04/0.86/0.8
Number of observables in the network	344/338/284/402
Number of unknowns in the network	147/147/135/150
Number of redundancies in the network	197/191/149/252
Mean number of images per target	6.6/6.5/6.5/7.5
Target Precision Summary	**1/2/3/4**
Mean precision of target coordinates (μm)	32.31/32.33/32.54/24.39
Relative precision for the network (1: X)	7000/6000/5000/7000

Table 4 - 4. *Summary of camera calibration for one C500 webcam at four different settings.*

4.2.3.2 Image quality

Figure 4 - 3 compares the image quality delivered by the older Pro 4000 (2007) and the newer C500 (2009) webcams:

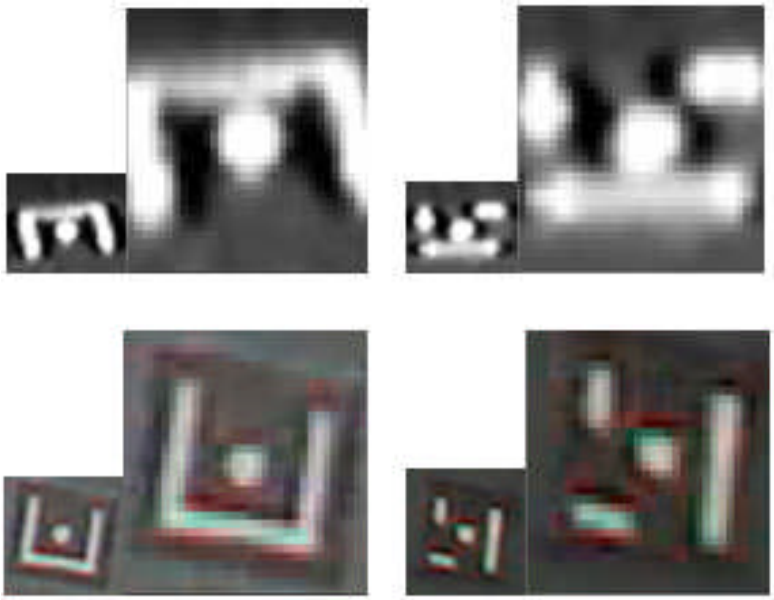


Figure 4 - 3. *Comparison of image quality between Pro 4000 (upper row: 1280×960) and C500 (lower row: 1280×1024) webcams.*

(Left: original target image; Right: Window Size (pixels): 32, 32)

4.2.3.3 Target coordinate uncertainty

The uncertainty of the 3D target measurement can be expressed through the mean precision values of intersected targets, as shown in Figure 4 - 4 (information corresponding to each data set is described in Section 4.2.2.3):

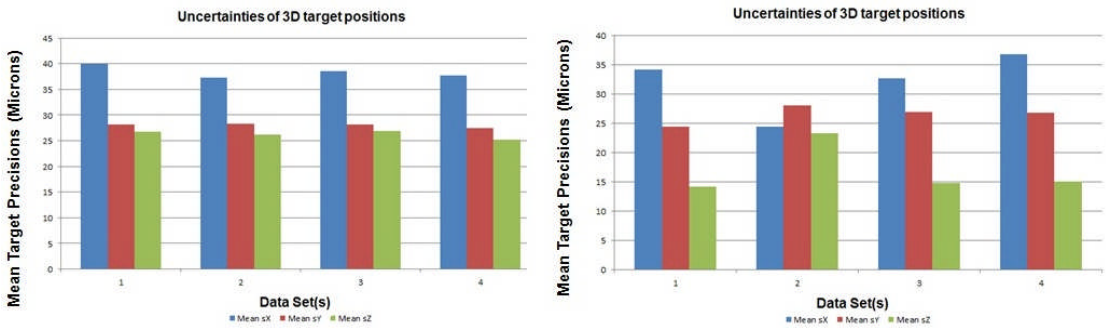


Figure 4 - 4. *Comparison of 3D target precisions of Pro 4000 (left) and C500 (right) webcams.*

Mean absolute target image residuals of the four data sets are shown in Figure 4 - 5:

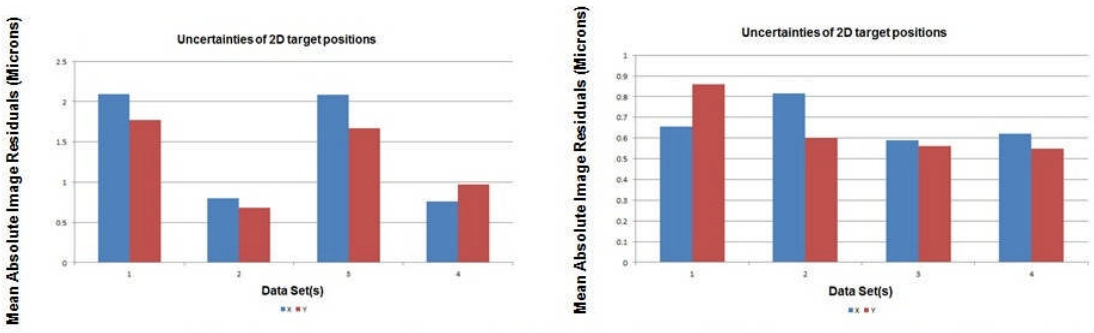


Figure 4 - 5. *Comparison of target image residuals of Pro 4000 (left) and C500 (right) webcams.*

4.2.3.4 Accuracy

To compare the accuracy among different data sets for each webcam model, summary statistics of the difference computed between each of the data sets and the standard values generated from the Nikon D100 reference photogrammetric survey are presented in Figure 4 - 6. Since all these data are on the same coordinate datum as that of Nikon D100, a check of the discrepancy values here is a valid method to compare the accuracy levels among different data sets.

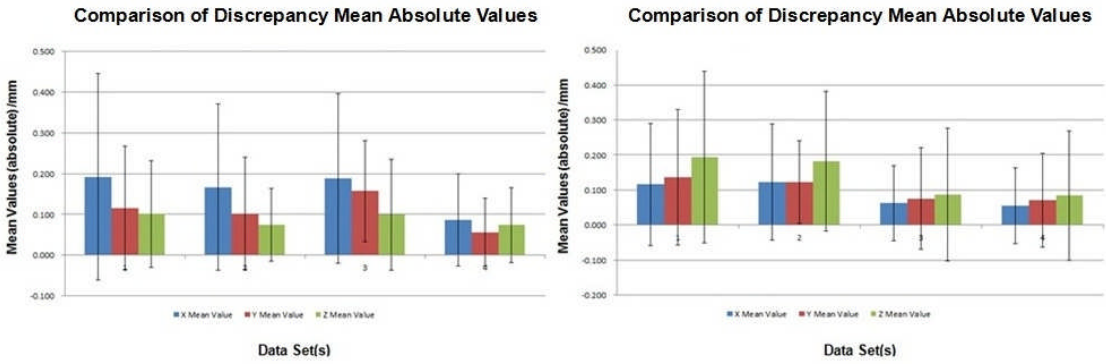


Figure 4 - 6. *Accuracy of target locations of Pro 4000 (left) and C500 (right) webcams.*

4.2.4 Discussion

In this research only the principal point of offset in x and y, principal distance and 3rd power term of radial lens distortion were considered due to the fact that only these parameters are significant (Section 3.2.3.3). Table 4 - 5 gives a sample output from one of the bundle adjustments:

Camera parameters	Value (mm)	Precision (mm)
Principal point offset in x	0.1104	0.0095
Principal point offset in y	0.0056	0.0115
Principal distance	6.5313	0.0143
3rd power term of radial lens distortion	-1.6741e-003	5.7467e-004
5th power term of radial lens distortion	6.4072e-004	3.3451e-004
7th power term of radial lens distortion	-9.0793e-005	6.4431e-005
1st term of tangential lens distortion	-2.4851e-004	8.8672e-005
2nd term of tangential lens distortion	-1.3493e-004	1.2445e-004
Orthogonality of image coordinate system	-3.9444e-004	1.1778e-004
Affinity of image coordinate system	-6.9066e-004	2.6284e-004

Table 4 - 5. *Sample output from bundle adjustment for Pro 4000 webcam.*

Logitech QuickCam Pro 4000 webcams

Figure 4 - 6 suggests that a network with individually calibrated webcams (4) has achieved higher accuracy compared with the network of webcams with uniform parameters (2). This is because the imperfection from the manufacturing process of these low cost consumer cameras has caused a variation of internal parameters across cameras even if they belong to the same model. Individual calibration enables individual precision for each of these parameters which can better constraint the network by reducing the ambiguity for target intersection and therefore achieve a higher accuracy level of target coordination.

In terms of precision, Figure 4 - 5 suggests that the higher resolution imagery (2, 4) has produced much smaller image residuals than the lower resolution imagery (1, 3). Image

resolution only describes how many samples are captured but not necessarily their quality being captured. Consequently it is possible for an image with higher resolution to deliver a poorer target image quality. Such a decrease in quality has a direct impact on target centroid location. One possibility for the decreased image quality could be the inability for its lens to preserve target contrast as resolution increases [McHugh, 2005].

Figure 4 - 5 shows that for the Logitech QuickCam Pro 4000 webcams, the uncertainty in the target image coordinates at a resolution of 1280×960 (2, 4) is about half that achieved with a resolution of 640×480 (1, 3). This suggests that the lens in this webcam has preserved the image contrast at the higher resolution of 1280×960. Moreover, for this camera any still images with more than its maximum hardware resolution (640×480) have software interpolation, which estimates pixel values at new pixel centres based on the values at surrounding pixels. For webcams, most probably bicubic interpolation algorithm (Section 3.3.1.3) is used. Although there could be a loss of overall image quality when performing such an interpolation, results from Figure 4 - 5 suggests that there is some benefit in the resampling process used within the Logitech Pro 4000 webcams.

Logitech C500 webcams

In terms of 3D target coordinate uncertainty, Figure 4 - 4 suggests that there is no difference between lower (1, 2) and higher (3, 4) resolutions. For target image coordinate uncertainty, Figure 4 - 5 suggests that there is no difference between the software (1 and 3) and hardware (2 and 4, respectively) images. That is for C500 webcams, there is no difference of image quality between software and hardware images in terms of photogrammetric target image measurement. However, when the resolution increases from 640×480 (1, 2) to 1280×1024 (3, 4), there is an average decrease of about 0.1 μm in target image coordinate uncertainty. In terms of 3D target coordination accuracy, Figure 4 - 6 again suggests that there is no difference between software (1 and 3) and hardware (2 and 4, respectively) images in terms of measurement accuracy. However, when the resolution increases from 640×480 to 1280×1024, there is an increase of about 0.05 mm for target coordinate accuracy. This has agreed with the conclusion drawn from Pro 4000 results that a higher resolution is

able to deliver a higher target 3D coordinate accuracy.

Comparison between Pro 4000 and C500 webcams

One of the most noticeable differences between the Pro 4000 and C500 webcams is the image quality. As shown from Figure 4 - 3, the boundary between target image and its background delivered by C500 is much clearer than those delivered by the Pro 4000 model, provided with the same relative distance between the webcam and the object. Further evidence is that when image resolution is 640×480, target image coordinate uncertainty for C500 is one third as that for Pro 4000 (Figure 4 - 5). The clearer images delivered by C500 in terms of boundary between targets and their background have made C500 model a better choice for target recognition for rapid target tracking applications.

Looking inside the webcams (Figure 4 - 1) suggests that the sensors used in both models have similar imaging areas, which should give the CCD sensor (Pro 4000) larger pixels and therefore a better light gathering potential as well as a higher signal-to-noise ratio, compared with the CMOS sensor in the C500. However, as technology for consumer level cameras has evolved rapidly, the maximum resolution of still image has increased from 1.3 megapixels (Pro 4000) to 5 megapixels (C500) (Table 4 - 1). The clearer images (Figure 4 - 3) suggest that the software enhancement on images from C500 has compensated the potential disadvantages of its CMOS sensor. Unfortunately how these images are enhanced for C500 model remains confidential to Logitech, as noted earlier there are many possible techniques used with cameras equipped with CMOS sensor for enhancing image quality and reducing noise (Section 3.3.1.1).

In terms of target image coordinate uncertainty for both camera models, Figure 4 - 5 suggests that when image resolution is 640×480, 2D uncertainty of target images with the C500 is one third that of the Pro 4000. When the resolution is higher (1280×960 for the Pro 4000 and 1280×1024 for the C500), the difference between the two models decreases, but the uncertainty of the C500 is still 0.2 μm less than that of Pro 4000. Also, Figure 4 - 6 suggests that the 3D target coordination accuracies for both models are

within 0.2 mm for lower image resolution and within 0.1mm for higher image resolution. This has agreed with the results (the same Logitech Pro 4000 webcam and 640×480 resolution setting) presented by Page *et al.* (2008) that it is possible to achieve measurement accuracy of 0.3mm with measurement distance of 300mm (Section 3.3.1). All these have demonstrated that for both webcam models, a higher image resolution is able to deliver target coordinates with lower uncertainty and higher accuracy.

Factors contributing to target measurement inaccuracy

There are some factors which can cause the discrepancy of target coordination:

1. Low spatial resolution

In the visual perception of the real world, contrast is determined by the difference in the colour and brightness of the object and other objects within the same field of view [Travnikova, 1985]. Image intensity variation of background is clearly shown for the Pro 4000 webcams (upper row of Figure 4 - 3), which has directly caused the confusion between the actual target and its background.

The lightness of an image captured by a camera is determined by the exposure of the image, which is determined by aperture, shutter speed, sensor sensitivity and any subsequent image compression (Section 3.3.1.2). When targets are over-exposed saturating the sensor, array read-out effects or even ‘blooming’ caused by the pixels in the sensor array exceeding their charge capacity will occur. As shown from upper row of Figure 4 - 3, both of the enlarged images are effectively “immeasurable” by standard centroiding techniques since the central dot and the code are joined and therefore many targets were manually measured rather than automatic centroiding in which the measurement error propagated to target 3D coordinates resulting a certain amount of inaccuracy.

2. Lens quality

When comparing different lenses the apparent quality of the image will be mostly determined by how well each lens preserves contrast as line pair frequency (the number of lines which fit within a given unit length) increases. Resolution limited by diffraction does not apply to this case due to the fact that resolution can only be limited by diffraction when the aperture is small, e.g. f/22. The 4.5mm f/2.2 glass lens [CNN review, 2011] for the Logitech QuickCam Pro 4000 model for example, notes that the spatial resolution directly links to the lens quality.

3. Image Measurement Residuals

Image residuals indicates the difference between measured and computed image locations. Acceptable image residual magnitudes are normally determined by the size of the pixels of the imaging sensor, the image measurement method and the target image quality. In this experiment some target images were manually measured and can therefore expect to realize RMS residuals of no better than 1/2 to 2/3 of a pixel. Figure 4 - 7 gives some examples of target images with large image residuals:

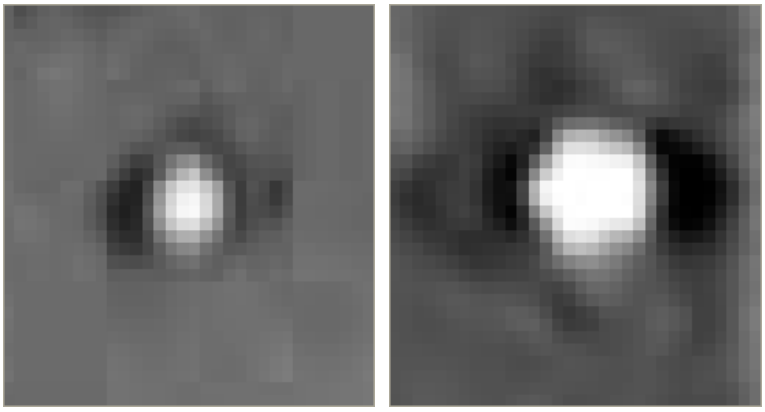


Figure 4 - 7. *Examples of target points with large residual values from Pro 4000 webcam. Mean image residuals: Left: (4.54µm, 3.3µm); Right: (4.39µm, 3.37µm); (Target Image Window Size (pixels): 32, 32)*

In the above cases the dark band on either side of the target is likely to be due to “ringing” where the high intensity signal influences the electronic readout of the surrounding pixels. The jpeg blocking is also very evident in the left target image. This increases the difficulty for the centroiding process to predict the target location and

therefore it is likely to produce a large image residual. Furthermore, this internal measure suggests that the imaging geometry and relatively small calibration object with a small number of targets on it (Figure 4 - 2) may not support the larger number of calibration parameters to be estimated in the full case as well as a much reduced set. Further investigation with a calibration object which more closely fills the field of view of the webcam images and has targets that are comparable with their optical quality is required (Section 4.3).

4.2.5 Summary

Although a higher accuracy level of target coordination can be achieved by individual calibration, individual camera calibration prior to each optical topography study for a system containing multiple cameras becomes impractical. This impracticality is due to the fact that manual cost is involved to measure the target images since the poor image quality delivered from this range of cameras has made fully automatic target centroiding measurement impossible. Regarding this an efficient calibration method will be to first create an image network of all the cameras with an initial uniform parameter set. Camera parameters will then be iteratively optimised through a self-calibrating bundle adjustment performed on the data processing of all cameras within the same adjustment (Section 3.2.3.2). However, given the significant improvement of image quality seen in the C500 webcam over the earlier Pro 4000 webcam, completely automation of target centroiding measurement should be possible with new generations of webcams. As a result the time required for camera calibration should not be considered as an issue for optical topography studies.

To summarise the study in this section:

1. The data are considered to be accurate enough for optical topographic medical tracking applications being better than the 1mm tolerance for topographic reconstruction (Section 2.2.3);

2. A significant improvement of image quality in terms of target recognition has been found between different generations of webcams. This demonstrates a substantial potential for future applications of automatic photogrammetric target measurements with low cost webcams;

3. The higher image resolution from webcams is able to deliver target measurements with better than 0.1mm accuracy for all X, Y and Z coordinates. Again this demonstrates a significant reduction in the gap between high end digital cameras and low cost consumer level webcams in terms of many photogrammetric applications requiring millimeter or even sub-millimeter accuracy level.

4.3 Self-calibrating bundle adjustment with a larger calibration object

4.3.1 Method

For practical imaging where multiple cameras are required, the accuracy and precision of the estimated coordinates will rely heavily on the stability of the cameras. As mentioned in Section 4.2.4, an imaging geometry of a relatively small calibration object with a small number of targets on it may not support the full image area to be used in the practical case. This section describes the capabilities of the C500 system in terms of 3D target measurement precision with a calibration object which more closely fills the field of view of the webcam images, under the maximum hardware resolution of C500 webcams (1280×1024).

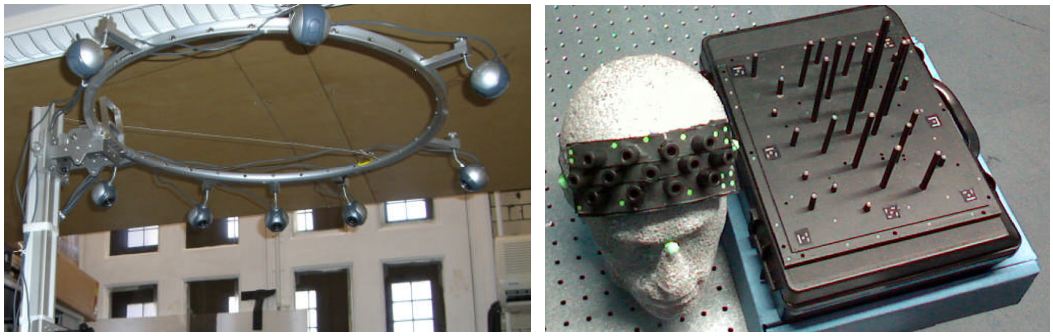


Figure 4 - 8. *A network of 8 Logitech C500 webcams (left) and a calibration object with larger volume (right).*

In order to study the precision of an eight camera C500 network, which has been designed for fast tracking of a target pad on a human head, the webcams were rigidly mounted in a circle with diameter of about 0.7m (Figure 4 - 8 left) at a range of 700mm. Such an imaging geometry provides convergent photogrammetric estimation of target coordinates and provides some tolerance to occlusion given small rotations of the subject. Imaging of a new calibration object (Figure 4 - 8 right) of similar dimensions to a human head was carried out at a range of approximately 600-800 mm providing a maximum target coverage within all camera viewpoints. A key advantage of the C500 webcam is that it is fixed focus and should be physically more stable than the focusable

Pro 4000 webcams. To calibrate the eight cameras, the calibration object was rotated into eight sequential positions. For each position images were taken with all eight webcams to provide a network of 64 images. A self-calibrating bundle adjustment was performed on the data processing of all eight cameras within the same adjustment.

4.3.2 Results

Table 4 - 6 summarises the calibration variables from the Nikon D100 network:

Program Control Variables	
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	2.43
Minimum images for a network target	4
Rejection criterion for image errors	5.0μm
Input Summary	
Number of camera calibration sets	1
Number of target image observations	777
Total number of exposures	14
Number of exposures in the network	14
Total number of targets	61
Number of targets in the network	61
Results for the calibration solution	
Unit weight estimate (sigma zero)	1.00
RMS image residual (μm)	1.95
Number of observables in the network	1568
Number of unknowns in the network	281
Number of redundancies in the network	1287
Mean number of images per target	12.7
Target Precision Summary	
Mean precision of target coordinates (μm)	51.32
Relative precision for the network (1: X)	8000

Table 4 - 6. *Summary of camera calibration for Nikon D100 network.*

Table 4 - 7 summarises the variables from the self-calibrating bundle adjustment:

Program Control Variables	**Values**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	0.34
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**Values**
Number of camera calibration sets	8
Number of target image observations	2367
Total number of exposures	64
Number of exposures in the network	64
Total number of targets	61
Number of targets in the network	61
Results for the calibration solution	**Values**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	1.05
Number of observables in the network	4719
Number of unknowns in the network	614
Number of redundancies in the network	4105
Mean number of images per target	38.3
Target Precision Summary	**Values**
Mean precision of target coordinates (μm)	31.12
Relative precision for the network (1:X)	13,000

Table 4 - 7. *Summary of self-calibrating camera calibration for eight Logitech C500 webcams.*

The uncertainties of 3D and 2D target positions are shown in Figure 4 - 9:

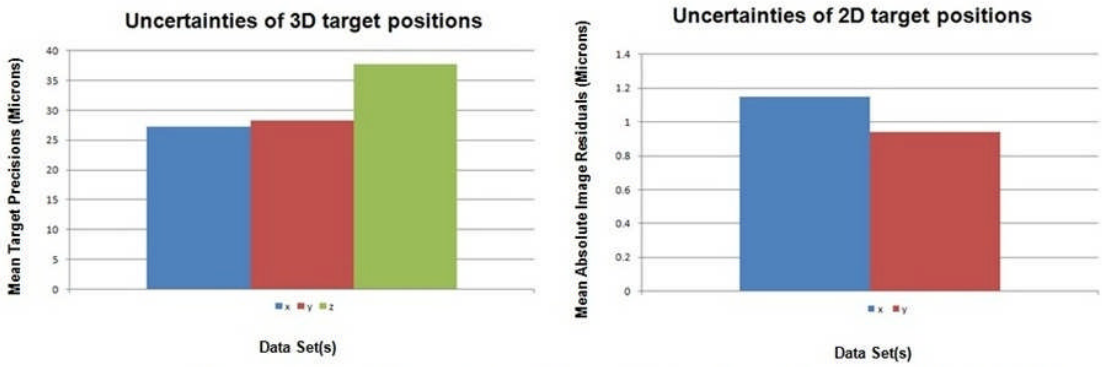


Figure 4 - 9. *Mean values of 3D target precisions (left) and target image residuals (right).*

4.3.3 Discussion

Compared with Figure 4 - 9 which has the same resolution setting ((4) on the right images of Figure 4 - 4, Figure 4 - 5 and Figure 4 - 6), a self-calibrating bundle adjustment containing all the imagery from all cameras is able to achieve a similar level of 2D and 3D target measurement uncertainty and target coordinate accuracy ((2) in Figure 4 - 18) as compared with results generated from individual camera calibrations. The imaging distance between each camera and calibration object has been doubled from 300mm to 600-800mm in this self-calibrating bundle adjustment in order to represent the most likely OT imaging case. The number of targets has also been increased from 32 (small calibration object, Figure 4 - 2) to 61 (large calibration object, Figure 4 - 8). Furthermore, the redundancies in the self-calibrating network with all the eight webcams have increased to 4105 (Table 4 - 7), as compared with a redundancies of 252 ((4) from Table 4 - 4) for the individual calibration. This level of redundancy in the network has allowed a better estimation of camera parameters as well as target coordinates in the self-calibrating bundle adjustment.

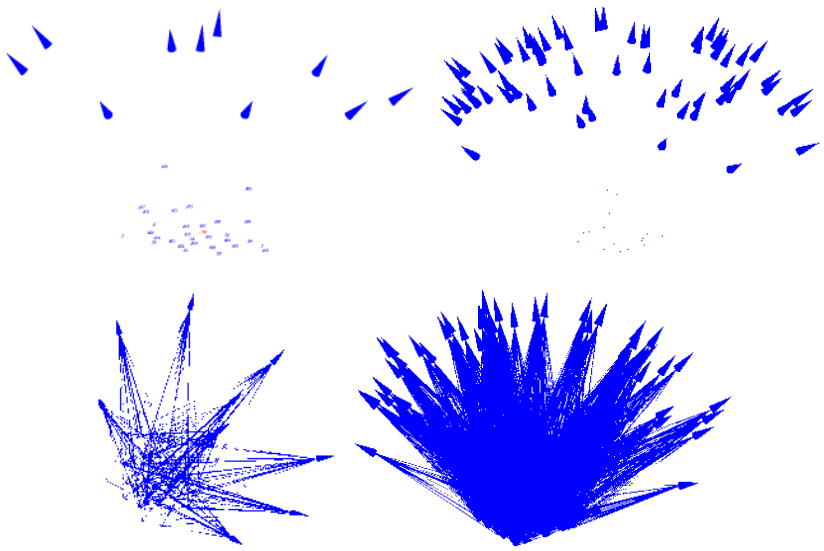


Figure 4 - 10. *Individual calibration network of 1 C500 webcam (left) Vs. Self-calibrating bundle adjustment network of eight C500 webcams (right). Upper left: A total of 10 images and the smaller calibration object in the network; Upper right: A total of 64 images and the larger calibration object in the network; Lower left: Mean number of images per target = 7.5; Lower left: Mean number of images per target = 38.3.*

4.3.4 Summary

This section has demonstrated that even at a longer range of between 600mm and 800mm, a network of eight Logitech C500 webcams, at their maximum hardware resolution of 1280×1024, is able to achieve an average precision of about 30µm in terms of 3D target coordination. The system has been calibrated under a self-calibrating bundle adjustment containing all the images from all webcams to achieve image residuals of about 1µm. This level of precision has saved the effort needed for individual camera calibrations prior to each optical topography study. Therefore a simpler and quicker set up of the system for optical topography studies can be achieved in this way. Further assessment of reliability (Section 4.4) for these webcams over a period of time is required in order to prove the camera stability in terms of target coordination accuracy. If stability can be proven and the tracking system is proven to be reliable over a relatively long period of time such as month(s), no calibration will be required between optical topography studies, greatly simplifying system use.

4.4 Camera Stability

4.4.1 Introduction

The interior orientation parameters of a camera are the prerequisite of target positioning. The simple construction of low cost webcams could be an advantage as the sensor and lens are physically connected when compared with high end digital cameras which use a combination of springs and hinges (Section 3.3.1). One interesting question is – how stable are the interior orientation parameters of these low cost webcams in terms of meeting the OT requirement for fast tracking? In this study, the stability of a number of interior orientation parameters were determined when camera calibrations were carried over a period of time. If the parameters are stable (i.e. the effect of differences are of the order of a few μm on the image plane), target coordination with networks of these cameras can be carried out without the need to recalibrate cameras prior to each optical topography study.

4.4.2 Logitech QuickCam Pro 4000 Webcams

4.4.2.1 Method

This work is confined to Logitech QuickCam Pro webcams, where the interest is to see how the parameters vary on day-to-day basis. To achieve this aim, similar repetitive camera calibrations were carried out on different days. The procedure used copied that in Section 4.2.2.3. For Set 1 (individual calibration), the calibration object was rotated each time a convergent image was taken by a webcam. Ten images were taken with each webcam. Individual calibrations for the four webcams were repeated on three consecutive days. For Set 2 (calibration of all six webcams together carried out in one day), the calibration object was rotated each time a total of six images were taken by the six webcams and there were twelve images for each webcam. Table 4 - 8 gives the details for this study:

Set 1 (Fixed Focus 1)			**Set 2 (Fixed Focus 2)**	
		Same camera used?		
Number of webcams	4		**Number of webcams**	6
		No	Cam 0	Trust WB-5600R
Cam 1	Logitech Quickcam Pro 4000	Yes	Cam 1	Logitech Quickcam Pro 4000
Cam 2	Logitech Quickcam Pro 4000	Yes	Cam 2	Logitech Quickcam Pro 4000
Cam 3	Logitech Quickcam Pro 4000	Yes	Cam 3	Logitech Quickcam Pro 4000
Cam 4	Logitech Quickcam Pro 5000	No	Cam 4	Logitech Quickcam Pro 4000
		No	Cam 5	Trust WB-5600R
Calibration Method	Individual calibration, 3 consecutive days		**Calibration Method**	Calibration of all cameras together, 1 day

Table 4 - 8. *Experiment details for camera stability study.*

The parameters to be studied include the image coordinates of principal point, the principal distance and radial distortion. In all cases calibration of the cameras was carried out using a self-calibrating bundle adjustment procedure with a coordinate datum based upon prior measurement of the smaller calibration object (Figure 4 - 2) with a Nikon D100 camera.

4.4.2.2 Results

Table 4 - 9 summarises the parameters from calibrations for Set 1 and Set 2:

Program Control Variables	**Set 1: Cam 1/2/3/4; Set 2**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	**Set 1**: Day 1: 0.73/0.63/0.5/0.45; Day 2: 0.54/0.86/0.63/0.52; Day 3: 0.59/0.84/0.53/0.44; **Set 2**: 1.88
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**Set 1: Cam 1/2/3/4; Set 2**
Number of camera calibration sets	**Set 1**: 1; **Set 2**: 6
Number of target image observations	**Set 1**: Day 1: 277/288/281/273; Day 2: 110/121/116/116; Day 3: 124/119/118/117; **Set 2**: 1906
Total number of exposures	**Set 1**: 10; **Set 2**: 72
Number of exposures in the network	**Set 1**: 10; **Set 2**: 72
Total number of targets	33
Number of targets in the network	**Set 1**: Day 1: 32; Day 2: 13; Day 3: 13; **Set 2**: 31
Results for the calibration solution	**Set 1: Cam 1/2/3/4; Set 2**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	**Set 1**: Day 1: 0.42/0.54/0.43/0.37; Day 2: 0.34/0.57/0.33/0.31; Day 3: 0.45/0.6/0.39/0.3; **Set 2**: 1.72
Number of observables in the network	**Set 1**: Day 1: 562/588/566/538; Day 2: 232/254/244/244; Day 3: 260/250/248/246; **Set 2**: 3723
Number of unknowns in the network	**Set 1**: Day 1: 165/168/168/168; Day 2: 111/111/111/111; Day 3: 111/111/111/111; **Set 2**: 562
Number of redundancies in the network	**Set 1**: Day 1: 397/420/398/370; Day 2: 121/143/133/133; Day 3: 149/139/137/135; **Set 2**: 3161
Mean number of images per target	**Set 1**: Day 1: 8.9/9/8.7/8.2; Day 2: 8.5/9.3/8.9/8.9; Day 3: 9.5/9.2/9.1/9; **Set 2**: 59.5
Target Precision Summary	**Set 1: Cam 1/2/3/4; Set 2**
Mean precision of target coordinates (μm)	**Set 1**: Day 1: 12.76/16.1/13.55/14.29; Day 2:

Relative precision for the network (1:X)	12.66/19.91/12.6/13.43; Day 3: 15.32/19.7/13.45/12.68; **Set 2**: 16.54 **Set 1**: Day 1: 17000/13000/16000/15000; Day 2: 17000/11000/17000/16000; Day 3: 14000/11000/16000/17000; **Set 2**: 15000

Table 4 - 9. *Summary of calibration parameters for Set 1 (individual calibration of 4 cameras over 3 days) and Set 2 (self-calibrating bundle adjustment of 6 cameras together in 1 day).*

4.4.2.2.1 Principal point offsets

Figure 4 - 11 presents the difference in the estimated principal point offset in the x direction for webcams 1 through 4 of Set 1 and webcams 0 through 5 of Set 2. Similarly Figure 4 - 12 presents differences in y for both sets. Since all webcams in Set 2 were calibrated at a different time (about one year later) from webcams in Set 1, the manual focus setting between the two sets is regarded as being independent.

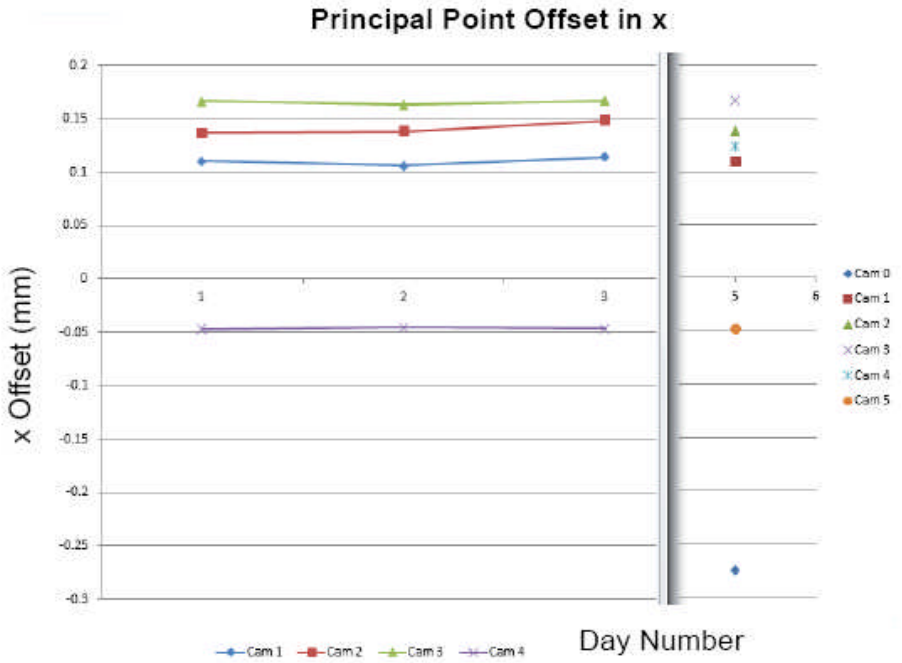


Figure 4 - 11. *Principal point offset in x for Set 1 over 3 consecutive days and Set 2 in 1 day.*

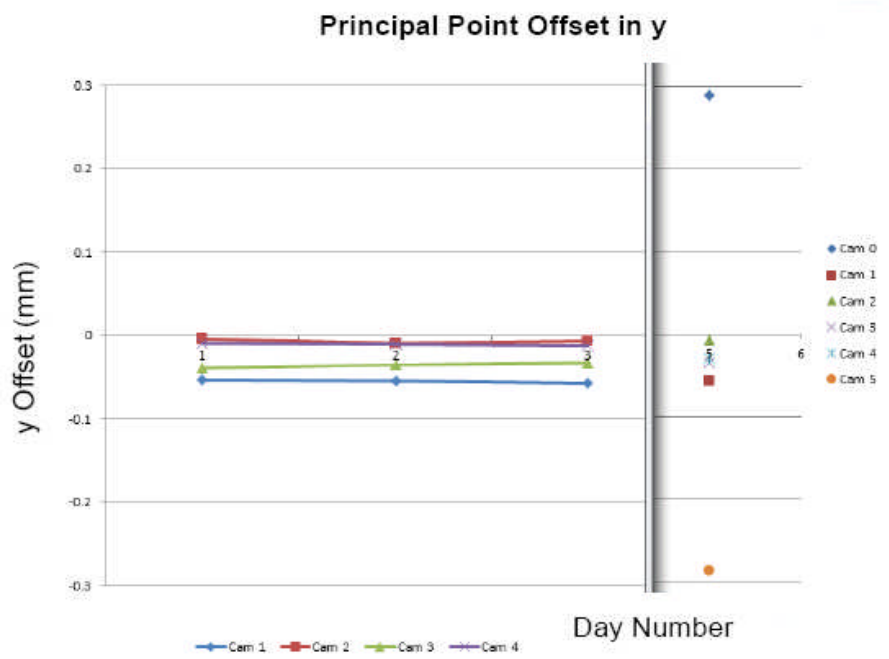


Figure 4 - 12. *Principal point offset in y for Set 1 over 3 consecutive days and Set 2 in 1 day.*

4.4.2.2.2 Principal distance

Figure 4 - 13 presents the difference of principal distances for webcams 1-4 of Set 1 and the difference of principal distances for webcams 0-5 for Set 2.

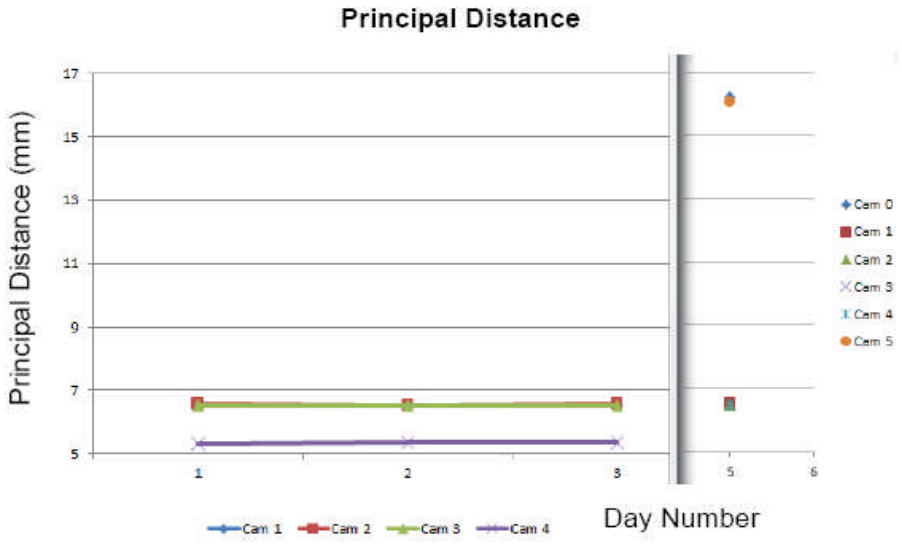


Figure 4 - 13. *Principal distance for Set 1 over 3 consecutive days and Set 2 in 1 day.*

4.4.2.2.3 Radial distortion

A typical result of radial distortion for Logitech QuickCam Pro 4000 cameras was selected to present in Figure 4 - 14:

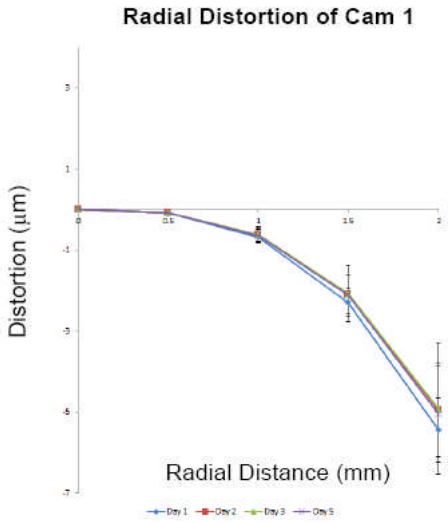


Figure 4 - 14. *Radial distortion of one Logitech Quickcam Pro 4000 camera (Day 5 is approximately 1 year after Day 1, 2 and 3 with different focus setting).*

4.4.2.2.4 Target coordinate accuracy

To compare the accuracy of output target coordinates among data sets from different days, summary statistics of the computed difference between each of the data sets (Table 4 - 9) and the standard values generated from the Nikon D100 photogrammetric survey (Table 4 - 2) are presented in Figure 4 - 15:

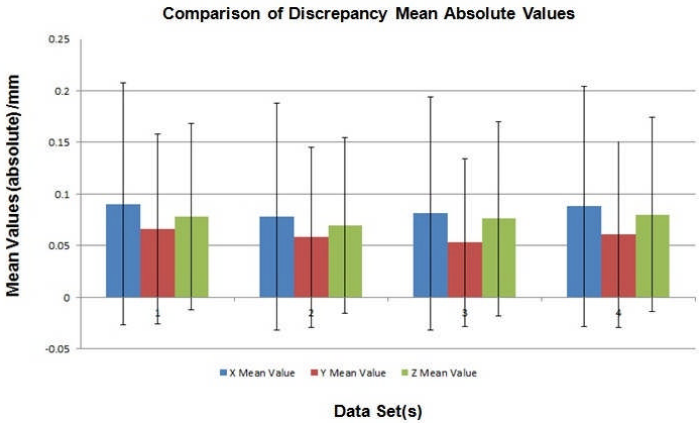


Figure 4 - 15. *Accuracy of output target coordinates (left to right: Day 1, 2, 3, 5).*

4.4.2.3 Discussion

Since this study is looking into how stable the camera interior orientation is over time, data from some typical interior orientation parameters are selected for analysis. For principal point coordinates, the difference in x and y directions was found to be within a range of 0.01 mm (Figure 4 - 11 and Figure 4 - 12). The difference in principal distance for each webcam was within a range of 0.05 mm (Figure 4 - 13). The difference in radial distortion for each webcam was within 0.6 µm at a maximum radius of 2mm (Figure 4 - 14). The magnitude of radial distortion, i.e. ~0.5 pixel at the radial distance of 2mm, has also agreed with the results presented by Page *et al.* (2008) (Section 3.2.2.1). These results suggest that the camera parameters for Pro 4000 webcams can be considered to be stable on a day-to-day basis since 3D target computation checks using the slightly different camera parameters result in almost the same target coordinate accuracies, where all coordinates are within 0.1mm accuracy level (Figure 4 - 15).

4.4.2.4 Summary

The data from this study has provided a reference for future optical topography study, where target coordination can also be carried out with parameters obtained from a calibration conducted several days or several weeks ago (provided that the relative positions among cameras remain unchanged during this time). Optical topography studies require better than 1mm 3D target coordination accuracy, which has been achieved in the presence of small variations in camera parameters on day-to-day basis. This suggests that future optical topography studies with these cameras can be much more flexible and time which would otherwise be spent on camera calibration can be saved.

4.4.3 Logitech C500 Webcams

4.4.3.1 Method

For the newer Logitech C500 webcams which will be used in optical topography fast tracking process, the interest is to see if the camera parameters vary significantly over a longer period and therefore the camera calibrations were carried out on two different dates with about one and half month's interval (08/06/2010 and 20/07/2010 respectively). The imaging geometry and the simultaneous processing is the same as described in Section 4.3.1. Table 4 - 10 gives the details for this study:

Logitech C500 Webcams (Fixed Camera Focus)		**Logitech C500 Webcams (Fixed Camera Focus)**
Calibration data: **08/06/2010**	Same camera used?	Calibration data: **20/07/2010**
Number of webcams: 8		**Number of webcams:** 8
Cam 1	Yes	Cam 1
Cam 2	Yes	Cam 2
Cam 3	Yes	Cam 3
Cam 4	Yes	Cam 4
Cam 5	Yes	Cam 5
Cam 6	Yes	Cam 6
Cam 7	Yes	Cam 7
Cam 8	Yes	Cam 8
Calibration Method A network geometry of 8 webcams		**Calibration Method** A network geometry of 8 webcams

Table 4 - 10. *Experiment details for camera stability study for Logitech C500 webcams.*

As noted in Section 4.2.4 for webcams in this thesis only the basic calibration set (the principal point, the principal distance and 3rd power term (K1) of radial distortion) are considered since only these parameters have a significant effect (Table 4 - 5). As before, calibration of the eight cameras was carried out using a self-calibrating bundle adjustment procedure with a coordinate datum based upon prior measurement of the larger calibration object (Figure 4 - 8) with a Nikon D100 photogrammetric survey.

4.4.3.2 Results

Table 4 - 11 summarises the calibration parameters for two self-calibrating bundle adjustments of all eight cameras conducted on two different days (denoted by A and B):

Program Control Variables	**A: 08/06/2010 B: 20/07/2010**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	**A**: 0.51; **B**: 0.34
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**A: 08/06/2010 B: 20/07/2010**
Number of camera calibration sets	8
Number of target image observations	**A**: 1056; **B**: 2367
Total number of exposures	64
Number of exposures in the network	64
Total number of targets	61
Number of targets in the network	**A**: 55; **B**: 61
Results for the calibration solution	**A: 08/06/2010 B: 20/07/2010**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	**A**: 0.75; **B**: 1.05
Number of observables in the network	**A**: 2135; **B**: 4719
Number of unknowns in the network	**A**: 588; **B**: 614
Number of redundancies in the network	**A**: 1547; **B**: 4105
Mean number of images per target	**A**: 19.1; **B**: 38.3
Target Precision Summary	**A: 08/06/2010 B: 20/07/2010**
Mean precision of target coordinates (μm)	**A**: 64.78; **B**: 31.12
Relative precision for the network (1:X)	**A**: 6000; **B**: 13000

Table 4 - 11. *Summary of calibration parameters for self-calibrating bundle adjustments of eight cameras together carried out on 08/06/2010 (A) and 20/07/2010 (B).*

4.4.3.2.1 Coordinates of principal point

Table 4 - 12 presents the difference in the output principal point offset in the x and the y directions for eight webcams over 42 days:

Principal point offset in x	08/06/2010		20/07/2010	
	x offset (mm)	Uncertainty (mm)	x offset (mm)	Uncertainty (mm)
Cam 1	0.006	0.006	-0.005	0.007
Cam 2	-0.080	0.006	-0.074	0.007
Cam 3	0.099	0.005	0.097	0.007
Cam 4	0.027	0.005	0.026	0.006
Cam 5	0.070	0.006	0.058	0.010
Cam 6	0.020	0.006	0.016	0.011
Cam 7	0.171	0.007	0.176	0.010
Cam 8	0.031	0.006	0.030	0.007

Principal point offset in y	08/06/2010		20/07/2010	
	y offset (mm)	Uncertainty (mm)	y offset (mm)	Uncertainty (mm)
Cam 1	-0.190	0.005	-0.183	0.007
Cam 2	-0.040	0.006	-0.035	0.008
Cam 3	0.038	0.007	-0.030	0.010
Cam 4	0.022	0.005	-0.030	0.006
Cam 5	-0.145	0.004	-0.141	0.004
Cam 6	-0.090	0.004	-0.084	0.004
Cam 7	0.074	0.005	0.080	0.005
Cam 8	0.157	0.008	0.157	0.011

Table 4 - 12. *Principal offset in x (left) and y (right) directions for eight C500 webcams.*

4.4.3.2.2 Principal distance

Figure 4 - 16 presents the difference of principal distances for all eight webcams:

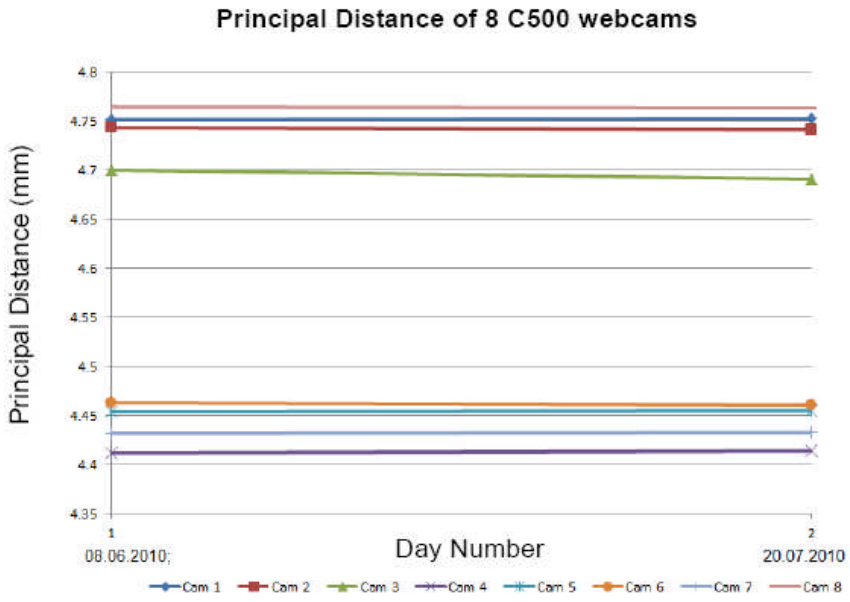


Figure 4 - 16. *Principal distance for eight C500 webcams.*

4.4.3.2.3 Radial distortion

Figure 4 - 17 presents the absolute discrepancies of radial distortions for eight C500 webcams over 42 days:

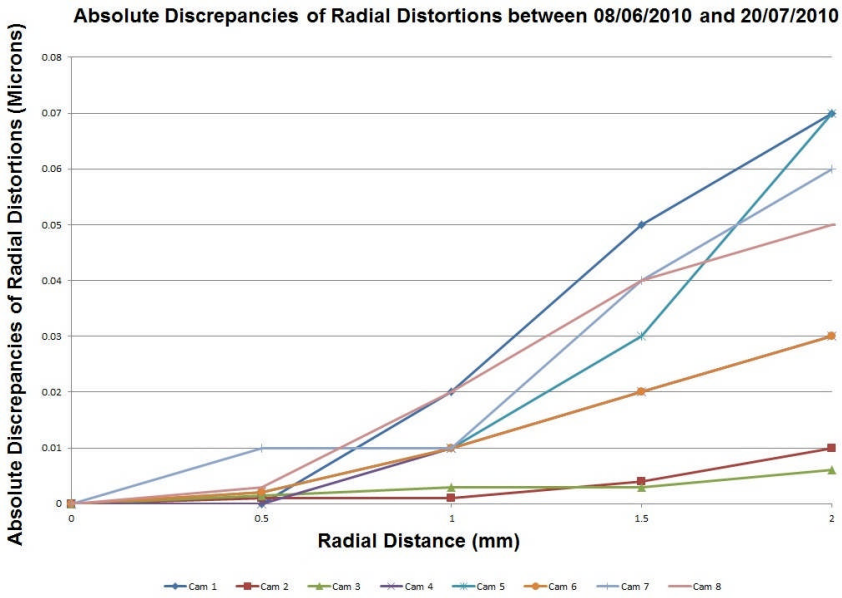


Figure 4 - 17. *Discrepancies of radial distortions of eight C500 webcams.*

4.4.3.2.4 Target coordinate accuracy

To compare the accuracy of output target coordinates among two data sets, summary statistics of the mean absolute discrepancies computed between each of the data sets (Table 4 - 11) and the standard values generated from the Nikon D100 photogrammetric survey (Table 4 - 6) are presented in (1) and (2) from Figure 4 - 18. In terms of reliability of the tracking system, estimated target coordinates of the large calibration object carried out on 21/05/2011 (about 10 months later, (3) from Figure 4 - 18) are compared with Nikon network, where the wheel supporting the cameras has been knocked by different researchers and the entire tracking system has been moved around into different places on the optical table during this period. The camera parameters used for 21/05/2011 were directly extracted from those derived from the self-calibrating bundle adjustment carried out on 20/07/2010, i.e. without any calibration on 21/05/2011.

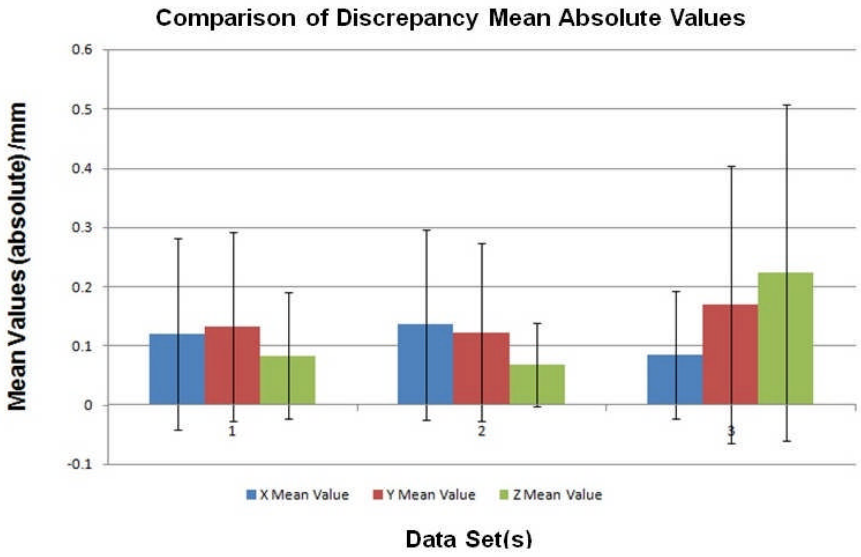


Figure 4 - 18. *Accuracy of output target coordinates*

(left: 08/06/2010; middle: 20/07/2010; right: 21/05/2011).

4.4.3.3 Discussion

This study is looking into how stable the C500 webcam internal camera parameters are over a relative longer period (42 days), as compared with the study carried out for Pro 4000 webcams (3 days). For the principal point, differences in x image coordinates for each webcam are within 0.02mm while differences in y are within 0.01mm (Table 4 - 12) over the 42 day period, where this magnitude of difference is of the same order as those for the Pro 4000 webcams observed over three days (0.01mm range, Figure 4 - 11 and Figure 4 - 12). The difference in principal distance over 42 days for each C500 webcam was within 0.01mm, as compared with the 0.1mm of Pro 4000 webcams over 3 days (Figure 4 - 16). This difference of an order of magnitude could be due to the fact that the C500 webcams do not focus and therefore have a more rigid physical relationship between the lens and the sensors, as compared with the adjustable focus for the Pro 4000 webcams (Table 4 - 1). These results demonstrate that the simple construction is able to offer webcams an advantage over other digital cameras using a combination of springs and hinges in terms of the internal geometry of sensor (Section 3.3.1).

However, interestingly principal distances can be separated into two clear groups with C500 webcams 1, 2, 3, 8 roughly 0.35mm greater than Cam 4, 5, 6, 7. However, Cam 1 to Cam 8 are all the same webcam models (Logitech C500) with fixed focus. The only

difference between Cam 1, 2, 3, 8 and Cam 4, 5, 6, 7 is that Cam 1, 2, 3, 8 are plugged into USB 2.0 slots, while Cam 4, 5, 6, 7 are plugged into PCI express USB 3.0 ports (Figure 4 - 23). To further verify this difference on principal distance, two individual calibrations were carried out independently where the same webcam was attached to each port in turn to eliminate any possible camera variations. Table 4 - 13 summarises the parameters from these two self-calibrating bundle adjustments:

Program Control Variables	**A: USB 2.0 B: USB 3.0**
Network datum definition type	Generalised internal constraints
Maximum iterations for a solution	10
Default target image precision by camera (μm)	**A**: 1.70; **B**: 1.42
Minimum images for a network target	4
Rejection criterion for image errors	2.5μm
Input Summary	**A: USB 2.0 B: USB 3.0**
Number of camera calibration sets	1
Number of target image observations	**A**: 282; **B**: 322
Total number of exposures	10
Number of exposures in the network	10
Total number of targets	61
Number of targets in the network	**A**: 42; **B**: 44
Results for the calibration solution	**A: USB 2.0 B: USB 3.0**
Unit weight estimate (sigma zero)	1
RMS image residual (μm)	**A**: 1.25; **B**: 1.16
Number of observables in the network	**A**: 507; **B**: 605
Number of unknowns in the network	**A**: 195; **B**: 201
Number of redundancies in the network	**A**: 312; **B**: 404
Mean number of images per target	**A**: 5.9; **B**: 6.8
Target Precision Summary	**A: USB 2.0 B: USB 3.0**
Mean precision of target coordinates (μm)	**A**: 139.75; **B**: 113.74
Relative precision for the network (1:X)	**A**: 2000; **B**: 3000

Table 4 - 13. *Summary of calibration parameters for two individual calibrations with the same webcam.*

Figure 4 - 19 shows the principal distances from these two calibrations:

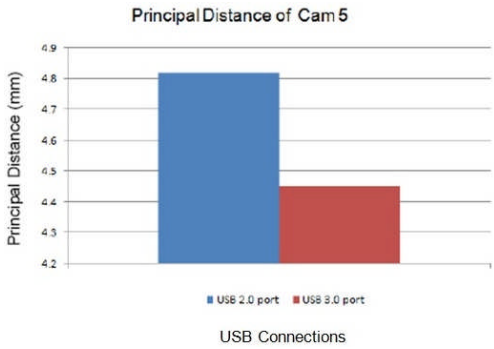


Figure 4 - 19. *Principal distance of a C500 webcam connected to USB 2.0 and USB 3.0 ports.*

Similar to Figure 4 - 16, there exists a difference of about 0.35mm of the principal distances even with the same camera, which has confirmed that the variations of principal distance between different connection ports are not related to the camera internal variations. The reason accounting for this change in principal distance goes to an image scaling occurred in the capture system which is linked to the change of image quality and automatic camera driver parameter settings managed by the DirectShow API. Furthermore comparison was made between the images captured from a webcam connected to PCI express board and a webcam connected to a normal USB 2.0 board (Figure 4 - 20). Although both images have the same resolutions (1280×1024), the image from a C500 webcam connected to a normal USB slot contains less details than the image when it is connected to PCI USB 3.0 slot:

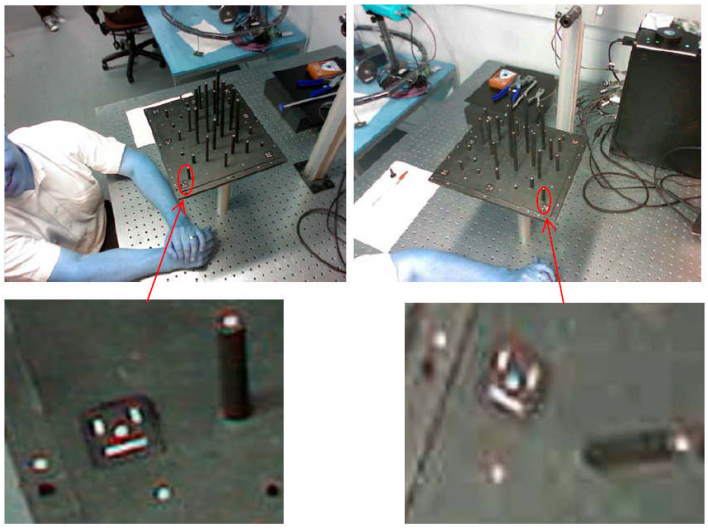


Figure 4 - 20. *Comparison of images extracted from image sequence.*

Upper left and upper right: An image with hardware resolution of 1280×1024;

Lower left and lower right: A portion of the image with 3 times enlargement (3840×3072).

Moreover, every time a webcam was connected to a USB 2.0 port in this single desktop, a warning message was shown on screen by the camera software as follows:

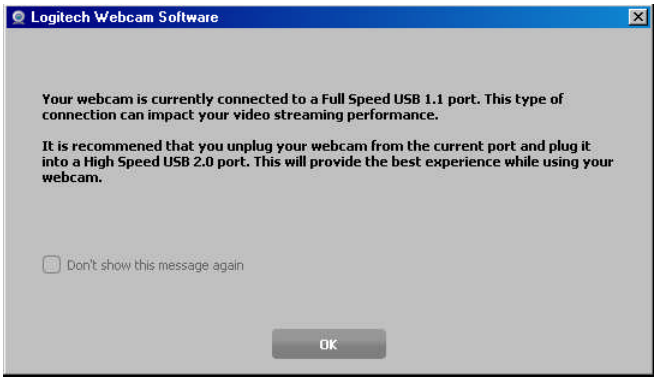
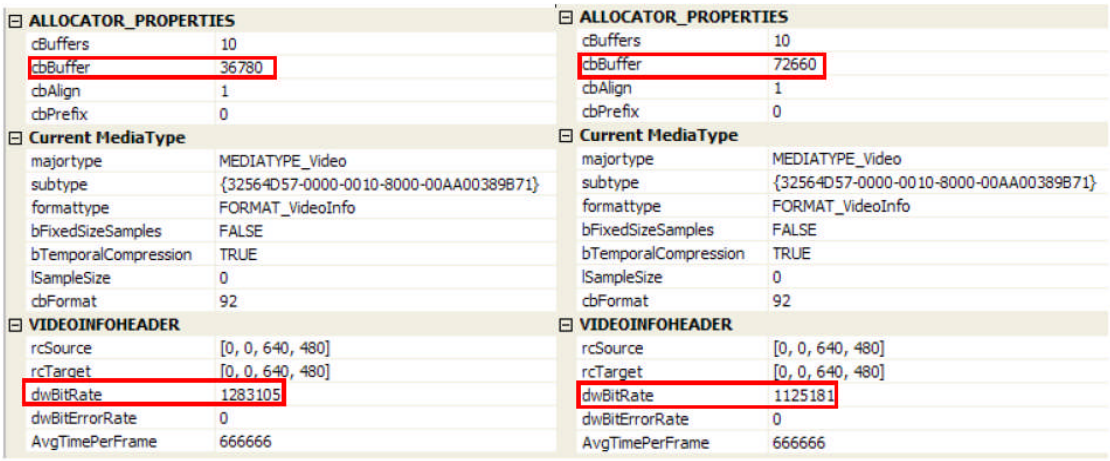


Figure 4 - 21. *Warning message when webcam is connected to a USB 2.0 port.*

This is because Logitech C500 webcam is designed with USB 2.0 interface, which means its maximum video transfer rate is not more than 480MB/s. Therefore even if it is connected to a USB 3.0 port, the camera driver still recognises it as a USB 2.0 port. When the webcam driver for C500 models detects different webcams being connected to USB 3.0 and USB 2.0 ports, due to different internal universal serial bus controllers between these two types of connections, the driver recognises the controller for those webcams connected to USB 3.0 ports as its native support connection type (USB 2.0) while recognises those USB 2.0 controllers as a lower level of connection it supports, i.e. USB 1.1. And therefore the above warning message shows whenever a C500 webcam is connected to a USB 2.0 port in this desktop. The C500 webcam driver recognises the connection type in such a way that the above warning message shows even if there is only one C500 webcam plugged into the desktop through any USB 2.0 port.

When an application tries to connect to the C500 webcams through DirectShow (Appendix 1), KsProxy, a wrapper filter provided by DirectShow to allow webcams to join DirectShow, is instantiated to send query to camera driver, which returns two different property sets which the camera driver support: one for USB 2.0 and one for USB 1.1. KsProxy is then configured to expose COM interfaces that correspond to these two property sets, which are then translated to camera driver in order to start video streaming from webcams. In this case the property sets received from the camera driver which controls Cam 1, 2, 3, 8 correspond to USB 1.1 while those controlling Cam 4, 5, 6, 7 correspond to USB 2.0.



ALLOCATOR_PROPERTIES	
cBuffers	10
cbBuffer	36780
cbAlign	1
cbPrefix	0
Current MediaType	
majortype	MEDIATYPE_Video
subtype	{32564D57-0000-0010-8000-00AA00389B71}
formattype	FORMAT_VideoInfo
bFixedSizeSamples	FALSE
bTemporalCompression	TRUE
lSampleSize	0
cbFormat	92
VIDEOINFOHEADER	
rcSource	[0, 0, 640, 480]
rcTarget	[0, 0, 640, 480]
dwBitRate	1283105
dwBitErrorRate	0
AvgTimePerFrame	666666

ALLOCATOR_PROPERTIES	
cBuffers	10
cbBuffer	72660
cbAlign	1
cbPrefix	0
Current MediaType	
majortype	MEDIATYPE_Video
subtype	{32564D57-0000-0010-8000-00AA00389B71}
formattype	FORMAT_VideoInfo
bFixedSizeSamples	FALSE
bTemporalCompression	TRUE
lSampleSize	0
cbFormat	92
VIDEOINFOHEADER	
rcSource	[0, 0, 640, 480]
rcTarget	[0, 0, 640, 480]
dwBitRate	1125181
dwBitErrorRate	0
AvgTimePerFrame	666666

Figure 4 - 22. *Difference of buffer sizes delivered from a C500 webcam connected to USB 2.0 port (left) and PCI express USB 3.0 port (right).*

While the data rates (dwBitRate which specifies the video stream's approximate data rate in bits per second [dwBitRate definition, 2011]) between USB 2.0 and USB 1.1 shows little difference, the size of each buffer (cbBuffer) set through USB 2.0 is nearly twice as that as in the USB 1.1 case (Figure 4 - 22). When video streaming reaches the video card, streamed video data is transferred from the output pin (the connection point between filters, Appendix 1) of one filter to the input pin of the next filter (e.g. image grabber filter). The two pins must agree on the number and the size of the buffers created by the allocator. Figure 4 - 22 shows the same number of buffers (i.e. cBuffers=10) created but completely different buffer sizes, while the buffer size from webcam connected to USB 3.0 port is nearly as twice as that connected to USB 2.0 port, which implies:

1. DirectShow assigns the pin from the raw video to provide the allocator, where the pin fills each buffer for USB 1.1 with nearly half of data as compared with that for USB 2.0, which account for the difference of image details shown in Figure 4 - 20;
2. The property set from camera driver has a direct impact on the determination of buffer size when two pins are connected. That is, the driver returns property sets for USB 1.1 connection for Cam 1, 2, 3, 8 while it returns property sets for USB 2.0 connection for Cam 4, 5, 6, 7, in which this difference between two property sets directly determine the buffer size to be retrieved by next filter.

Due to the differences in buffer sizes, it is very likely that the image grabber samples images with two slightly different scales (about 4.45/4.8=93%), which accounts for the

0.35mm change in principal distance (Figure 4 - 19). This provides an important conclusion that for a multi-camera system, it is critical to calibrate the cameras as a system but not as individual sensors.

For the desktop computer (Dell Precision 490) used in this research, each USB 2.0 board is able to support only one C500 webcam video stream and each PCI express USB 3.0 board (2 port USB 3.0 hub to PCI-E express card adaptor) is able to support two C500 webcam video streams. Therefore to set up an 8-camera tracking system, Cam 1, 2, and 3 are individually connected to three USB 2.0 boards and Cam 8 is connected to a USB port which comes with the desktop computer itself. On the other hand, each PCI express board contains 2 USB ports and tests have shown that 2 webcams can be used with 1 single PCI express board. As shown in Figure 4 - 23, Cam 4 and 5 were connected to a PCI express board while Cam 6 and 7 were connected to the other PCI express board. It is worth to note that each of the boards that had been added has multiple USB ports but common interface hardware between all ports through which the data must pass. This configuration has resulted in the four cameras connected to USB 2.0 ports not able to deliver the image quality as they should (left images of Figure 4 - 20) but imagery filled with nearly half of the original details (right images of Figure 4 - 20).

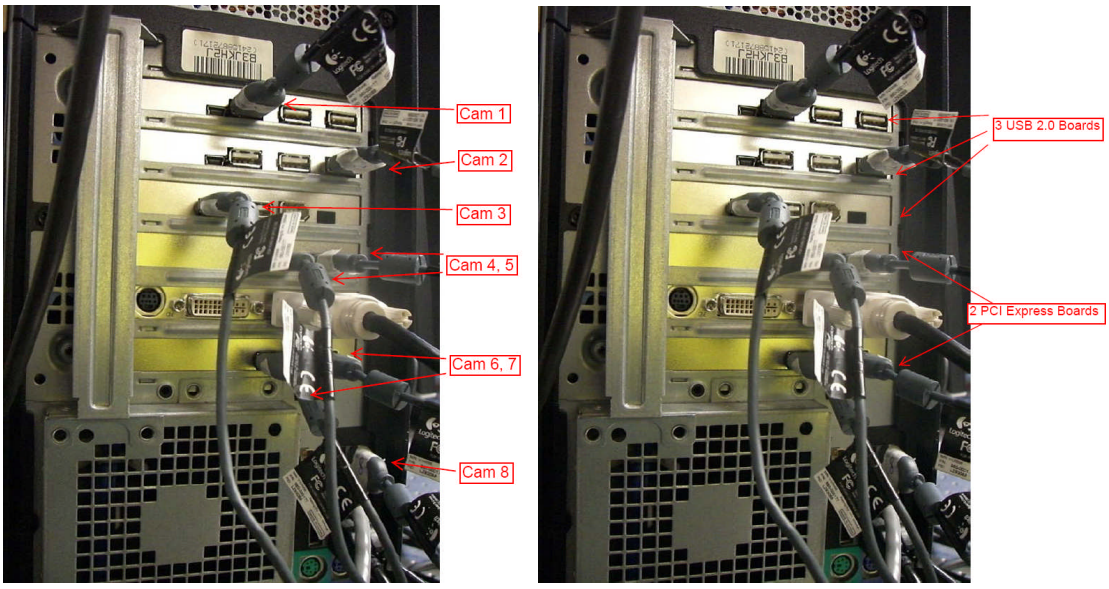


Figure 4 - 23. *8-USB Connection to Dell Precision 490 (Intel Quad Core 2.0GHz, 4GB Ram).*

The difference in radial distortion over 42 days for each C500 webcam was within 0.1µm, as compared with 0.6 µm at a radius of 2mm for the Pro 4000 models (Figure 4 - 14). However, Figure 4 - 17 suggests that the lens distortion varies between the different C500 webcams (from -5µm to -30µm at a radius of 2mm). This is despite the fact that Logitech officially declared that C500 webcams use a precision engineered glass lens. Despite this, lens distortion for each individual C500 webcam (Figure 4 - 17) can be corrected by the self-calibrating bundle adjustment.

Output 3D target coordinates generated from the self-calibrating bundle adjustments from 08/06/2010 and 20/07/2010 respectively are presented in (1) and (2) of Figure 4 - 18. Once again results suggest that the small variation in camera parameters between the two data sets have no effect on the target coordination accuracy of the tracking system. While output target coordinates of data set (3) were generated ten months after the self-calibrating bundle adjustment (20/07/2010), accuracy levels within 0.25mm for all coordinates was still achieved. This magnitude is still well below the accuracy level required for optical topography tracking. The slightly larger discrepancy for 21/05/2011 is caused by two factors. First, webcams and the wheel have been knocked during their use by different researchers over a ten month period of time. Secondly, there is a possibility that the physical shape of the steel rod mounting these webcams had slightly changed due to a large difference of temperature across summer and winter seasons so that there have been a slight difference of bending angles. All these could have resulted in a slight change of the relative positions among the eight webcams which accounts for the small decrease in accuracy level ((3) in Figure 4 - 18). Nonetheless, these results have demonstrated that provided with good internal sensor geometry of webcams (Section 4.4.3.3), maximising their exterior orientation stability by re-engineering the mechanical mounting has enabled webcams to deliver consistent 3D coordination accuracy (Section 3.3.1).

4.4.3.4 Summary

The main obstacle for bundle adjustment with these low cost webcams is the manual measurement of target images since the poor image quality delivered by these webcams has currently made a fully automated target image measurement process difficult with existing target image measurement software. In particular target image quality for coded target recognition was found to be insufficient and that an appropriate set of coded targets suited to webcam work is required for full automation. For a self-calibrating bundle adjustment where eight images are taken for each webcam and there are a total of eight webcam, about one minute is needed to measure target images for each individual image and therefore about one hour is needed to generate sufficient target images for the bundle adjustment. However, the automation of the calibration process for webcams would be possible given merger of the target detection techniques developed later on in this thesis (Section 4.5.4). The continuous enhancement of image quality in webcam technology will also have an influence.

In terms of reliability of the tracking system for optical topography applications, results have demonstrated that the camera parameters are stable and the system is reliable over a long period of time to the required 1mm level of accuracy. As a result the need to repeat self-calibrating bundle adjustments for each use of the system can be eliminated over a long period of time. This reliability has made the multi-camera tracking system almost immediately ready for optical topography application. The camera parameters used for the phantom and human studies performed on 03/08/2010 and 04/08/2010 (Section 5.2 and 5.3) were directly extracted from those obtained from the self-calibrating bundle adjustment on 20/07/2010.

4.5 Fast Tracking System

4.5.1 Introduction

One of the advantages of optical topography over conventional tomographic functional mapping methods such as fMRI is that slight movement is allowed during imaging. Combined with the highly portable imaging equipment and the non-invasive imaging technique, more flexible functional studies can be designed for optical topography (Section 2.1). However, although the optical sensing pad is usually firmly attached to subject's head during optical topography study, relative movement between pad and head can still happen with the result that study data are invalidated. For example, in a topography study in the Baby Laboratory at Birkbeck College of London, we found that sometimes the baby touches their mother who usually sits behind the baby. This could cause the dislocation of the optical sensing pad and therefore the functional study has to be terminated. This limitation has leaded to the fact that many repetitive topography studies have to be conducted (without knowing if dislocation actually happens) before useful data can be extracted for analysis.

(a) (b)

Figure 4 - 24. *Optical sensing pad used in the Baby Lab, Birkbeck College of London. (a). Inside look of the optical sensing pad; (b) The sensing pad is attached on the baby's head.*

Regarding this issue, if the operator can be informed if differential movement is detected during the topography study, the operator can either decide to abort or continue the study. Provided that the following requirements are fulfilled, optical topography experiment time can be saved whilst obtaining useful data:

1. 2D and 3D target positions (coordinates) on the head and optical sensing pad should be computed in nearly real time. The relative positions on the pad and head can then be compared with the previous positions to see if there is any differential change;
2. Coordinate update should be quick enough to suit the fast monitoring purpose.

Before going into the details of the system design, some factors to determine the characteristics of a multi-camera tracking system suited to clinical conditions need to be considered first:

1. For portability of the system in future use, the materials used to build the system need to be light but rigid as well, leaving aluminum as the best choice;
2. Since the camera parameters and the orientations for target tracking need to be extracted from the associated self-calibrating bundle adjustment, the accuracy of target coordination heavily relies on fixed relative positions among these cameras during the tracking. Therefore webcams need to be firmly attached onto the system;
3. A convergent multi-camera network is required to deliver strong network geometry in terms of target coordination (Section 3.2.3). This implies that an object with circle shape (wheel) should be used to mount the webcams;
4. Given the wheel based camera mounting used to deliver a multi-camera convergent geometry, to prevent occlusion of targets by the wheel itself, all the webcams need to be situated underneath the wheel;
5. The height of the wheel needs to be adjustable in favour of positions of test subjects;
6. The wheel size should allow enough length for the USB cable of every mounted webcam to connect to the USB port of the desktop.

In the following sections key components of the tracking system are described.

4.5.2 Design of the tracking system

Taking advantage of the clip underneath the C500 webcam, shells of C500 webcams were opened to replace the original clip with a steel rod so that webcams can be rigidly fixed in positions around the edge of a bicycle wheel ring, as shown in Figure 4 - 25. The rigidity of the webcam mounting on the wheel ring is critical since the relative positions among webcams should theoretically be unchanged after camera calibration to ensure

consistent camera orientations between calibration and tracking. The supporting rods were further bent outwards to increase the overlapping area being viewed by all eight webcams due to the limiting size of the wheel ring (622mm internal diameter). A flexible screw was designed to keep each camera's position fixed after their relative positions had been adjusted to provide a convergent geometry in favour of 3D target coordination.

Figure 4 - 25. *Workflow of the initial setup of the tracking system*.

The fast tracking system is composed of eight Logitech C500 webcams (£44 each) mounted upside down at the edge of an alloy bicycle wheel rim (£40). The wheel can be adjusted on a supporting stand for height to achieve a suitable tracking distance between webcams and subject. The cameras are positioned so that four webcams are close to each other to ensure the most important targets around the sensing pad can be tracked. The remaining four webcams are positioned more sparsely around the ring to ensure less important targets facing these directions can be intersected from at least two images (Figure 4 - 26). For example, in order to localise brain activation area, target locations on an optical sensing pad positioned on the upper left area of a head in order to study left temporal brain activation are the most important. Conversely the system can be adjusted, or simply rotated so that target locations on an optical sensing pad at the back of the head to study human visual cortex activation can be observed.

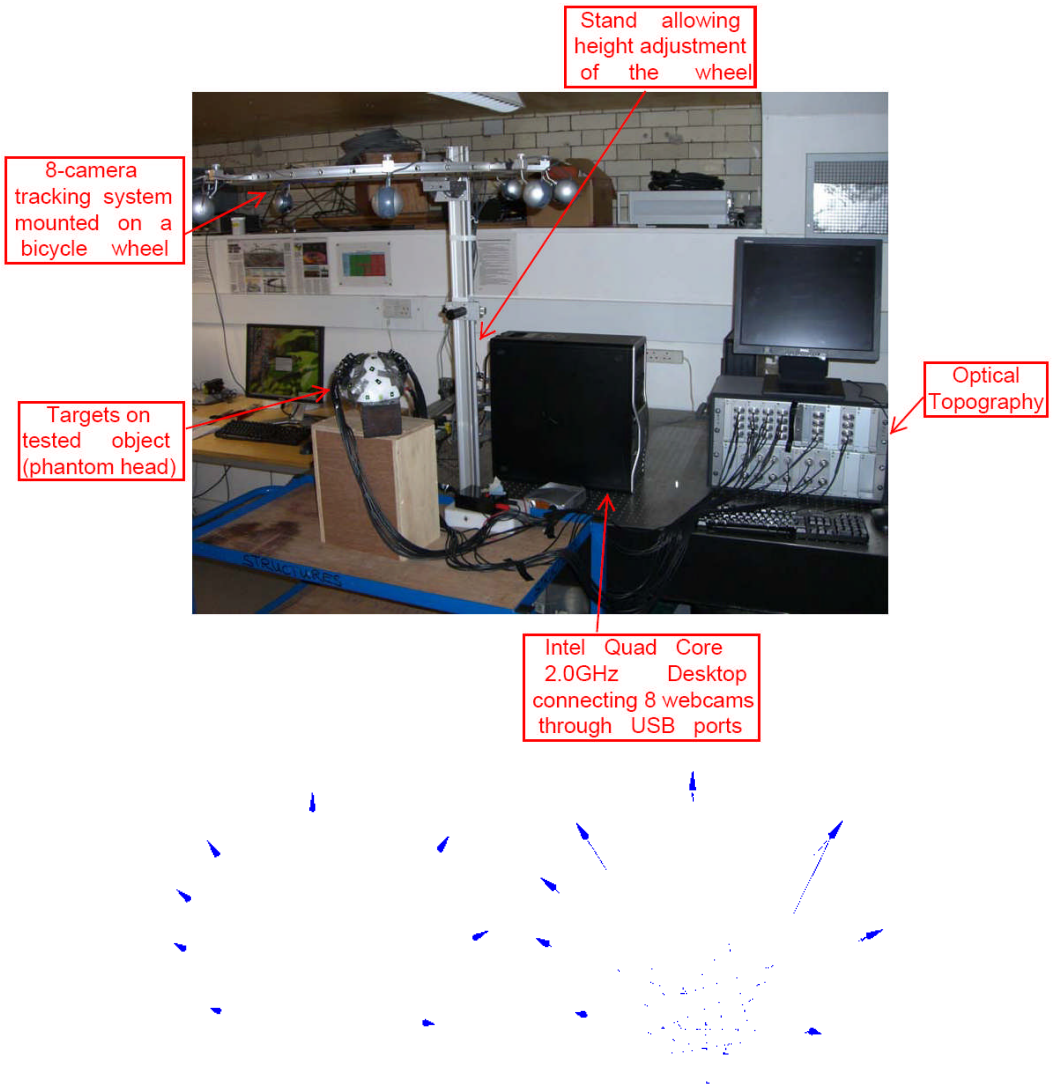


Figure 4 - 26. *Experimental setup and convergent geometry of the 8-camera tracking system.*

Given the convenience of the wheel rim, but its limited diameter, the maximum bending range of webcams on the wheel and the approximate size of human head, a distance of 0.6 to 0.8m between webcams and targets was found to provide the maximum number of targets (situated over human scalp) viewed by all the webcams (Section 4.3.1). During these studies image quality was found to be more consistent when the white balance is switched off with software control to ensure the same settings across all webcams. The system design (Figure 4 - 26) is a prototype with a total cost of less than £400. The design also offers the possibility to mount the ring of cameras on a stand fixed to the back of a clinic chair in future so that tracking can be done anywhere in clinic, at the patient's bedside or inside a testing room.

4.5.3 Synchronization of webcams

The goal of the fast tracking system is to perform 3D target coordination as fast as possible to allow tracking of any movement of the human head during the study. Any time delay between image frames captured by different webcams could directly cause an inaccuracy in the 3D coordinates being intersected if the subject's movement velocity becomes comparable to the time delay.

The images captured by USB webcams are executed in series and there exists a time delay between different frames. In order to minimize the time gap, the frame capture process was implemented independently of camera initialization and image processing. This ensures that image capture starts only after all connected webcams have been initialized and set up and the image processing starts only after all connected webcams have grabbed new image frames and saved them into a buffer.

A time counter was used to record and calculate the time difference between successive image frames being captured by different webcams. Test and retest results have shown that the average time delay between image frames captured by different webcams in one set (all images in one set are used for 3D coordination) is between 0.005s and 0.007s. Provided that the movement of the test subject is small compared to those doing sports such as running, the time synchronization issue can be ignored in the study.

Methods do exist to synchronize different cameras, but these are not generally available on webcams. In hardware, genlock can be used to connect different cameras, whilst in software, parallel computing using the parallel pipelines of a computer's graphics card as a trigger might also be implemented. Since the research focus is not about the synchronization and the time difference in order of millisecond is considered to be small enough to be ignored in this application, and will not be discussed in the following sections. However, research to completely synchronize webcams is an interesting topic for the future.

4.5.4 Target detection

After the image capture environment has been established, useful information needs to be extracted from the complex scene for 3D target coordination. In order to achieve this, two methods have been developed. The first is a target detection method which can be applied in a broad sense whilst the second is specially designed to tackle target detection under dim environments with insufficient lighting such as that often found during optical topography studies where normal lighting could seriously diminish the infrared signal emitted from the optical fibres. Details regarding the first method have been given in Section 3.4.4, the second method, which is built on top of the first, is described in the following section with example images.

4.5.4.1 Background

Tests [Wong *et al.*, 2009] demonstrated that the edge image from a Canny edge detector [Canny, 1986] as a mask is capable of defining the target image boundary for a subsequent centroiding process. The Canny method was found to work well when targets of interest constitute the majority of the image grey values of interest within an image (e.g. 90%). However the method is quite sensitive to background noise if the background scene is also included in the image (Figure 4 - 27).

The target detection method described in Section 3.4.4 applied a morphological operation prior to Canny edge detection in order to filter background features. However, when the earlier study in Section 3.4.4 was extended to more complex environments, although the erosion operation was able to eliminate some image noise, objects which

are similarly sized to target images and form an obvious contrast against the background passed the criteria set by the Canny edge detector and were recognised as targets giving false positives [Wong *et al.*, 2010].

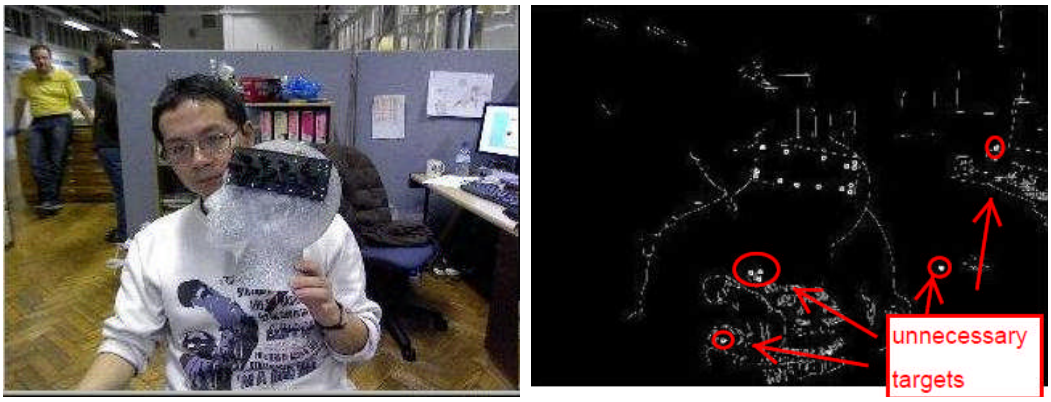


Figure 4 - 27. *Target detection in simulated clinic-environment [Wong et al., 2010].*

In more extreme cases, such as Figure 4 - 28 where a human head model with an OT optode pad was imaged in a clinic-simulated laboratory environment, the contrast formed by the depth difference between the dark holes in the brighter optical table surface allowed the Canny edge detector to extract the edges of these holes. Due to the similar size and shape of the holes and the target dots, the edge information further passed the blob detection algorithm (Flowchart 3 - 2), even though it incorporated circularity criteria for target shape and size. Finally these holes were all recognised as targets and their individual centroids were computed. In the next stage the coordinates of both real and incorrectly recognised targets were input into our subsequent 3D correspondence computation process and must be filtered in 3D space in a computationally more expensive process [Atkinson, 1996; Luhmann *et al.*, 2006].

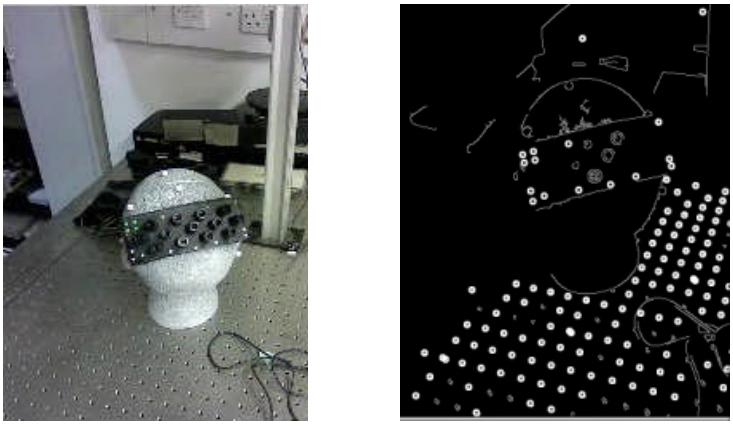


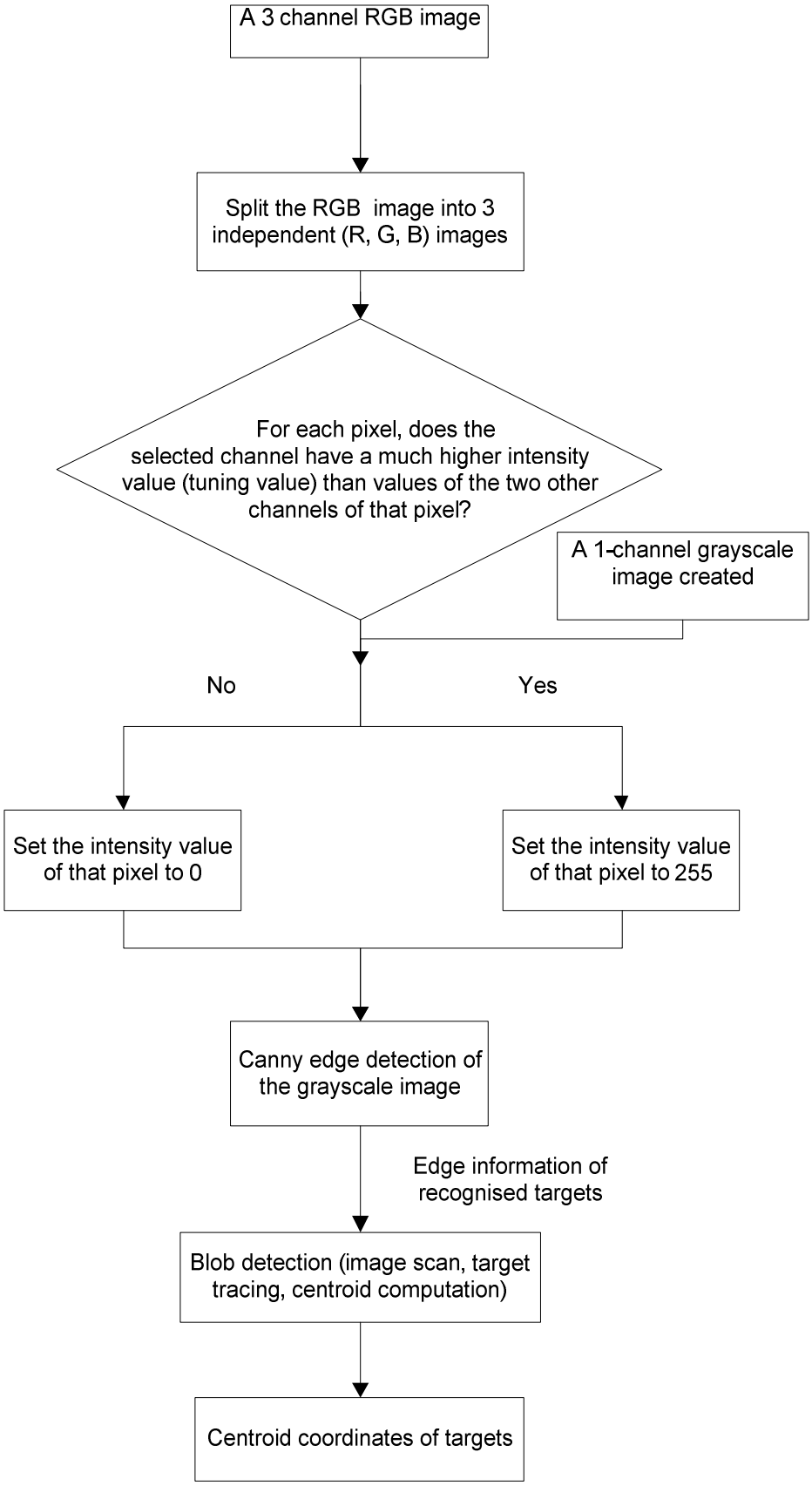
Figure 4 - 28. *Holes on an optical table were false recognised as targets [Wong et al., 2010].*

Incorrectly recognised targets are an obstacle to our goal of localising OT optodes and incorrect measurements must be rejected before 3D correspondence. This elimination simplifies the correspondence computation as each measured target needs to potentially be compared with every measured target in all the other images. Minimising the required correspondence search space is very important for fast processing (Section 3.5.3).

4.5.4.2 Method

Regarding the drawbacks of the target detection method, further steps are necessary on top of the morphological filter and Canny edge detector so that new criteria can be used to distinguish between the real and unwanted targets. However any new steps need to be as computationally simple and efficient as possible in order to minimise the time delay for rapid processing.

As a first step, coloured targets that will be predominantly imaged by a single colour channel can provide a first pass filter for target detection. One impact is that less clinical lighting control is required than for multi-channel imaging as specular patches which might be incorrectly identified as targets are usually imaged in all three colour channels and can be quickly filtered out. This is considered to be crucial for in-clinic work since optical topography studies are usually carried out in locations such as at the patient bedside where lighting conditions are wide ranging. In terms of channel selection, green is allocated by the Bayer filter with twice as many detectors as the other channels (ratio 1:2:1 for R, G, B respectively) for human visual purpose (Section 3.3.1.2). The missing pixels from each channel are then interpolated from neighbourhood pixel(s) in the demosaicing process (Section 3.3.1.4) to build up the complete image. To minimize the possible incorrect image information introduced by interpolation, the colour channel with the least interpolation should be selected, which is green. Flowchart 4 - 1 shows the logic of the colour solution.



Flowchart 4 - 1. *Colour solution logic [Wong et al., 2010].*

4.5.5 Results

Although the default setting for the tracking system is able to effectively detect targets in most environments, several parameters need to be tuned to allow target detection suited to a specific testing environment:

	Hardware resolution	Canny upper threshold	Canny lower threshold	Tuning threshold
Normal targets	✓	✓	✓	
Coloured tuning targets	✓	✓	✓	✓

Table 4 - 14. *Adjustable parameters in two target detection methods.*

Discussion of the results found in fast tracking systems are divided into two sections. In the first section results of target detection for normal targets are presented while in the second section, the comparison of target detection results between the normal-target method and the tuning-target method is given. Moreover, since the tuning-target method was used for optical topography study, more results regarding this specific method will also be given in the second section as a reference for future study.

4.5.5.1 Target detection with normal targets

In this section target detection results after applying the Canny edge detector are given. Results reveal the drawbacks regarding this edge detection technique. Since the Canny edge detector is very sensitive to background scene, adjustment of its threshold is the key to achieve an acceptable separation between foreground and background. Figure 4 - 29 shows some typical results with different threshold settings for the Canny edge detector used in this study:

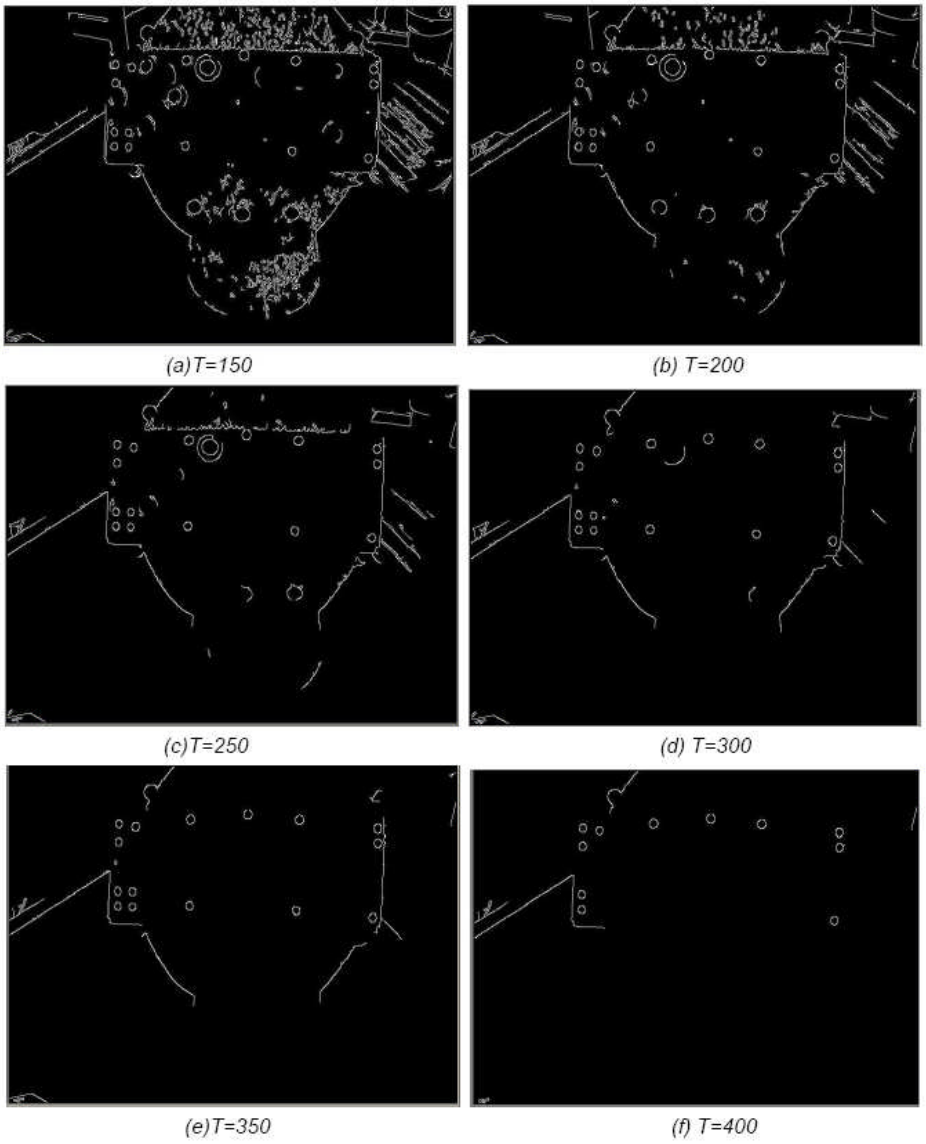
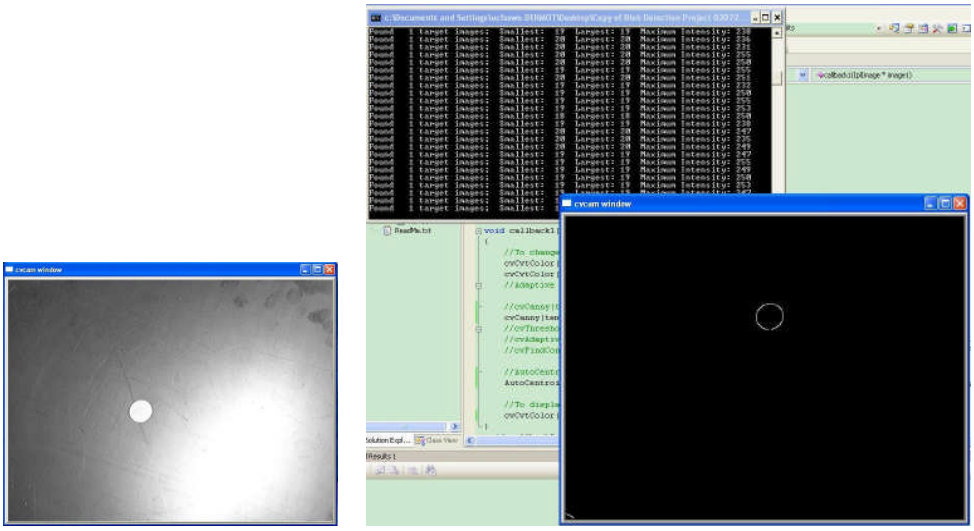


Figure 4 - 29. *Results of Canny edge detection with different thresholds on the targets located on an optical topography pad positioned on a polystyrene head.*

Although the Canny edge detector is able to isolate required targets from a complex background as shown above, under- or over- threshold settings could easily lead to undesired target detection results. For example, from Figure 4 - 29 it can be seen that the first three images with thresholds of less than 300 have left unwanted scene information even after filtering. As a result unwanted targets are highly likely to be found. The smaller the threshold value, the more unwanted targets detected. However, all 15 targets located on the sensing pad were detected and meanwhile most of the background information was filtered when the threshold value is 300, which was considered to be the suitable detection threshold value, as shown in the fourth image of Figure 4 - 29. Any threshold values higher than 300 were able to better filter the

background/unnecessary information, but the cost is the filtering of edges of the real targets at the same time. As shown in the last image in Figure 4 - 29, edges of the two targets on the bottom of the pad (with less contrast compared with other targets) were completely filtered out when the threshold value reached 400. Therefore care needs to be taken when choosing the appropriate threshold value with regard to a particular testing environment. In practice the threshold adjustment process typically took less than five minutes in order to reach the suitable setting and therefore applying the Canny edge detector rather than the conventional threshold method is considered to be beneficial.

The larger the threshold value, i.e. the difference between threshold 1 and threshold 2 in the Canny edge detector, the more likely edge disconnection will occur (Section 3.4.4.2). However, the cost of reducing the difference between the two thresholds is an increase of unnecessary background information including noise. Introducing the dilation and erosion (Section 3.4.4.3) not only reduced the noise from the shrinking effect of erosion, but also connected gaps (broken edges) between disconnected edges to enhance the shape of the objects of interest. A comparison between results of the conventional Canny edge detector and the same edge detector but with dilation and erosion added is shown in Figure 4 - 31. Results are illustrated after introducing a blob detection algorithm on the edge detection result (Flowchart 3 - 2), as this enables a better comparison of the difference.



(a) (b)

Figure 4 - 30. *Edge of a blob detected by the Canny edge detection.*

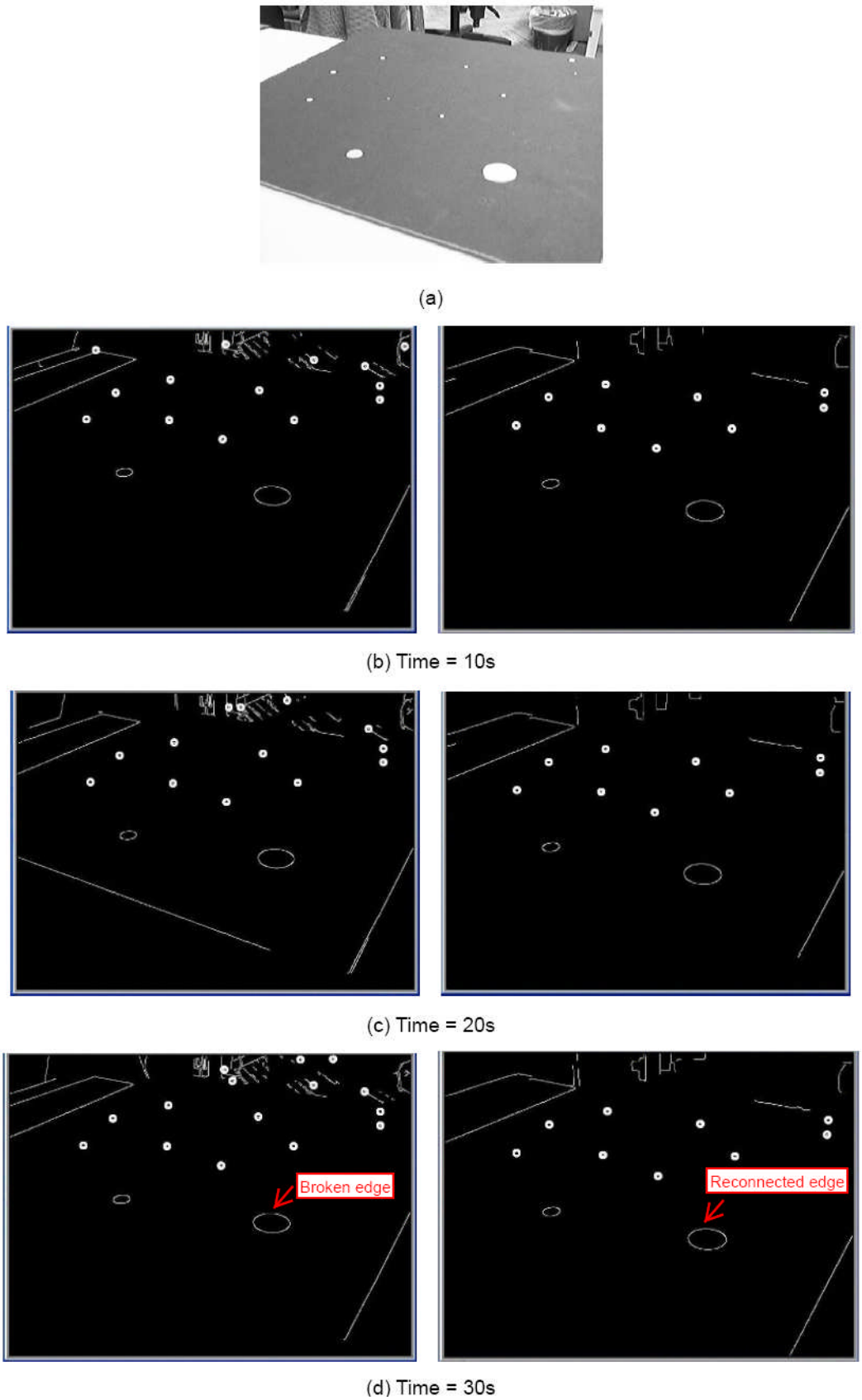


Figure 4 - 31. *Comparison of Canny only (left) and morphological operations + Canny (Right).*

In Figure 4 - 30 and Figure 4 - 31, the blob area limits were set between 10 and 80 pixels. Canny thresholds one and two were set to be 200 and 400 respectively. Figure 4 - 30(a) shows a frame captured by the webcam and Figure 4 - 30(b) is a screenshot of the detection result. The result shows that one target was consistently found in every frame. Figure 4 - 31(a) also shows a frame captured by the webcam. For Figure 4 - 31(b)-(d), the left hand side shows screenshots of the blob detection results after the Canny operation whilst the right hand side shows the blob detection results after the morphological operations (dilation and erosion) followed by the Canny edge operation. We can observe image noise, which was wrongly detected as target blobs (white dots at the top of the image) on the left hand side, had been successfully filtered out so that the right hand side correctly recognized a total of nine blobs. The two blobs in front were not recognized as blobs since their areas are over the limits, i.e. 10 to 80 pixels.

In order to be a potential target, one of the criteria set by the blob detection algorithm is that the blob needs to be in a complete shape. Any incomplete shape such as a circle with broken edge(s) could terminate the edge tracing process (Section 3.4.4.4) and will not be recognised as a target. The morphological operators (Section 3.4.4.3) can effectively compensate the high noise sensitivity of the Canny detector by promoting the reconnection of broken edges (Figure 4 - 31). Tests have shown that broken edges are very common after an edge detection process, for example the biggest blob had a disconnected edge after Canny (second last image, Figure 4 - 31). Therefore morphological operation is considered as necessary as a companion to the Canny detector to reconnect those broken edges, as shown in the last image that the disconnected edge has been reconnected.

4.5.5.2 Target detection with coloured tuning targets

In Section 4.5.4.1 the drawbacks of the normal-target detection method, which was discovered by accident when the polystyrene head was tested on an optical table with many circular threaded holes, were discussed. Regarding the drawbacks, the tuning-target method was developed (Section 4.5.4.2) in order to better isolate targets to be detected from complex background, even if objects with similar size forming similarly sharp contrast with background exist within the view of webcams. In the following comparison of the two methods with the same object and under the same testing environment will first be given, followed by the appropriate setting for this method.

Comparison of normal-target and colour tuning-target methods

For better comparison of the two methods, 7 out of 15 targets on an optical sensing pad were painted to green in colour by a common highlighter, as shown in Figure 4 - 32. Since the tuning (green) targets also formed a sharp contrast with the black background of the optical sensing pad, they should theoretically be detected as targets by Canny edge detector in the same way as white targets. However, only the coloured tuning targets should be detected as targets when the tuning-target method is applied, provided the green component of the coloured tuning targets is higher than the colour comparison threshold that has been set.

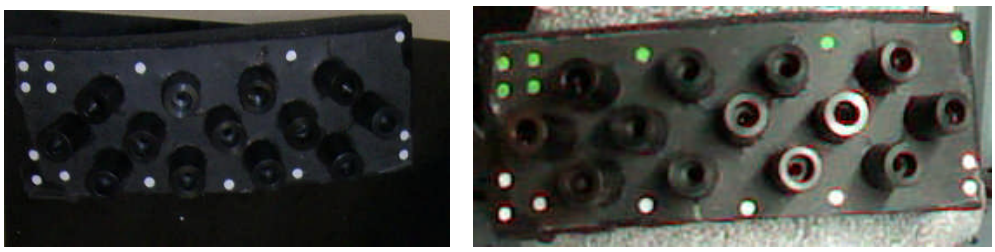


Figure 4 - 32. *White targets (left) and 7 coloured tuning targets (right) on a sensing pad.*

Figure 4 - 33 compares the difference between the two methods in terms of target detection. For target detection with normal targets, background edges normally exist (1st row and 3rd row of Figure 4 - 33) and therefore the chance to recognise unnecessary targets is high since the target detection for normal targets relied on the object size as well as the contrast formed between the object and its background. It can be seen that

with the normal targets there are seven and five unnecessary targets being recognised at image resolutions of 640×480 (1st row of Figure 4 - 33) and 1280×1024 (3rd row of Figure 4 - 33), respectively. Compared with the normal target detection, the colour tuning-target method produced a much cleaner image prior to target detection because the criteria for an object to be recognised as a target consist of three information sources: the particular colour (green in this case), the target image diameter and area and; the sharp contrast formed between the object and its background. As can be seen from Figure 4 - 33, all eight coloured tuning targets were recognised by the C500 webcams at both resolutions because the background information could be completely filtered out before target detection algorithm started.

The advantage of using a higher hardware resolution is again illustrated by comparing the 2nd and 4th rows of Figure 4 - 33. The four close coloured tuning targets on the sensing pad were recognised as targets by both resolutions. However, the chance for edges from neighbouring targets to be mixed with each other is much lower with a higher resolution. This is because with a higher resolution setting, there are more pixels representing the gap between two neighbouring targets.

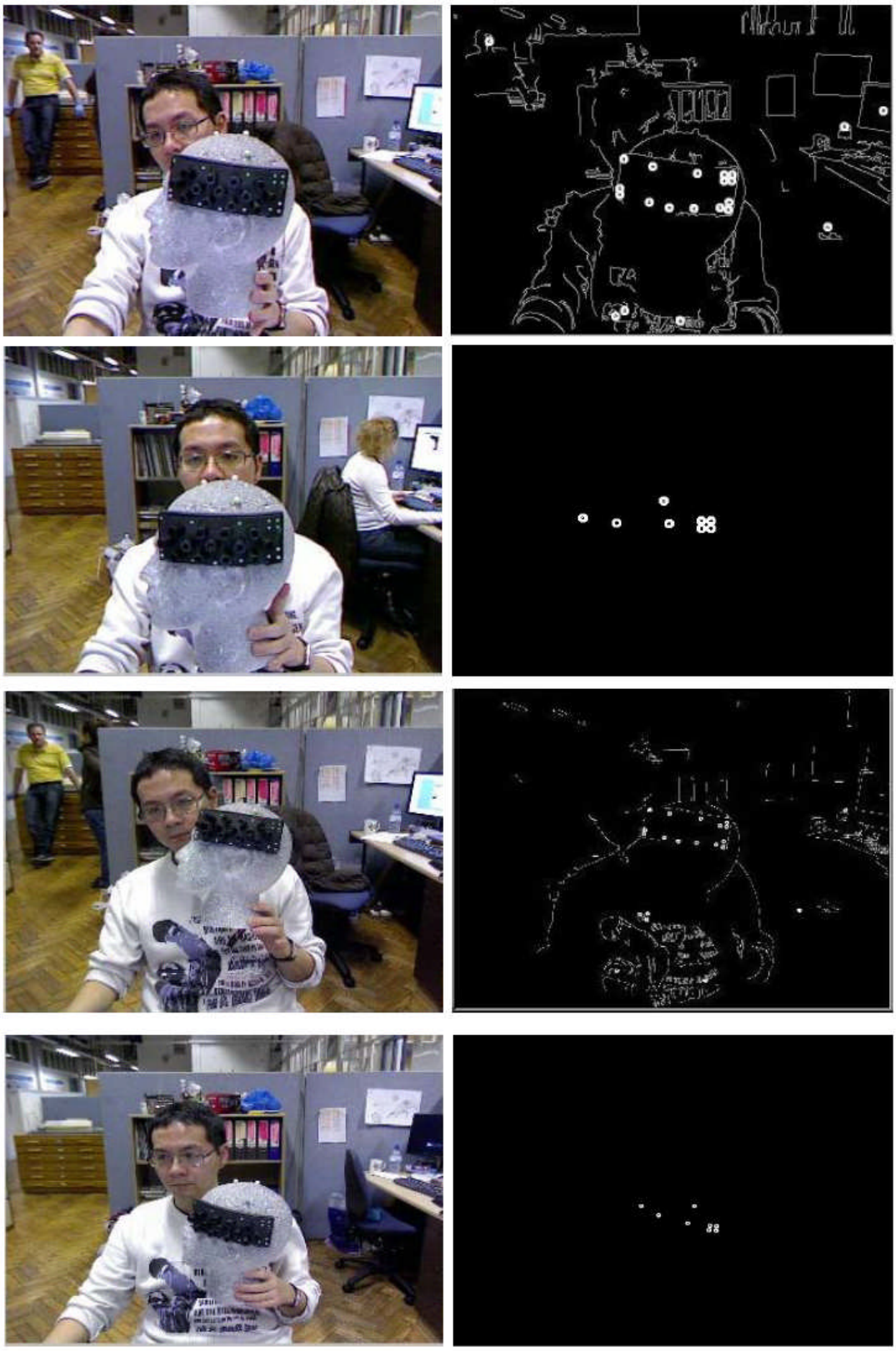


Figure 4 - 33. *Target detection comparison.*

Left: Original scene; Right: Target detection results.

*Resolution of 640*480: Target detection of normal (1st row) and coloured tuning targets (2nd row);*

*1280*1024: Target detection of normal (3rd row) and coloured tuning targets (4th row).*

Setting for colour tuning-target method

A. Dilation operation

While target detection tests were carried out with the colour tuning-target method, it was also important to find out if this detection method could benefit from the morphological operation in the same way as normal target detection. Figure 4 - 34 shows a typical sample result when the tuning threshold (Flowchart 4 - 1) was set to 45:

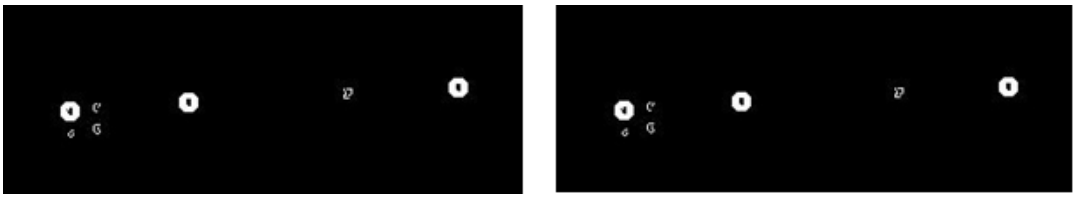


Figure 4 - 34. *Tuning target detections without (left) and with (right) closing operation, tuning threshold = 45.*

The threshold was set to 45 based on the lighting and background clutter in laboratory environment in order to completely filter out background noise. However, the complete filtering out of background noise overlaps with the beginning of losing detected targets. This is because whilst the green component for each pixel inside the coloured tuning target has a higher intensity than the other two components, it varies on a pixel-to-pixel basis. A high tuning threshold such as 45 (0-255 range, 8 bit per component for RGB) could filter out pixels with less green component and cause target edge disconnection. Such a case is shown in the left image of Figure 4 - 34, where 4 out of 7 were not recognised as targets. Lowering the tuning threshold enables all the coloured tuning targets being detected. However, the lower the tuning threshold, the more unwanted targets are detected.

Moreover, tests have shown that applying morphological (both dilation and erosion) operation did not produce a significant improvement in terms of target detection (shown on the right image of Figure 4 - 34), as compared with the significant improvement for normal-target detection. This raised a question of interest – Why does the morphological operation improve target detection results after Canny with white targets but not with coloured tuning targets? This question required an analysis of the difference between edge detection effects, if any, between the normal white-colour targets and the

green-colour tuning targets. By enlarging the edges of those non-recognised targets (not detected by blob detection), it was discovered that there is a pattern difference between the detected edges between the two approaches. As shown from Figure 4 - 35, although disconnection occurs for both edges, the edge after Canny operation for normal targets is an incomplete circle shape (or ellipse when the planar targets were viewed by camera from certain angles, e.g. 45 degree). In the case of the coloured targets, irregular shapes were observed in the edge data after the Canny operation was applied. Furthermore, the irregularity of edge shape for one particular coloured tuning target varied on a frame-to-frame basis making the edge shape after edge detection unpredictable though the physical coloured tuning target itself was a planar circle.

By assuming all the other processes exactly the same for these two approaches, the only difference falls on the target colour, which again led us to refer back to the colour sensitivities of the camera image sensor to different colours. Green has twice as many detectors as red and blue (ratio 1:2:1 for RGB) in a Bayer filter arrangement (Section 3.3.1.2). This colour filter arrangement makes the sensor respond with the highest sensitivity to the green component of each pixel inside the coloured tuning target, as compared with the other two colour components. Although the tuning-target method was mainly designed for target image detection under dim environments, both the white-colour and green-colour targets absorbed a certain amount of light (mainly the room light in the laboratory environment), therefore a re-distribution of all colour components for each target pixel occurs. However, a difference in the filtering process is present. For normal targets filtering criteria depends on the average pixel intensity value. In the case of the coloured tuning targets filtering criteria build upon the green-component intensity value. The re-distribution of all colour components with the Bayer algorithm can reduce the intensity value of the green component against the other two components. In such a case this pixel would be judged as background and assigned with an intensity value of 0 within the tuning filtering process. The result is the unpredictable edge shape following the Canny detector, as shown on the right image from Figure 4 - 35:

Figure 4 - 35. *Comparison of edge disconnections after Canny detector for white-colour targets (left) and tuning threshold filtering followed by Canny detector for green-colour targets (right).*

On the other hand, the closing morphological operation, including dilation followed by erosion, is applied as a structuring element to each image pixel. Small broken edges (left image in Figure 4 - 35) are usually reconnected after dilation is applied. When the erosion was operated after the dilation, only those isolated pixel(s) (usually for random background noise) which had been recognised as foreground are removed leaving the reconnected edge unchanged. However, due to the variation in edge shape and gap size for the coloured tuning targets, a cancelling effect can occur for the edge(s). As an example (Figure 4 - 34), the same number of targets were detected (three out of seven) before and after applying the closing operation.

In order to minimise this cancelling effect, one solution would be to eliminate erosion from the morphological operation by leaving the reconnecting effect after dilation intact. Results, shown in Figure 4 - 36, demonstrate that by applying the dilation operation alone, this detection method was able to allow more coloured tuning targets to be detected and meanwhile achieve a more consistent tracking across image frames [Wong *et al.*, 2010].

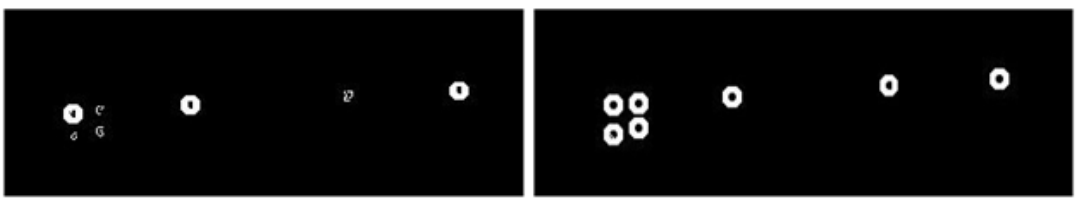


Figure 4 - 36. *Target detection without (left) and with (right) dilation, tuning threshold = 45.*

A larger structuring element mask for dilation has a higher capability to connect edges since the pixel neighbourhood search scope of is increased. However, a mask that is too large, e.g. 5×5 pixels, could easily mix the edge from one target with another,

resulting in two close-by targets being detected as one. Mask size for dilation therefore needs to be carefully chosen. Tests have shown that a 3×3 structuring mask for dilation enabled better target detection results without causing mixing of edges, provided that the target diameter was 4mm and the target-to-camera distance was about 700mm. The dilation process can therefore be used as a tool to compensate the drawbacks (removal of some edges of real targets) brought from a higher tuning threshold (in order to completely remove background noise).

The dilation operation compensated, though not 100%, the shrinking effect of the tuning target edge being detected by enlarging the overall edge (the shrinking effect occurred for most coloured tuning targets since there were always not 100% pixels within a coloured tuning target being recognised as target pixels after the tuning threshold filtering). However, since the detected edge was not exactly the same as the original tuning target edge, especially the non-circle/irregular edge shape, there would be a certain degree of error introduced into the target centroid. The effects of these operators on the actual measured locations of detected targets are reported in Section 4.5.6.2. Also, the small difference in edge shape from one image to another due to the unpreventable light change over time can result in slightly different 3D centroid locations being computed on a frame-to-frame basis. Further investigations will be reported in the accuracy assessment section of the tracking system in Chapter 5.

To summarise, due to the high sensitivity to green, it is desirable to remove as much background noise as possible before the edge detection. A high tuning threshold has been used to achieve that goal. However, edge disconnection and shrinking occurred as a side effect of applying a high tuning threshold. In order to minimise these side effects, dilation operation with an appropriate mask was used. Results have shown a higher consistency in terms of target detection across image sequences after the dilation was applied. A further advantage of using dilation is the reduced time used to process each image compared with both dilation and erosion.

B. Tuning threshold

After dilation was introduced into the image processing for coloured tuning targets, the next is to compare target detection results with different tuning thresholds. Figure 4 - 37 shows typical results:

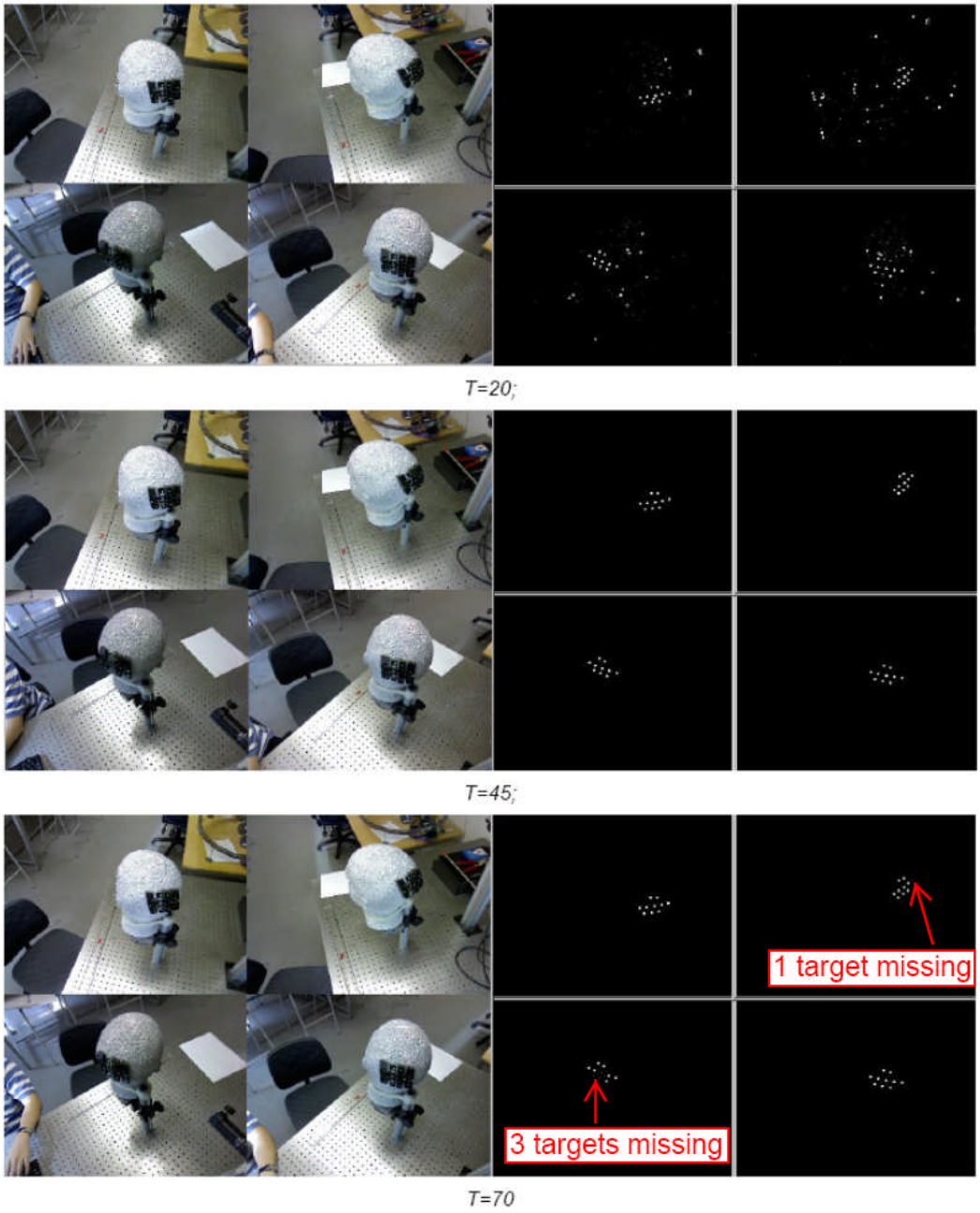


Figure 4 - 37. *Live target detection results from different tuning thresholds.*

It can be seen from Figure 4 - 37 that whilst a low tuning threshold setting allowed all targets (provided that there was no strong light absorption or light reflection on targets) to be detected, a lot of background noise was introduced to the edge detection process causing many unnecessary targets being recognised. There are two disadvantages

from this result. First, the processing time for target matching to compute 3D target locations increases since each detected target from one camera must be compared with each detected target from all the other cameras in the 3D correspondence solution; secondly, intersection ambiguities could easily result from close together targets. For example, if target A and B were close to each other where B was actually image noise, while target C is the target recognised by the other camera, provided that the distance between A and B is within the 3D ambiguity threshold set, then instead of correctly matching target A with target C, target B could possibly be used together with target C to compute a false correspondence. Furthermore, if a non-target object was recognised by two or more cameras simultaneously, 3D locations of this object would be intersected and computed. This would again not only increase the processing time but also result in incorrect and unnecessary 3D locations being output from the system.

Increasing the tuning threshold allows more background noise to be filtered out. However when the tuning threshold was increased to 75, some real targets were lost from the system (Figure 4 - 37 with T=75). Investigation of the missing targets from the second and third camera frames shows that they were located at either corner or edge on the sensing pad and at viewing angles of much less than 45 degrees. The spherical shape of human head and the curved shape of the sensing pad (in order to firmly attach to the head surface) increase the possibility of target occlusions provided with a circle-shape convergent geometry. For example, the second and the third cameras from Figure 4 - 37 had much better views of targets than those for the first (lower left) and the fourth (upper right) cameras. The more acute viewing angle between the real target and the camera, the fewer image pixels present in the image. This decrease in the number of target image pixels has increased the possibility for the target to be completely filtered out prior to edge detection.

While the issue of viewing angle could not be improved without adding more cameras, the tuning threshold can be adjusted to compensate as much as possible the drawback brought by the viewing angle. The solution found was to lower the tuning threshold without increasing the background noise. An example is shown from the middle row of Figure 4 - 37. When the tuning threshold was adjusted to 45, the missing targets from the first and fourth cameras (T=75) were recovered. The recovery of missing targets

from some cameras with acute viewing angles was critical because for a convergent geometry, the more cameras simultaneously intersect a target, the higher accuracy of the target's 3D locations can be computed (Section 3.2.3.1).

The proper settings for coloured tuning targets together with the corresponding effects are summarised in Figure 4 - 38. By utilising the dilation operation with a proper tuning threshold which needed to be slightly adjusted from scene to scene, the multi-camera tracking system was able to achieve better consistency in terms of target detection across image sequences, when the same tuning threshold was used without dilation. This consistency ensured the output of 3D target locations throughout the image sequences since for each set of image frames from all cameras, one set of 3D target locations were computed based on the targets being detected from that particular set of image frames. Experiments with a 2008 model Dell Precision 490 desktop, fitted with Intel quad core 2.0GHz Xeon processors and 4GB memory, delivered sets of 3D target coordinates at two FPS with a four camera tracking system.

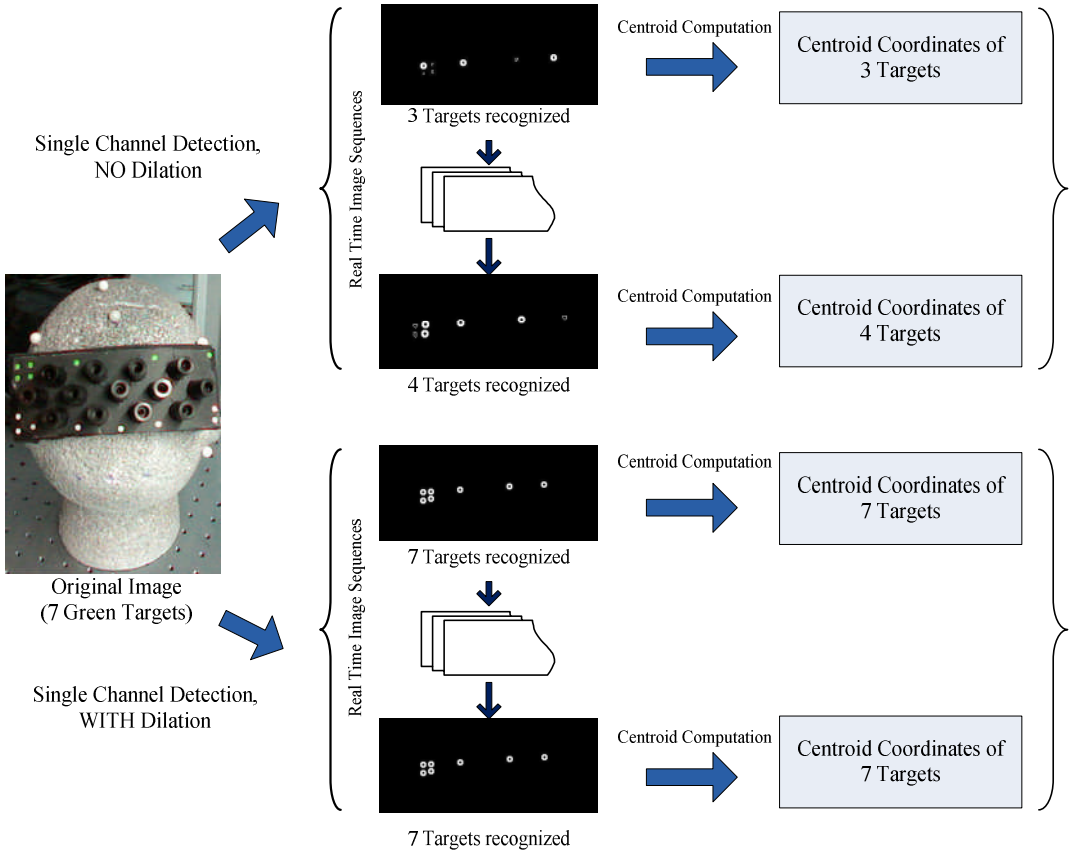


Figure 4 - 38. *Results of tuning-target methods without and with dilation.*

4.5.5.3 Consideration of eight webcam tracking system

While the initial tests with a four webcam configuration showed a consistent 3D coordination, the next step would be extending the system to eight cameras (Figure 4 - 25 and Figure 4 - 26) in order to have a better target coverage in 3D space. However, as shown in Figure 4 - 20 and Figure 4 - 22 in Section 4.4.3.3, an eight camera system with C500 webcams connected to USB 2.0 ports suffers from automatic video streaming buffer size limitations. As a result target image quality is much reduced as shown in Figure 4 - 40, where no targets were detected by Cam 1, 2, 3 and 8 despite being connected to USB 2.0 ports.

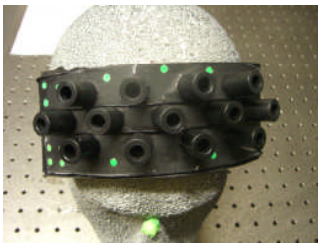


Figure 4 - 39. *Green coloured planar targets used on an optical sensing pad.*

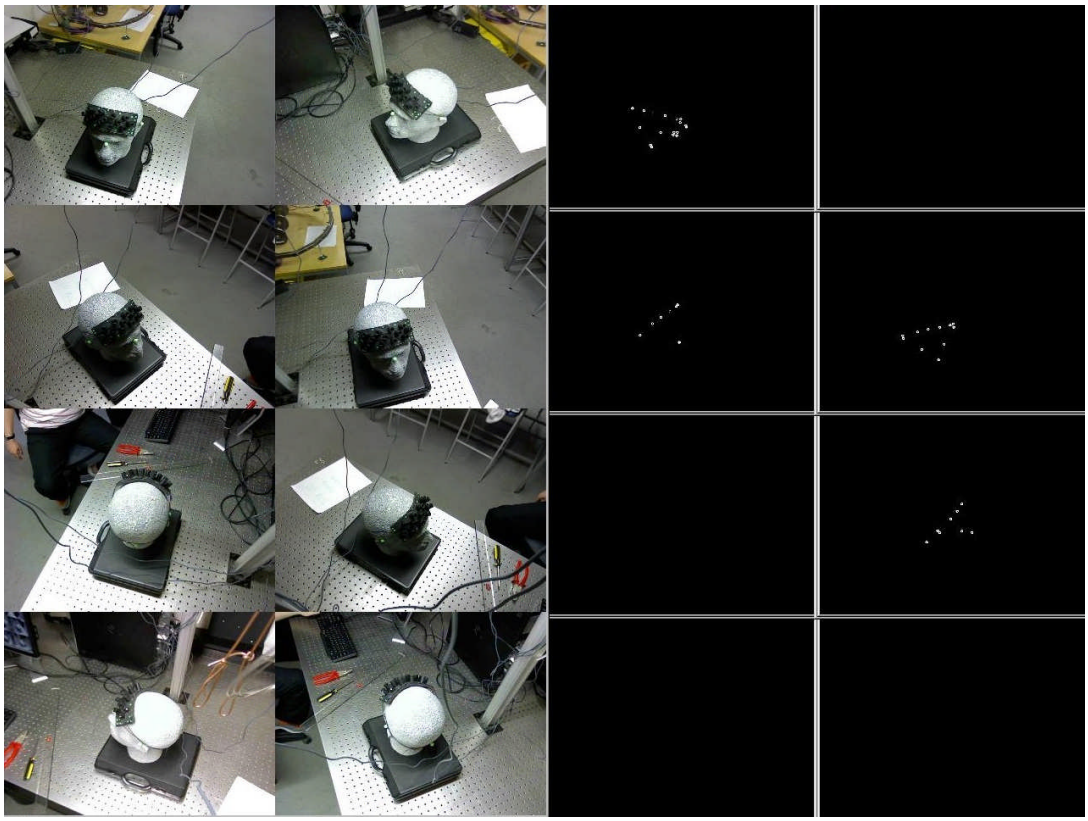


Figure 4 - 40. *Live target detection results with 8-camera configuration.*

(Lowest left: Cam 1; Highest right: Cam 8.)

Similar to Figure 4 - 20 in Section 4.4.3.3, a comparison was made between the images extracted from sequences captured by Cam 7 (USB 3.0 port) and Cam 8 (USB 2.0 port). The image from USB 2.0 port appeared much more blurred than the image from USB 3.0 port, with square grids resulting from high compression clearly visible. The observed effect is illustrated by magnifying a target image captured by Cam 8, as shown in Figure 4 - 41. Cameras connected to the PC with the same settings deliver different image qualities. Those images of higher quality pass the target detection criteria, while those of lower quality fail the same criteria. In such a case no targets can be reliably detected for Cam 1, 2, 3, and 8.

Figure 4 - 41. *Comparison of image qualities captured by Cam 7 (left) and Cam 8 (right). Images have been enhanced for visual purpose of grids.*

Since the image quality of the image sequences varied with the USB configuration set by camera driver, which is highly device dependent, adjustment for each individual connected camera is impractical without a detailed re-write of the camera drivers. However experiments were carried out to determine if software-control of the Logitech camera driver could compensate, to some extent.

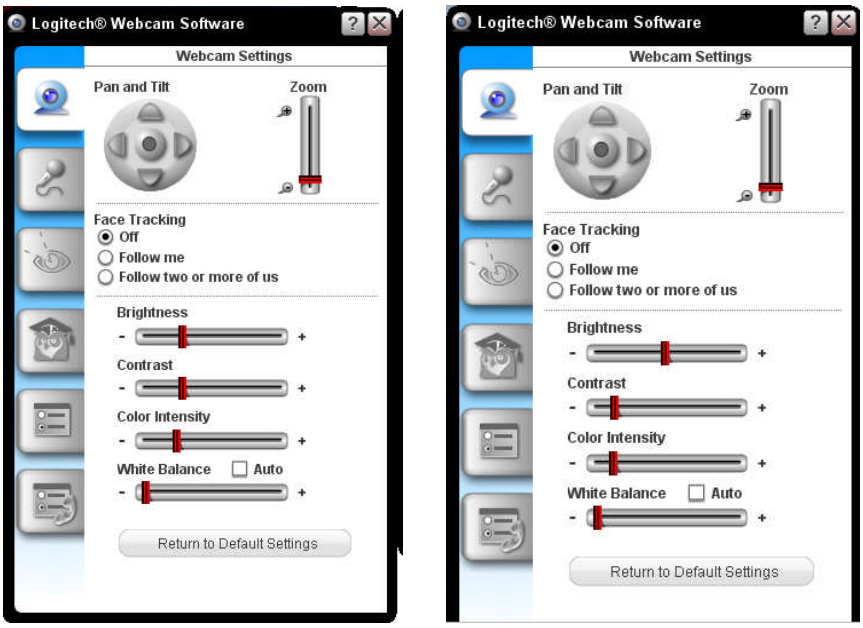


Figure 4 - 42. *Adjustments made for webcams connected to USB 2.0 ports (left) and default setting for webcams connected to USB 3.0 ports (right).*

Figure 4 - 42 shows the adjustments made for Cam 1, 2, 3 and 8, as compared to the default setting for Cam 4, 5, 6 and 7, provided all other settings exactly the same. Lowering the image brightness and meanwhile increasing the contrast allowed a sharper boundary between the foreground targets and the background sensing pad. These settings favoured edge detection and thereafter the edge searching process, while increasing the colour intensity ensured better colour filtering results, by leaving objects with strong green component even in a darker view after the tuning threshold filtering. Figure 4 - 43 shows the coloured tuning targets located on the optical sensing pad, which was firmly attached to the phantom for human brain activation simulation, and Figure 4 - 44 shows the live target detection results in the phantom optical topography study after the adjustments through camera drivers had been made:

Figure 4 - 43. *Coloured tuning targets on optical sensing pad for phantom study.*

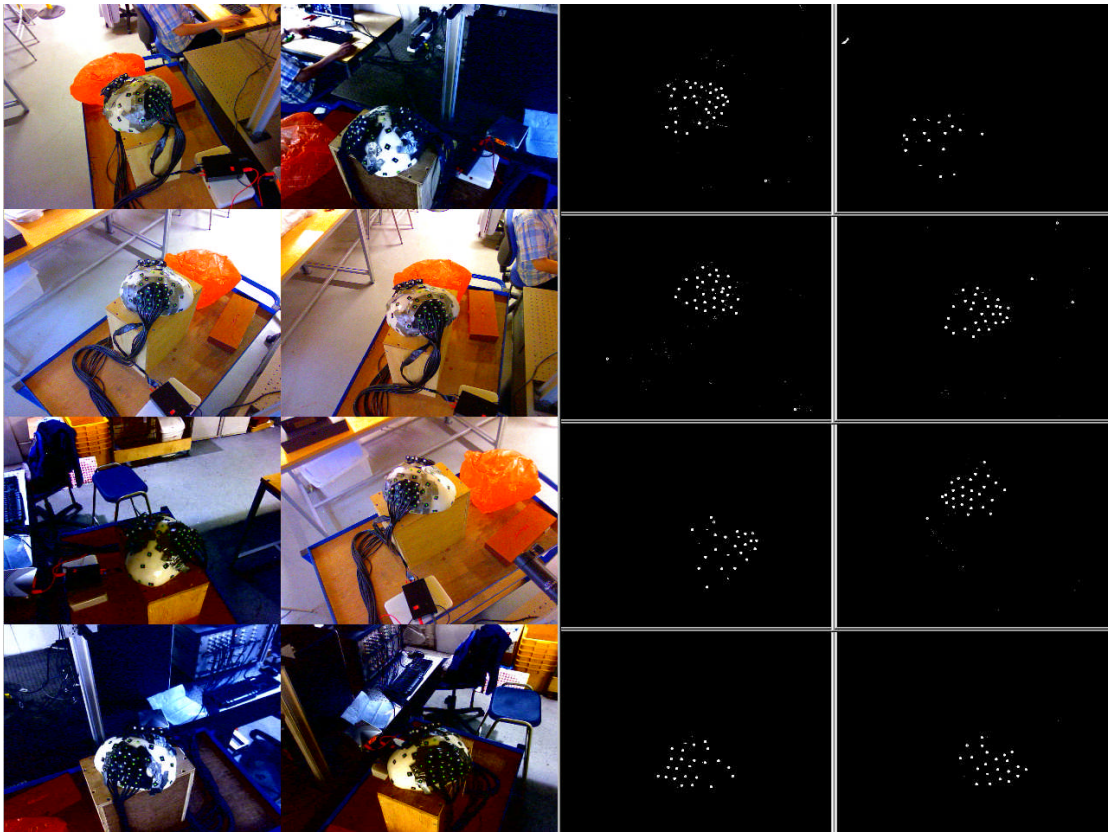


Figure 4 - 44. *Live target detection for phantom study with 8-camera configuration (T=35), after slight adjustments has been made through camera drivers for Cam 1, 2, 3 and 8.*

By comparing the target detection results seen in Figure 4 - 40 and Figure 4 - 44, Cam 1, 2, 3 and 8 demonstrate a significant target detection improvement after camera driver adjustment. In order to allow for the four webcams connected to USB 2.0 ports, the tuning threshold for all eight webcams was lowered to 35 in the phantom study. As expected, some unnecessary targets existed for Cam 4, 5, 6 and 7 (the tuning threshold of 45 shown in Figure 4 - 37 is for the four webcams connected to USB 3.0 ports). Given careful setting (contrast, brightness and colour intensity) and different tuning thresholds applied to each connected webcam individually taking both the connection types and the individual scene into consideration, better target detection results can be obtained. Specifically more detected targets for webcams connected to USB 2.0 ports and no unnecessary targets for webcams connected to USB 3.0 ports can be achieved for this eight-camera tracking system. However, since the possible time used for individual camera adjustment contradicts the goal of easy and quick system setup for clinical use as well as the robustness of the system, the above mentioned approach was not implemented further.

4.5.6 Discussion

4.5.6.1 Target detection with normal targets

The results in Section 4.5.5.1 have demonstrated that the Canny edge detector can effectively extract useful information from a relatively noisy background. Experiments show that for normal targets such as white targets or retro-reflective targets, the threshold adjustment to a particular range suitable for the testing environment is crucial for good edge detection. In the laboratory lighting environment where the images were captured, a lower threshold ranged between 250 and 350 would optimize the edge extraction. Either too low or too high settings for both thresholds decrease the efficiency of edge detection. A Canny edge detector with too low thresholds maintains a substantial amount of background noise, whilst too high thresholds easily hide the useful edge information. Combining all these factors, in future where possible, retro-reflective targets should be used with a proper light control so that targets can be better distinguished from the background as compared with usage of white targets. However this would require each camera to be fitted with a lighting system.

The Canny edge detection combined with the blob detection algorithm can effectively detect normal targets in the scene in most cases. However, experiments have shown that the following factors may lower the blob detection performance:

Noise

Since the edge detector is sensitive to object boundaries in the scene, any unnecessary objects appearing as dots inside the detected scene can easily bias the target detection result. These can arise from scratches on a desk surface, or even a hand moving across the image. As shown from Figure 4 - 45, the number of detected targets varied from 1 to 3 in consecutive image frames. This is due to the two small dots near the real target which are capable of forming a complete shape fulfilling the blob detection algorithm and appear as two individual blobs.

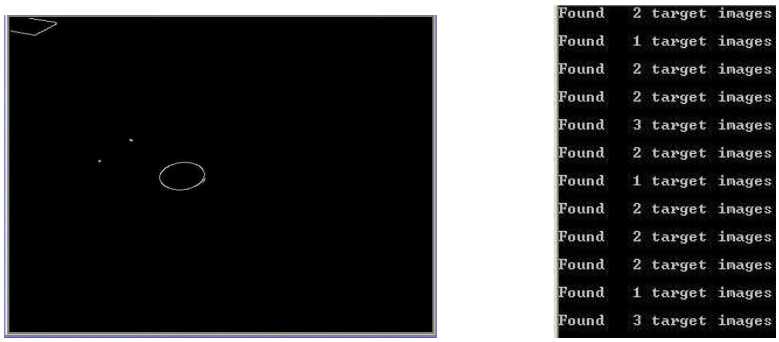


Figure 4 - 45. *Scenario of inconsistent target detection across image sequence due to noise.*

Light reflection

In the real world, such as the laboratory environment, certain degree of light reflection on objects is unpreventable. A target could easily disappear from the scene due to light reflection hiding boundary information. For example in Figure 4 - 46, the blob was intermittently detected because the blob edge was not in a complete shape when light reflection caused a substantial decrease in contrast between the blob and its surrounding background. In more extreme cases, blooming due to too many photons hitting individual photosites in the sensor is apparent and the threshold is unable to distinguish between edge pixels and the background.

Figure 4 - 46. *Situation when light reflection occurs.*

For the Canny edge detector, the larger the difference between the lower and upper thresholds, the more noise eliminated from the scene. However, the connectivity of the extracted target edges is also reduced resulting in missed targets. Morphological operations enable noise reduction (noise pixels tend to be shrunk/filtered by erosion operations) whilst keeping the target edges connected (pixels tend to be added to the broken edge by the dilation operation). However, experiments demonstrate that the morphological operation with a large structuring element mask easily links targets located close to each other (upper row, Figure 4 - 47). Same effect can also be caused by unnecessary iterations of morphological operation (lower row, Figure 4 - 47).

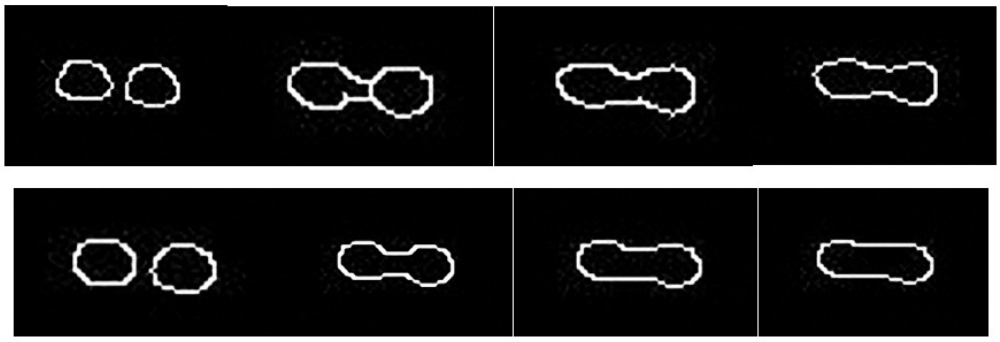


Figure 4 - 47. *Comparison of different morphological operations.*

(target diameter: ~8mm; target separation: ~2.5mm; target-to-camera distance: ~600mm)

Upper row (from left to right, 1 iteration): mask size of 3×3; 5×5; 7×7; 9×9;

Lower row (from left to right, mask size 3×3): iteration of 1; 2; 3; 4.

Combining all these factors, making use of the robustness in terms of edge detection from the Canny edge detector, retro-reflective targets with an appropriate illumination as well as a suitable setting of morphological operations should allow an effective target detection, not only in the clinical environment but for most cluttered scenes.

4.5.6.2 Target detection with coloured tuning targets

Lighting

Results from Section 4.5.5.2 have demonstrated the colour tuning-target detection method is capable of overcoming the drawbacks from normal-target detection method. This advantage is particularly visible for similar sized objects forming sharp contrast against the background, where these objects can easily be recognised as targets with the normal-target detection method. Furthermore, the green-colour tuning-target method has a far better performance in dim environment because of the highest sensor sensitivity to green. This is beneficial for many clinical studies, such as optical topography since dim environment is required to minimise the interference with the infra-red light source emitted from the topographic sensors. Despite the above advantages, tuning-target detection can still be influenced by the lighting in extreme cases. If targets are facing a strong light source such as a window with sunshine outside, as shown in (b) from Figure 4 - 48, a lot of photons fall on the sensor area which covers both the targets and the background. This phenomenon will result in no targets being

detected, either for normal targets or for coloured tuning targets. The other extreme case is when targets are facing a dark area with no light source, as shown in (c) from Figure 4 - 48, as we can observe live images from the left hand side that, targets are not able to absorb even minimum amount of light and therefore will be judged as background.

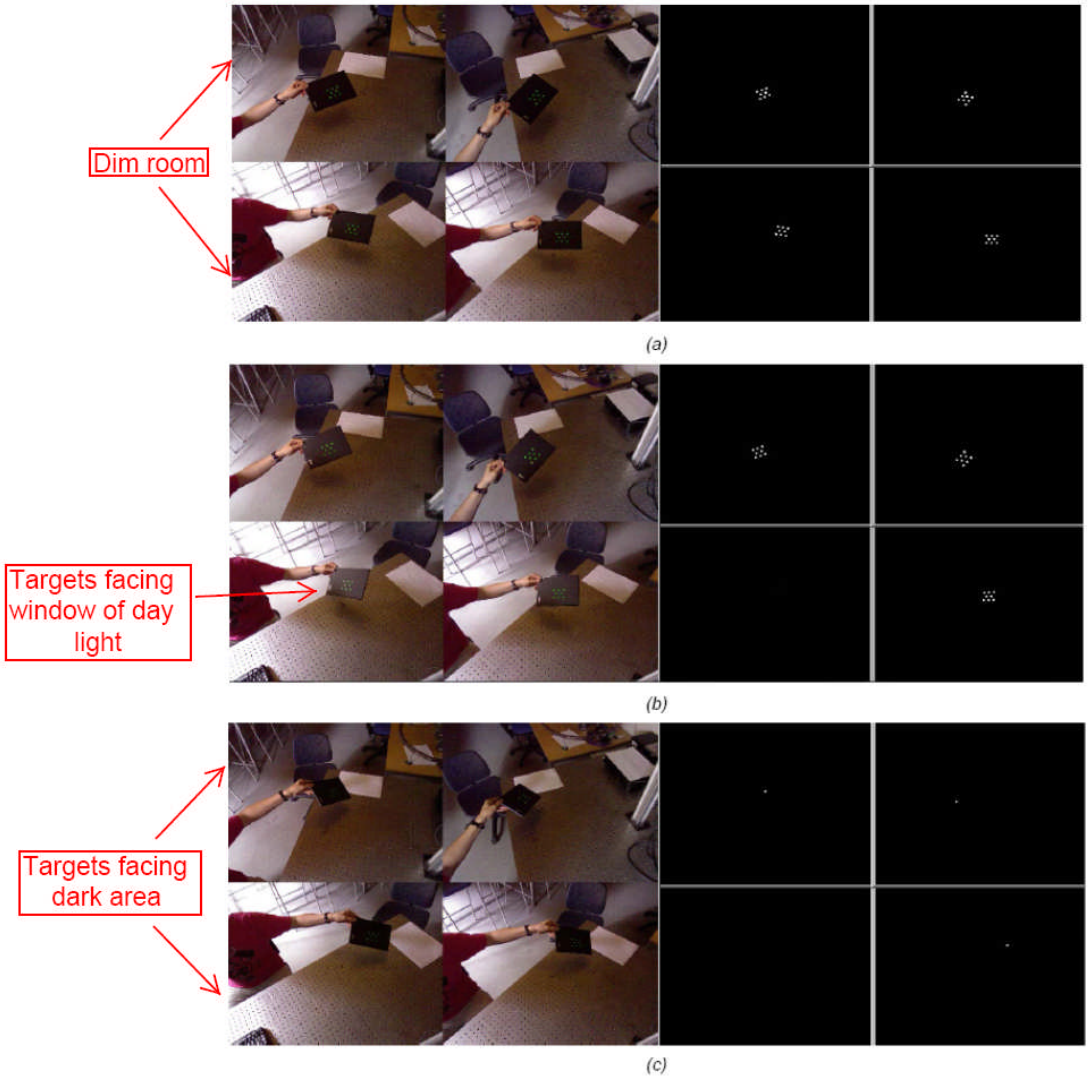


Figure 4 - 48. *Extreme cases of lighting for tuning target detection.*

For optical topography studies involving human subjects, targets are usually distributed around the human scalp, where each target in 3D space is subject to a different amount of light from different angles. Therefore it is difficult to achieve a scenario in reality where each target is seen and meanwhile recognised as a target by at least two cameras. Such a situation makes a 100% successful rapid tracking system for human studies very challenging. Details about the target tracking performance for both the phantom and human studies will be reported in Chapter 5.

Target centroid accuracy

The fact that detected edge for coloured tuning targets was not exactly the same as the original target edge (Figure 4 - 35) would cause a small shift of target centroids computed (Table 4 - 15). Similarly the slight different edge shape being detected from one image frame to another (Figure 4 - 49) due to variation in ambient light reaching each photosite on the imaging sensor for each frame will give rise to error. Such small variations in lighting are at a level which cannot be readily detected by the human eyes.

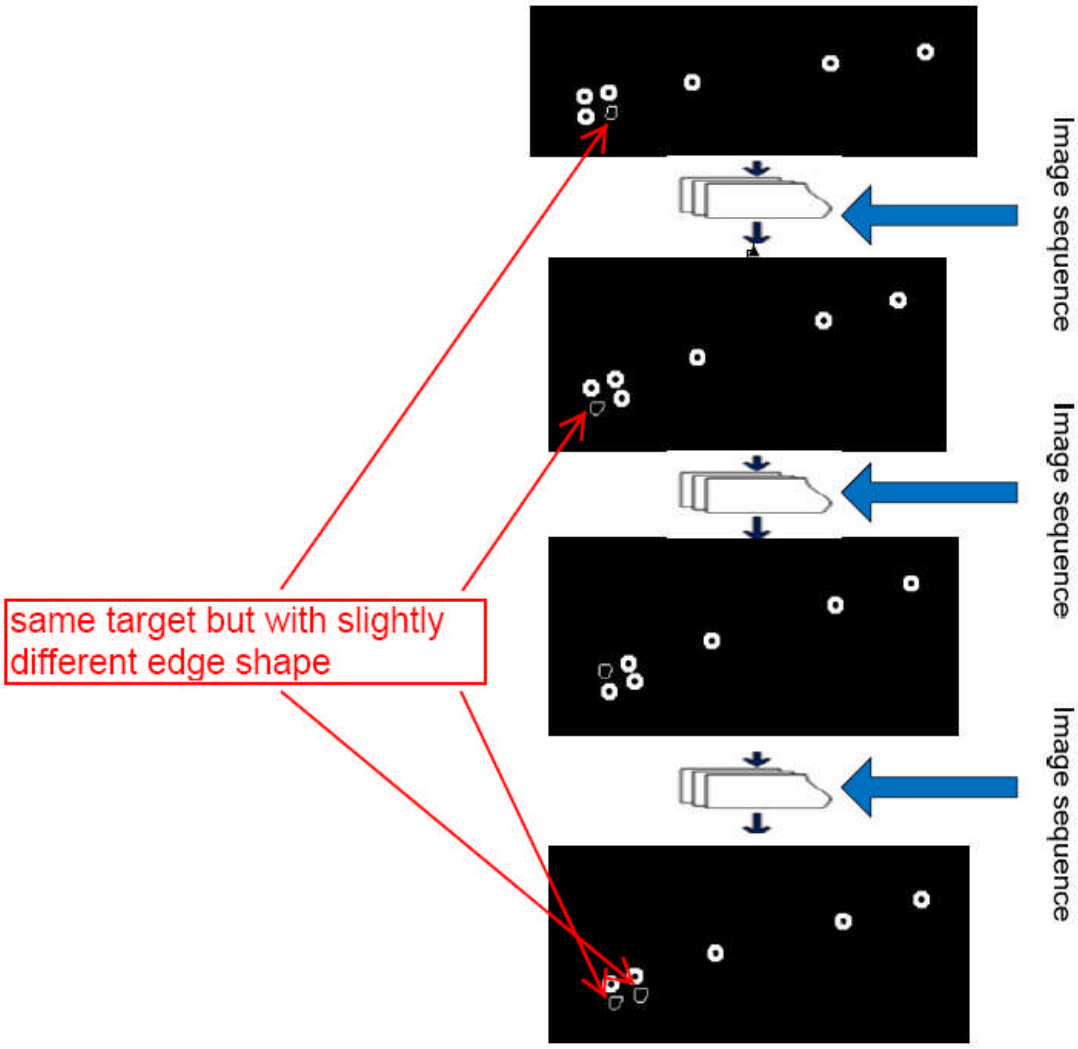


Figure 4 - 49. *Slightly different edge shapes of a target across image sequence.*

To quantify the magnitude in terms of 3D accuracy, tests (Figure 4 - 50) were performed to find out the mean 3D target coordination discrepancies over fifty frames caused by different patterns of target edge detected by the two target detection methods, i.e. normal-target detection and colour target-tuning detection.

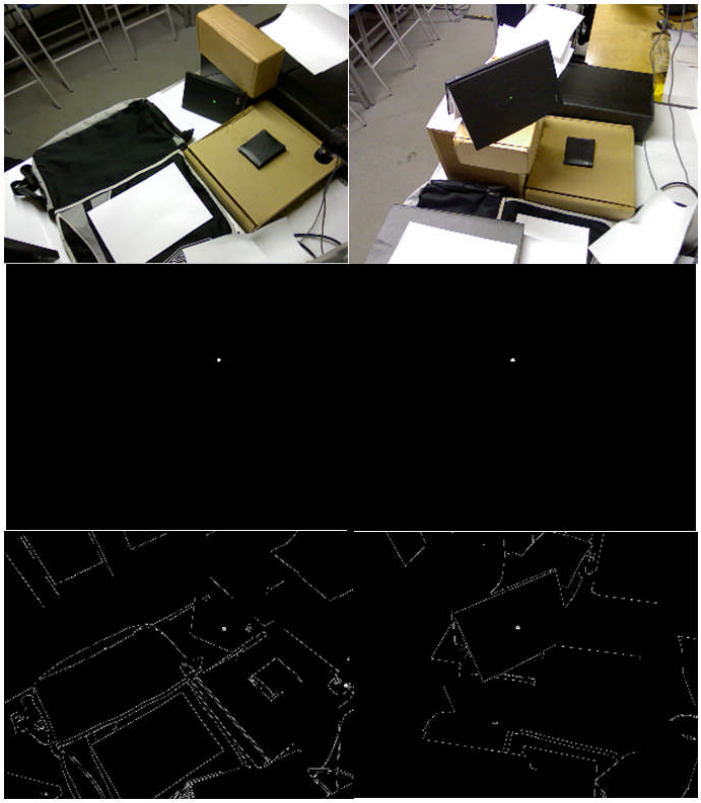


Figure 4 - 50. *Target detections at target-to-camera distances of 900mm (left column) and 600mm (right column).*
1st row: original scene; 2nd row: detection result from colour tuning-target method;
3rd row: detection result from normal-target method.

Table 4 - 15 summarises the results. Method A represents the tuning-target detection method whilst method B represents the normal-target detection method. T represents the coloured tuning threshold whilst T1 and T2 represent the lower and higher threshold for the Canny edge detection respectively.

Camera model used	Logitech C500 webcam			
Number of cameras used	4			
Number of rays used	4			
Colour of target	Green			
Number of frames	50			
Standard/Reference of coordinates	Coordinates intersected by VMS			
Target-to-camera distance	~900mm		~600mm	
Target radius	~8mm		~3.5mm	
Method	A	B	A	B
Threshold	T=45	T1=200; T2=600	T=45	T1=200; T2=600
Absolute Mean Discrepancies	X=0.08 Y=0.12 Z=0.17	X=0.01 Y=0.02 Z=0.03	X=0.06 Y=0.07 Z=0.05	X=0.01 Y=0.01 Z=0.01
3D Mean Discrepancy	0.22	0.04	0.11	0.02

Table 4 - 15. *Comparison of 3D target coordination discrepancies between the colour tuning-target detection (Method A) and normal-target detection (Method B).*

4.5.7 Summary

A new rapid tracking system has been built based on a platform of eight Logitech C500 webcams implementing Videoinput library, Intel OpenCV and OpenGL libraries as well as Microsoft Directshow (Appendix 1, 2, 3). A four camera tracking system is able to produce 2-FPS 3D target coordination while the eight camera system produces 0.5-FPS 3D target coordination prior to the code being optimised. For example text concerning every intermediate result was printed on screen for checking and all live images from all webcams were rendered onto the screen during the tracking process. A comparison showing the speedup of the tracking process when development information was eliminated from the tracking process is given in Section 5.3.4.4.

The normal-target detection method has subsequently been applied in several applications, for example at Airbus where targets on a rotating carbon fibre beam were tracked. The colour tuning-target detection has also been successfully applied in both phantom and human volunteer optical topography studies. Detailed accuracy assessment and validation of the tracking system from both the phantom and human subject studies are reported in Chapter 5. The robustness of the tracking system has allowed users to adjust the minimum number of parameters (Table 4 - 14), normally within five minutes requiring no knowledge of the programs inside, to an acceptable level for the tracking performance under a particular scene. Where these parameters were manually adjusted in this research, they can be automatically adjusted by the computer itself in future. For example, a quick learning time can be allowed for the computer to compare the target detection results with different thresholds it selects and threshold parameters which fulfil criteria set by the computer are considered as suited to that specific test scene.

4.6 Conclusion

The following summarises the key results drawn from this chapter:

- Studies to compare both hardware (video mode) and software (still mode) images have confirmed that there is no difference between these two types of images in terms of photogrammetric measurement. Therefore any conclusion drawn from software images in future directly applies to hardware images. Please note that the internal camera parameters may change between these methods and that a calibration under one set of capture parameters will not necessarily be valid for another situation. This leads to the need for a system calibration;
- Internal parameters of webcams are demonstrated to be stable over time, where tests have been performed in daily, monthly and even yearly basis leading to a conclusion that webcams are able to deliver consistent and reliable 3D coordinate measurements;
- Webcam technology continues to evolve demonstrating significant improvement in image quality and stability even over the short duration of this research. The study of webcams of different generations demonstrates great potential for these systems to meet the demands of a wide range of future measurement applications;
- Although not a key aspect of this research, the improving image quality delivered from webcams and a design of coded targets suited to webcams, probably based on an enlargement of existing designs should allow a fully automatic self-calibrating bundle adjustment to be achieved. The key benefit would be a removal of the need for preparation work prior to target measurement in clinic and therefore allow even more flexibility for webcam photogrammetric measurements in medical community;
- Conventionally fixed image thresholds are used to separate foreground target image information from the background in order to identify candidate target images prior to photogrammetric image measurement. However, the conventional method cannot effectively isolate targets in the presence of a relatively high intensity or variable intensity background. The developed edge detection methods have proven an effective filter in many environments, from a dim testing room where

lighting needs to be minimised to bedside with light sources from different directions. The developed methods have made target extraction more robust and much faster than before since only useful data will be passed to a subsequent blob collection function;

- A rigid multi-webcam tracking system has been built with a total cost under £400 and proved to be reliable in terms of delivering a consistent target tracking accuracy over time. Where a four-webcam system is sufficient to coordinate most targets located on a surface, an eight-webcam system has been built to cover targets in 3D space with occlusion. The tracking system has been proved to work in both general environments and particularly well in more cluttered environments such as those that might be found in during an optical topography test;
- Currently with the maximum hardware resolutions from mainstream webcams (1280×1024), the developed rapid tracking system is able to deliver 3D target coordination at 2 FPS for a four-camera system and 0.5FPS for an eight-camera system with a single desktop computer (Intel Xeon 2.0GHz, 4GB Ram). Furthermore the tracking rate can be significantly increased (roughly 4×) with a lower resolution (640×480) while a 0.2mm accuracy (compared with Nikon network) can still be maintained. Although general movement can be tracked at the current stage, there is still capability for speedup of 3D target coordination for the automated tracking system with coding optimization. Plus, automatic target tracking process without any human interaction should be available in near future with a careful design of programming protocols.

In summary, it has been demonstrated that current webcam technology has made consumer level webcams a good choice for 3D coordination of targets located on a medical device (optical topography sensor) within sub millimetre accuracy for volumes similar in size to the human head. This has opened a new opportunity for photogrammetric measurements with much lower cost while high accuracy is maintained at the same time.

Chapter 5 Practical Application and Validation

5.1 Introduction

This Chapter discusses the practical application of the developed fast multi-webcam tracking system. It is presented as a series of short chronological experiments, each with its own discussion and summary. An overall chapter conclusion places the results in the context of a solution for optical topographic pad tracking. The studies discussed in this chapter, each with their own focus and inter-relationship with each other, are briefly described as follows:

1. Coordination of head targets (Section 5.1)

The main focus in this study is to investigate how many cameras are necessary for optical topographic pad tracking. In particular, a minimal two-camera setup is selected to compute 3D target coordinates by intersection and compare their discrepancies against a six-camera setup which represents the systems available at the commencement of this research. A reference object which had been previously coordinated with an established Nikon D100 photogrammetric network allowed an accuracy check to be performed for the six-camera system.

2. Study with a polystyrene head and calipers (Section 5.2)

Image quality optimization and object coverage led to a four-camera setup being selected to perform fast tracking of a targeted topographic pad on a polystyrene phantom head. Discrepancies between the live tracked coordinates and those computed with VMS intersection are presented. An accuracy assessment, independent of photogrammetric comparison was performed by comparing the separation between targets located on caliper jaws orientated in different directions within the measurement volume.

3. Optical topographic study with a phantom head and volunteerss (Section 5.3)

This section describes an eight-camera setup and its practical use for a phantom study and a study of two human subjects in collaboration with Department of Medical Physics and Bioengineering of UCL. The phantom study compares the tracked coordinates from

the system with those computed from post processed VMS intersection. Internal phantom anatomy was co-registered with optodes utilising 10-20 bony landmarks as common points. These co-registered anatomical positions estimated from photogrammetric bundle adjustment were compared with those estimated from optical image reconstruction. The human study looks into the differential movements between the topographic pad and head by co-registration of targets over a time series of continuously tracking images, where this technique allows optical topography clinicians to validate optical imaging studies in nearly real time sense.

5.2 Coordination of head targets

5.2.1 Introduction

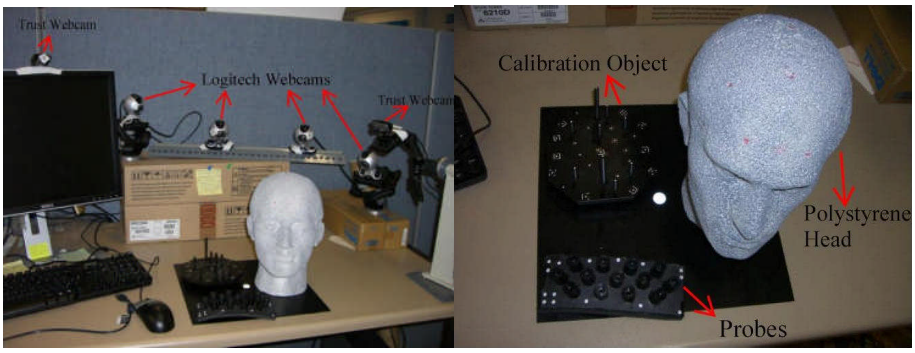
Low cost webcams have a USB interface which imposes limitations on data transfer speed and camera connectivity. Therefore it is desirable to minimise the load on the USB bus and subsequently on the computer for processing where possible. A key question prior to developing a fast optical topographic pad tracking system is “How many cameras are necessary to give a maximum coverage of key targets and accommodate occlusion whilst delivering 3D coordinates with an acceptable accuracy?” In order to experimentally assess this question for optical topography tracking, photogrammetric target intersections were computed in VMS (Geometric Software) to explore the differences in 3D target coordination between a minimal two-camera stereo setup and a more comprehensive six-camera setup which contains the maximal number of available webcams at the commencement of this research. 3D target coordinate discrepancies between both two and six-camera setups and a reference Nikon D100 network was used to assess performance.

5.2.2 Method

5.2.2.1 Experiment setup

This experiment was carried out in the formative stages of this project as practical imaging geometries were being worked out prior to designing and building the camera framework used for the clinical trials. Figure 5 - 1(a) shows the overall experimental setup and Figure 5 - 1(b) shows the objects used in the experiment. Six webcams were positioned in a crown shape around the testing object to provide the convergent geometry. The distance between cameras and testing objects ranged between 300mm and 500mm. The cameras used in the experiment were two Trust WB-5600R webcams and four Logitech Quickcam Pro 4000 webcams representing the systems available at the commencement of this research. Noting that both sensors have a 4× image

interpolation (actual pixels on sensor: 692×504 for Logitech cameras (Table 4 - 1); 1280×1024 for Trust cameras according to manufacturer sheet), where the single shot image resolution of Trust and Logitech cameras is 2560×2048 and 1280×960 (4μm) pixels respectively.

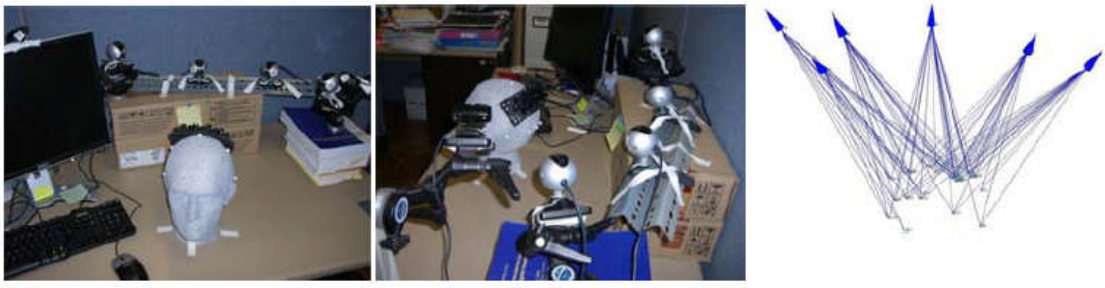


(a) (b)

Figure 5 - 1. *(a) Experimental setup; (b) Objects used in the experiment.*

5.2.2.2 Target intersection

The polystyrene head with a sensing pad on it is representative of the volume and configuration of a typical adult subject (Figure 5 - 1(b)). Provided with geometrically stable cameras and camera fixture, the unknown target positions on the head surface and the pad can be intersected with the camera orientation parameters derived from a self-calibrating bundle adjustment (Section 4.4.2), where the parameter uncertainties were propagated into the least squares intersection. An alternative is to run the bundle adjustment, i.e. having some coded targets and correcting for orientation changes during tracking. However, this method was not implemented in this research because the targets would add an extra dimension to the clinical work.



(a) (b) (c)

Figure 5 - 2. *(a) Front view of the network configuration; (b) Back view of the configuration; (c) Relative positions between intersected targets on the pad/head and six webcams.*

5.2.3 Results

The 3D positions of the targets were intersected first using all six cameras and then a subset of two which were located in the middle of the camera crown and providing the most complete target coverage on the sensing pad. These data allowed a comparison of the intersected target coordinates on the sensing pad (most important in terms of source localisation) between a six-webcam setup and a selected stereo pair of webcams (Table 5 - 1). As a reference, the discrepancies between 3D target coordinates located on a stable reference object introduced into the field of view and imaged by the six webcam configurations (Set 2 of Table 4 - 9) were compared (Table 5 - 2). The 3D target coordinates of the reference object had been previously computed using a photogrammetric network taken with a Nikon D100 camera with inter-target distances provided by a pair of digital calipers (Table 4 - 2).

		Discrepancy (mm)			
	Target Id	X	Y	Z	3D
Targets on head	603	-0.68	1.06	0.07	1.27
	604	-0.14	0.06	-0.44	0.46
	613	-0.01	0.36	-1.31	1.36
	614	-0.06	0.55	-0.25	0.61
	615	-0.23	1.06	0.14	1.09
	616	2.11	2.36	4.34	5.37
	Mean (absolute)	0.54	0.91	1.09	1.70
	Standard Deviation	0.98	0.81	1.99	1.84
Targets on pad	501	-0.09	0.14	-0.18	0.25
	503	-0.04	-0.03	-0.16	0.17
	504	-0.07	0.89	-0.57	1.06
	505	-0.93	0.95	-1.16	1.76
	507	-0.20	0.58	-0.50	0.79
	508	-0.17	0.41	-0.48	0.65
	509	-0.30	0.08	-0.48	0.58
	510	-0.07	0.30	-0.22	0.38
	523	0.01	0.47	-0.39	0.61
	531	-0.20	0.01	-0.37	0.42
	532	-0.20	-0.04	-0.66	0.69
	Mean (absolute)	0.21	0.35	0.47	0.67
	Standard Deviation	0.26	0.35	0.28	0.44

Table 5 - 1. *Target coordinate discrepancy between the six- and two-webcam intersection data.*

	Discrepancy in mm			
	X	Y	Z	3D
Mean (absolute)	0.08	0.07	0.08	0.15
Standard Deviation	0.11	0.10	0.09	0.07

Table 5 - 2. *Target coordinate discrepancy between the reference object and the six-camera network.*

5.2.4 Discussion

Regarding the global registration of target positions to the head, most 3D coordinate discrepancies in target position between the six-camera and two-camera solutions are within 1mm. A single large 5.4mm discrepancy (Target 616 in Figure 5 - 3) is included in the mean and standard deviation figures. The cause of this discrepancy is due to partial occlusion and low contrast as discussed below.

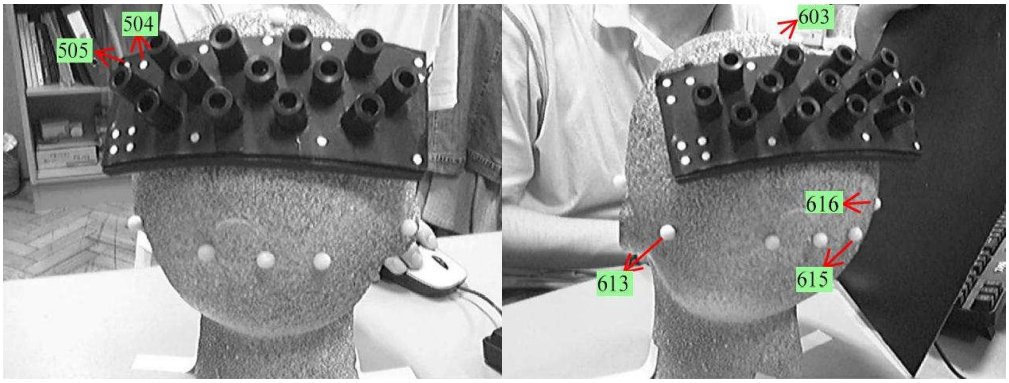


Figure 5 - 3. *Images captured by a stereo pair of Logitech Quickcam Pro 4000 webcams.*

The accuracy of 3D coordination is not only dependant on the imaging geometry in relation to the spatial positions between optodes and occlusion, but also on target image contrast. These factors are particularly important given unavoidable variations in lighting and background image content under clinical conditions.

1. Target Image Contrast

The contrast between the target image and its background affects how effectively the target boundary can be automatically identified and hence the image measurement used to compute the centroid. The white dots attached to the grey polystyrene head do not produce strong contrast with the result that target location is affected (Figure 5 - 4).

Figure 5 - 4. *Examples of target images with low contrast.*

(manual target centroid measurements were applied for all the above target images since none are readily capable of being automatically measured by the centroiding algorithm from VMS)

Target image quality accounts for the fact that targets on the head have a larger range of coordinate discrepancies when compared with the targets on the pad (Table 5 - 1). High contrast against the background is a critical factor for target positioning accuracy [Darrell *et al.,* 1994; Iwai *et al.,* 1999; Hilton, 1999].

2. Partial occlusion

Partial occlusion of a target in an image will introduce a bias in the computed 3D target coordinates if the image measurement is not rejected. As an example, target 616 (Figure 5 - 3, also highlighted in Table 5 - 1) where the right image only contains a small portion of the target image. The incomplete target measured by the stereo pair of cameras increases the difficulty and error in determining the 3D target position as there is no effective redundancy in the measurements. However at least two images from the six-camera geometry contain the complete shape of target 616 and therefore the intersection on this target is possible. Existing software considers the circularity of the target images, but this is not robust with web-cam images as partially occluded targets can appear as smaller circles due to image blur and compression. Such target images could be eliminated by checking predicted image dimensions as a function of range from the camera given knowledge of the expected target location, but this is not currently implemented in either VMS or the implementation of fast tracking developed here. The large discrepancy in 3D coordination of target 616 between the two sets of cameras is attributable to both contrast and occlusion. Another example of occlusion is indicated by target 505 where partial occlusion occurred in the left image of Figure 5 - 3. The problem of occlusion usually becomes more significant as more targets are involved and therefore, where possible, more than two cameras are desirable for a more complete coverage of targets at various 3D locations.

Considering if these data will be fit for purpose requires an understanding of the ability to repeatable place markers on the human head in order to define the 10-20 coordinate system. It was noted that for EEG, the accuracy of source localization critically depends on the accurate and reliable placement of the electrodes onto the 10-20 coordinate system (Section 2.2.3). Test-retest measurements following the repeated establishment of physical locations for 10:20 points exhibit errors of up to 7mm, and inter-subject variability of up to 7.7 mm have been reported even when electrodes were applied by an experienced senior registered EEG technologist [Towle *et al.*, 1993]. In addition, the

amount of placement error was dependent on the electrode positions, with more lateral electrodes displaying more uncertainty than electrodes placed on the midline. A major part of this problem is that no anatomical landmark could be determined with less than 5 mm of uncertainty. Additional within- and across-subject error can also result from the experience of the experimenter, and the length and type of a subject's hair. Regarding the accuracy of source localization (position of optodes) and determination of anatomical landmarks such as nasion and inion, a 1mm level of fitting should be sufficient for the optical topography application of this research.

5.2.5 Summary

Laboratory experiments have demonstrated that whilst a stereo pair of low cost Logitech QuickCam Pro 4000 webcams were able to intersect targets to within 1mm of that possible with an image network of six similar webcams, partial target image occlusions occurred more frequently and are therefore considered to be an issue for the stereo case. This problem can be largely reduced by a convergent imaging geometry containing more cameras directed towards the sensing pad location.

5.3 Study with a polystyrene head and calipers

5.3.1 Introduction

Section 5.1determined that a stereo pair of cameras are not able to reliably coordinate targets situated on both optical sensing pad and human head. However there is a trade off since the more USB cameras that are connected into the same PC, the lower the frame refresh rate of the tracking system. In order to check and enhance performance, a study prior to the practical optical topography studies was conducted. The aim of this test was to determine accuracy and speed in terms of rapid target positioning with a four-camera setup. As such the test is divided into two, firstly a check against a model head and optical sensing pad and secondly comparisons against 3D target positions made with a pair of digital calipers. This investigation provides a reference for the accessibility, portability and accuracy of such a four-camera tracking system.

5.3.2 Method

5.3.2.1 Experimental setup

Experiments have been carried out to find out the consistency and accuracy with which a set of four Logitech C500 webcams can intersect 3D targets in locations representative of an optical topography experiment. Table 5 - 3 details key parameters for this study. Note that the Logitech C500 webcams and USB3.0 cards represent the technology level available towards the end of this research project.

Cameras used	4 Logitech C500 webcams with PCI USB 3.0 slots
Interior/exterior orientation parameters	20/07/2010 (Section 4.4.3.2)
Wheel ring (for mounting webcams, Figure 4 - 25) diameter	600 mm
Vertical distance between ring and tested object (top head)	~370 mm
Distance between camera and tested object (top head)	~550 mm
Planar target diameter	4mm
Target detection tuning threshold	45
Frame refresh rate (FPS)	2

Table 5 - 3. *Key parameters to set up a four-camera C500 tracking system.*

The four webcams were positioned in a half circle shape above the testing object to optimise the convergent geometry. The hardware image resolution provided by the Logitech C500 cameras is 1280 by 1024 (4μm square pixels).

5.3.2.2 Phantom test

In this study 10 coloured targets on a small sized sensing pad, designed in 2008 by the Biomedical Optics Research Laboratory at UCL, were attached to the back of the polystyrene head. The experimental setup is identical to that given in Table 5 - 3. Figure 5 - 5 shows a screenshot of targets coordinated with the tracking system utilising the target detection method with coloured tuning targets (Section 4.5.4.2).

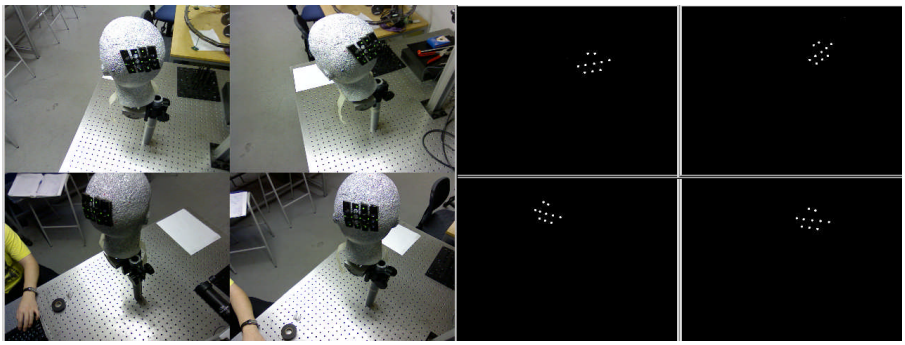


Figure 5 - 5. *Views (left) and targets detected on the sensing pad (right) in a four-camera tracking system.*

To assess the accuracy of 3D coordination with the tracking system, tracked target coordinates were compared with those intersected from VMS (Geometric Software), where the same set of camera parameters (20/07/2010 from Section 4.4.3.2) were used and the four images (one per webcam) were hardware images (Section 4.2.2.3) extracted from video streams. Object positions were assumed to be fixed during this study. Note that not all target images were readily capable of being automatically measured by the centroiding algorithm from VMS and therefore manual target centroid measurements were applied for some of the target images.

5.3.2.3 Caliper test

Besides the simulated study with the optical sensing pad, direct accuracy assessment with a caliper was also conducted. The information for this study is shown in Table 5 - 3. One coloured target was attached to each side of the caliper and the target detection method used was the same as above (Section 4.5.4.2). Whilst the absolute distance between the centres of the two attached targets was hard to measure to high accuracy, changes in target separation, obtained by opening the caliper jaws were used instead to access the target coordination capability of the four-camera tracking system. Change in target separation was simply measured with the in-built electronic meter to a high accuracy with a standard deviation of 20μm according to the caliper manufacturer's calibration data. While the distance between the two sides was adjusted, slopes between the two coordinated targets tracked were compared with the adjustment made.

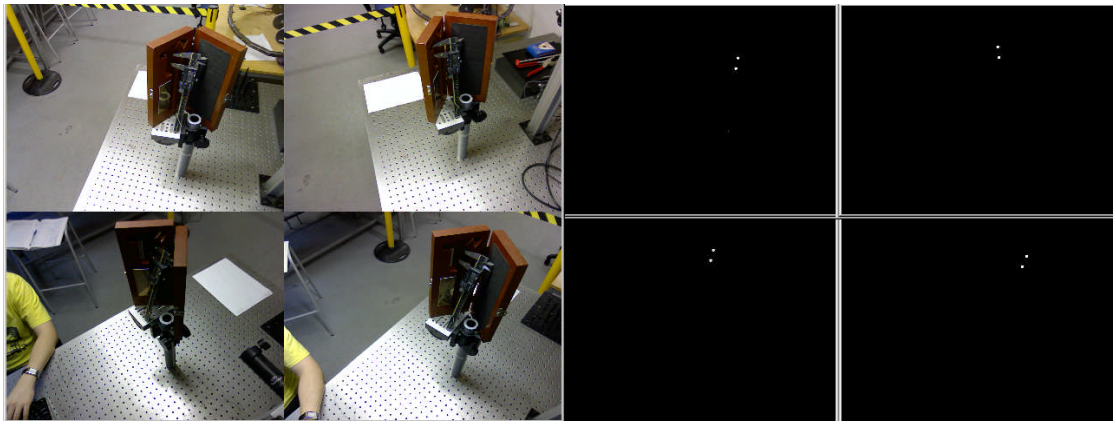
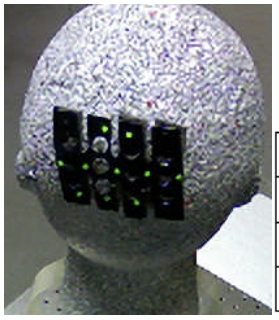


Figure 5 - 6. *Views (left) and targets detected on caliper (right) in a four-camera tracking system.*

5.3.3 Results

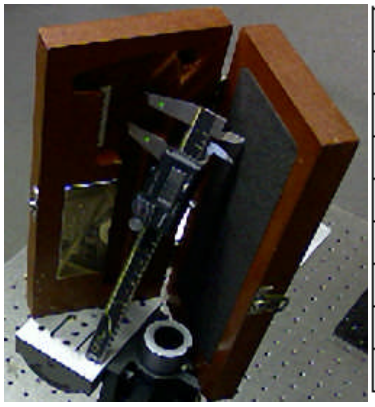
In both the phantom and caliper tests, all rays from all four cameras were used in the 3D target coordinate intersection computation. Table 5 - 4 presents a summary of target coordination comparison between the automated four-camera tracking system and established intersection processing with VMS:



Discrepancy (10 targets)	X	Y	Z	3D
Mean (absolute) (mm)	0.13	0.20	0.26	0.35
Standard Deviation (µm)	8.8	13.2	9.9	18.7
Mean Image Residual (µm)	0.6	0.2		

Table 5 - 4. *Target coordinate discrepancies between the live tracking system and VMS intersections. Left: A camera view of the 10 targets on optical sensing pad.*

Table 5 - 5 gives a summary of relative separations between targets on the opening caliper measured between the tracking system and the in-built electronic caliper:



Relative separation between targets (mm)	Discrepancy (mm)
2.00	0.18
4.00	0.03
14.00	0.02
19.00	0.29
29.00 (1)	0.17
29.00 (2)	0.32
29.00 (3)	0.27
Mean Discrepancy (mm)	0.18

Table 5 - 5. *3D discrepancies between the electronic meter and the live tracking system. Left: A camera view of the opening caliper.*

5.3.4 Discussion

In this study, coloured tuning targets (Section 4.5.4.2) have enabled a consistent target tracking across image frames (as previously illustrated in Figure 4 - 38). This method has proven reliable even if the target image contrast is low, for example even in the presence of bright reflections in close proximity to the target image (Figure 5 - 7). Under such circumstances the white target detection process will fail whereas the colour tuning-target detection method is still able to detect the colour component and filter out the surrounding area in order to provide a more reliable target image detection. Note however that as described in Section 4.5.6.2, this higher reliability of the colour tuning-target detection method might come with the cost of slightly lower accuracy compared with the normal target detection method.

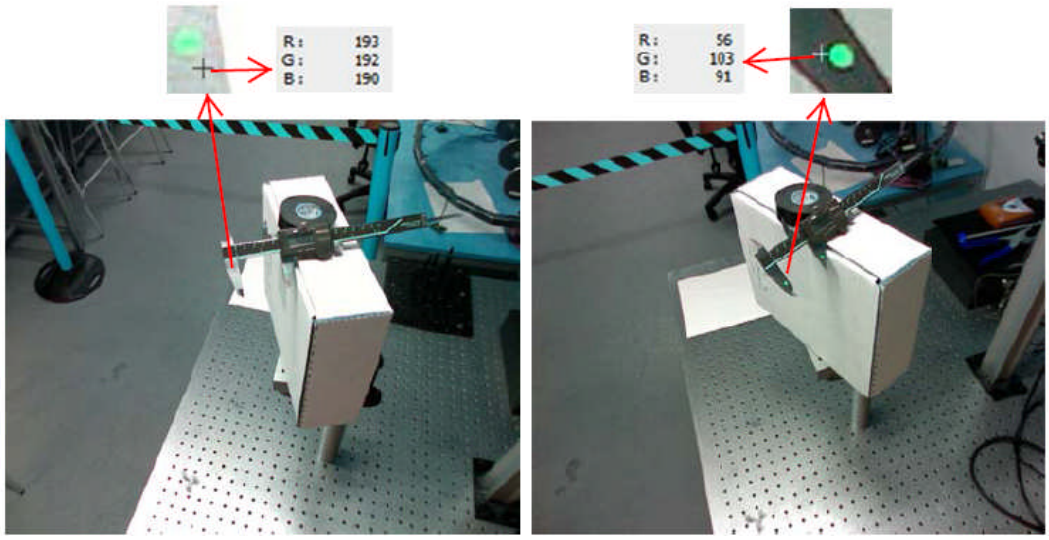


Figure 5 - 7. *Specular reflection on the metal caliper jaws (left) and a better viewing angle under the same lighting conditions where the green targets can be clearly observed (right).*

In cluttered environments with multiple views it is nearly unpreventable that view(s) from one or more cameras will contain target images with a certain amount of light reflection. A suitable method to tackle this problem is necessary for tracking consistency across frames. In the clinical environment, the scene is usually composed of more than one light source such as computer monitor (Figure 3 - 13). In such cases the colour tuning-target detection method developed in this research can largely eliminate the concern of different light sources and has proven particularly useful for target detection.

Furthermore, as mentioned in Section 5.1.4, the contrast between the target image and its background directly influences the successful rate of target detection. Coloured targets are easier to distinguish from their background with various types/colours when compared with normal white targets, which will not be detected under the circumstance as shown on the left image of Figure 5 - 6. The exception is provided by retro-reflective targets, however these require appropriate axial lighting adding to system cost and typically have a useful light return that is limited by acceptance angles of +/- 50 degrees (Section 3.4.2).

In terms of target coordination accuracy, although the mean discrepancy is 0.18mm for the caliper test with each target intersected by all four rays, there are maximum variations in the computed discrepancies of 0.3mm across measurements where the minimum discrepancy is 0.02mm and the maximum discrepancy is 0.32mm (Table 5 - 5). Figure 5 - 8 compares the images associated with this difference from the same camera. The green channel histograms of the selected targets shown below explain this variation.

For the caliper test with the maximum discrepancy, a white box was placed under the caliper in order to roughly match the relative distance between the targets and all cameras (Table 5 - 3). The white surface gave rise to strong light reflection in close proximity to the target (highlighted on the upper right image of Figure 5 - 8). Although the targets could be detected under these circumstances, the number of pixels containing green component inside the target is significantly less in the presence of a bright surrounding surface (lower row of Figure 5 - 8). In this way, edges of the target image recognised by the tuning method were different from the true/original edges. Furthermore the squared centroid computed based on this edge information is considered to be unreliable and therefore there is a discrepancy of 0.32 mm compared with the caliper measurement. Although such a discrepancy is still within the level of acceptance for the topographic reconstruction, for the best tracking accuracy, over-exposure objects in the proximity of targets should be avoided. As shown in Figure 5 - 8, the light reflection from the metal table in such a range of distance (>500mm) is considered to be acceptable, while the light reflection from the white box underneath (<100mm) will result in less accurate target coordination.

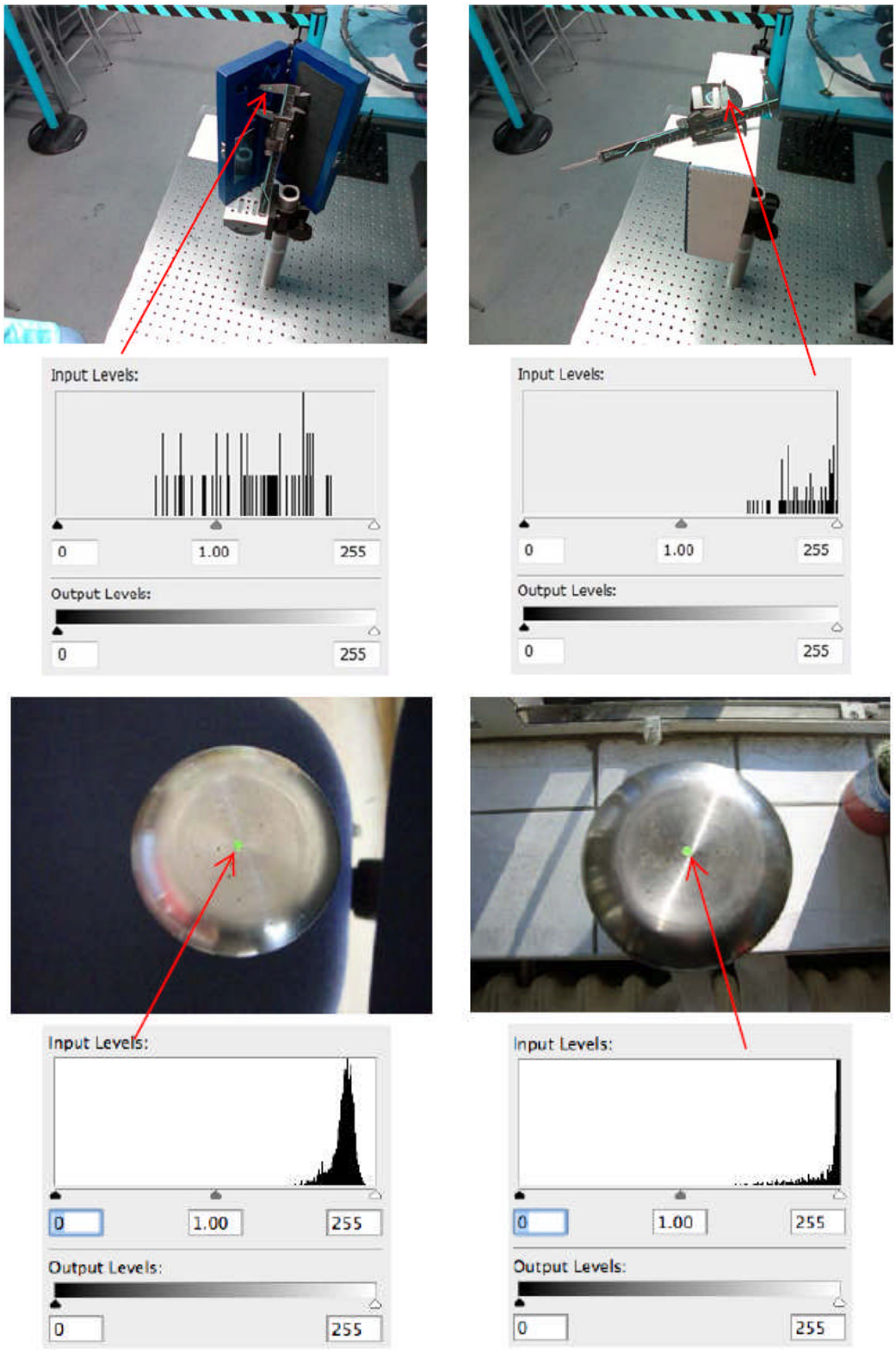


Figure 5 - 8. *Upper: Views of Camera 3 with minimum (left) and maximum discrepancy (right) during live tracking and their associated green channel histograms of selected targets. Lower: Target image taken with Casio Exilim EX-Z500 (resolution: 2304×1728 pixels) in the absence (left) and in the presence (right) of a bright surrounding surface and their associated green channel histograms.*

For target coordination on the phantom, static images from the tracking system were captured during the tracking process for intersection post-processing in VMS. This VMS intersection provides a reference to assess the accuracy of the output 3D coordinates from the live tracking system. Results have demonstrated 3D discrepancy better than 0.4mm. While this discrepancy does not represent the absolute accuracy assessment, 3D discrepancy better than 0.2mm between the VMS intersection and the standard Nikon D100 photogrammetry survey (Table 5 - 2) is provided as a reference.

The larger 3D discrepancies for the phantom test (0.35mm) compared to the caliper (0.18mm) are attributable to the target orientation with respect to the cameras. In the phantom study, the circular targets had a diameter of 4mm and were attached to the optical sensing pad, which was then bent to match the shape on the scalp (Figure 5 - 5), whilst the two circular targets on the caliper were on the same 2D plane. Caliper targets are pointing directly at the tracking cameras resulting in their imaging as nearly a full circle. However, whilst targets on the curved sensing pad, especially those on an edge or corner, could be recognised as targets by the coloured target detection method, they may not be detected as a complete circle. In this case, the squared centroiding for subpixel target location (discussed in Chapter 3) did not include all pixels in the target but a portion of pixels inside the target resulting in a target centroid bias which accounts for the higher discrepancy in the phantom study compared with the caliper study.

As shown in Figure 5 - 5, a four-camera tracking system where the cameras are placed in a crown shape enabled full target coverage. For both the phantom and caliper test, every target was seen and detected by at least two images from the four-camera geometry and therefore the four-camera fast tracking system was able to provide a reliable target intersection solution. Another point to note is that all the four Logitech C500 webcams were connected to four PCI express slots for this study (the maximum number of PCI express slots the desktop used in this research can accept) and therefore the four-camera tracking system delivered 3D target coordination at a rate of 2 frames per second, i.e. a full image processing (Section 3.4.4 and 3.5) leading to 3D target coordinates being output completed every half a second. Taking both processing speed and occlusions into account, a four-camera tracking system is a good choice to deliver 3D target coordinates without too much time delay. However, as shown in Figure

5 - 5, only the back of the scalp in the phantom study can be covered by the four-camera geometry. Therefore if targets are distributed over a 3D object such as a full human scalp, four cameras are still not enough to fully cover all targets. In that case more cameras located in a full circle are considered to be appropriate for overcoming (not eliminate) the occlusion problem by enabling target intersections from at least more than two cameras.

5.3.5 Summary

The experiments discussed in this section are considered as the pre-study for target coordination in the clinic like environment. Both the phantom and caliper studies demonstrate that a tracking system with coloured target detection built from four off-the-shelf low cost Logitech C500 webcams is able to overcome difficult target illumination conditions particularly close to image saturation, to a certain extent and meanwhile deliver 3D coordination accuracies better than 0.4mm discrepancy, with VMS intersection results and caliper measurements as the reference. The coordination inaccuracy is mainly caused by a strong light reflection onto targets and the flatness of surface where targets are attached. Therefore it is suggested that where possible, highly light-reflective objects should be kept far away from targets being tracked and more cameras should be used to enable detection of a target from more directions. However, whilst the increase of equipment cost will be marginal, a lower refresh rate of the tracking system will result.

5.4 Optical topography studies

5.4.1 Introduction

Section 5.2 demonstrated that a setup of four cameras is not able to fully cover targets distributed over a human scalp. Therefore in the optical topography studies the tracking system was extended to include eight Logitech C500 webcams. This section discusses the target coordination of such a system for both the phantom and volunteer studies. The problems encountered for the optical topography applications together with the suggested solutions will also be delivered.

5.4.2 Method

The main focus of the phantom study was to investigate target coordination accuracy of the eight-camera tracking system assuming the phantom was fixed in position throughout the tracking process. The main focus of the volunteer studies were to detect any differential movements between the optical sensing pad and head by registering the continuously changing positions of pad and head during the study to their starting positions. Both the phantom and human studies were based on the camera calibration made three weeks prior to the optical topography study (Section 4.4.3), in order to provide a reference of time needed for recalibration of an on-going 3D measurement system (Section 3.3). The secondary focus of the optical topography study is to investigate the robustness in terms of target tracking of the developed method in the clinic like environment under no specific light control, where the light source was mainly from room light and sunlight through windows.

5.4.2.1 Experiment setup

Parameters of tracking system for optical topography studies are shown in Table 5 - 6.

Cameras used	4 Logitech C500 webcams with(PCI express) USB 3.0 ports + 4 Logitech C500 webcams with USB 2.0 ports
Interior/exterior orientation parameters	Obtained from camera calibration in Chapter 4
Wheel ring diameter	600mm
Vertical distance between ring and tested object (top head)	~370mm
Distance between camera and tested object (top head)	~550mm
Planar target diameter	4mm
Spherical ball bearing diameter	4.76mm
Frame refresh rate (FPS)	0.5 (potential to be higher)
Coloured target-tuning threshold	35
Light control	No
Light source	Room light + sun light
Human interaction during tracking	No

Table 5 - 6. *Parameters of the eight-camera tracking system for optical topography studies.*

The eight webcams were positioned in a full circle shape above the testing object for better target coverage in 3D space. The hardware image resolution provided by the Logitech C500 cameras is 1280 by 1024 (4μm square pixels). Software adjustments were made to these webcams (Cam 1, 2, 3, 8) through camera drivers to enhance the

target contrast while default settings were applied for those (Cam 4, 5, 6, 7) connected to USB 3.0 ports (Section 4.5.5.3).

Since the cameras had been pre-calibrated, the tracking process started immediately after the targets were attached to appropriate positions on the tested objects. The slowest part of the setup process was the ten minutes required to attach targets onto the subject at corresponding 10-20 positions (Section 2.2.3) after measuring the circumference of the head. Targets were also attached on top of optodes to provide features for optical sensing pad for tracking. Since the optical topography studies carried out in this research were mainly for verification of the tracking system, many more targets than required were used for 3D coordination accuracy assessment for the tracking system. As a result the setup for future optical topography studies should be much quicker.

The live views from all cameras were monitored during each functional study through a single PC monitor on a frame-by-frame basis. Every result including the 2D image coordinates of detected targets for each camera was displayed on the monitor. Output 2D correspondence matching data was also displayed on the monitor as a check to make sure the tracking system was acquiring the right data to compute the target 3D coordinates. These overlays have resulted in a lower refresh rate of the 3D coordination from the tracking system. A comparison of the tracking system used for this optical topography study with the one where unnecessary components were removed is provided in Figure 5 - 77 to give an idea how much acceleration the tracking system can have without a large amount of data printed out on screen.

Static images from the tracking system were captured during the tracking process for intersection post-processing in VMS in order to provide a reference to assess the accuracy of the output target 3D coordinates. However, one criterion for this accuracy assessment is that the images captured for VMS intersection processing were the same as the images used for target 3D coordinate computation by the tracking system, i.e. the targets need to be in the same 3D positions. Studies with the phantom provided an improved and more independent assessment of the 3D coordination accuracy of the developed system as compared with those with volunteers, in which there existed a

continuous head movement. Under these circumstances, the focus of volunteer studies was the investigation of differential movement between scalp and optical sensing pad during the study.

5.4.2.1.1 Phantom

A solid phantom (Figure 5 - 10) which is able to reliably provide optical change simulating brain activation (developed and built by Biomedical Optics Research Laboratory of UCL) was used in this study [Hebden *et al.*, 2008]. The simulation of the optical absorption changes from brain activities is made of thermochromic pigments containing both colorant and organic acid within a low melting point solvent. These materials are then mixed with the polyester resin within a plastic cylindrical tube (Figure 5 - 9). Since the thermochromic pigments undergo an optical change when its nominal activation temperature is reached, a heating power of about 1.2W is provided with a small 11Ω resister while the temperature can be simutaneously monitored by measuring the resistance of an embedded pre-calibrated 4.7kΩ bead thermistor (Figure 5 - 9). When the activation temperature is reached, the switch of the resister is opened a few seconds to prevent the damage which might be caused because of overheating. Each target containing such pigments was a cylinder with 8mm diameter and 8mm height which was designed to produce a signal of the minimum spatial dimensions that the OT system can detect [Hebden *et al.*, 2008].

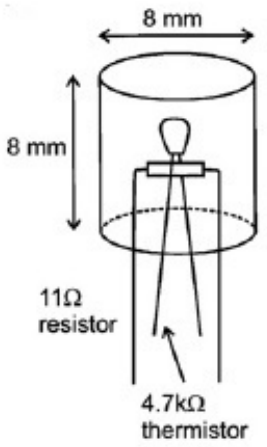


Figure 5 - 9. *Materials used to generate optical change within a solid phantom.*

(reproduced from Hebden et al. (2008))

Tests have shown that the transmittance decreases, i.e. undertaking optical absorption, at near-infrared wavelengths when the activation temperature of the thermochromic pigments is reached. However, it should be noted that the temperature distribution is unlikely to be even across the entire tube containing these pigments. Such distribution can also be different between heating and cooling under the same temperature [Hebden *et al.*, 2008].

The experimental setup used for the phantom study is shown in Figure 5 - 10. The monitor above the optical topography signal generator (Figure 5 - 10) was used to monitor the wavelength of each channel during the study. The monitor on the left was used to monitor the live tracking process. The phantom head was adjusted to match the height parameter from Table 5 - 6. A total of fifty two green coloured targets were used to deliver features of the head surface and the optical sensing pad.

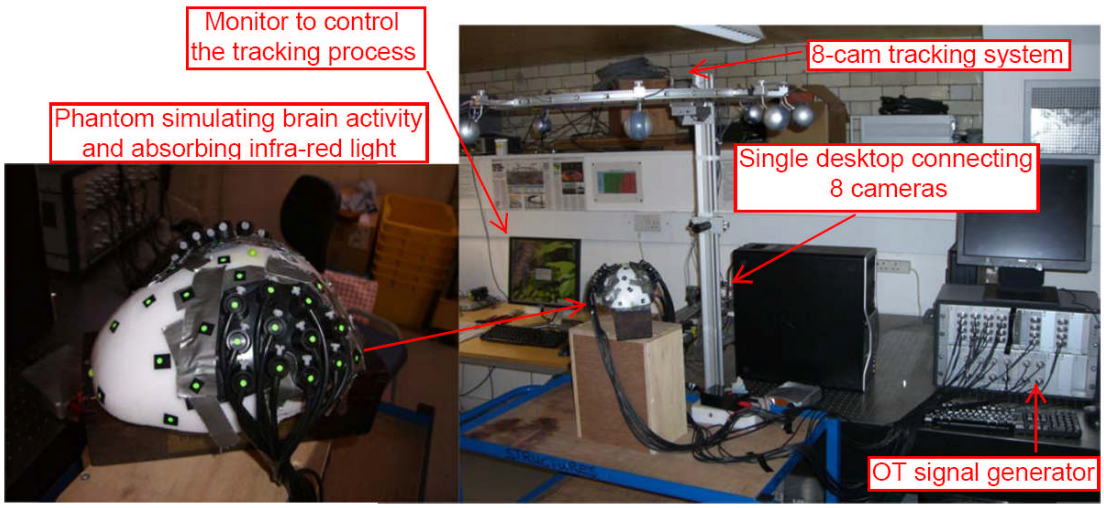


Figure 5 - 10. *Tracking system setup for the functional study of phantom.*

Three studies of regions representing the temporal left, temporal right and visual cortex were performed individually, where each of these areas is shown in Figure 5 –11. One target was attached on top of each optode for better characterisation of the sensing pad. There were eleven sensors on the sensing pad for both temporal left and right studies and thirteen sensors for the visual cortex study. Bony landmarks corresponding to human subjects were marked with spherical metal ball bearings (Figure 5 –11) in order to provide omni-directional targets.

Since the area of interest for each individual study is the most important part for 3D

coordination, these areas were oriented towards the four cameras connected to USB 3.0 ports (Figure 5 –11). Please note that the C500 webcams were not USB 3.0 systems but that the USB 3.0 bus allowed the cameras to perform much more reliably (Figure 4 – 20 and Figure 4 - 22) given the six independent USB chipsets connected to the desktop computer (Figure 4 – 23). Other targets were tracked by the other four cameras which were more sparsely distributed and connected through USB 2.0 ports.

The design of the tracking system can be referred back to Section 4.5.2. However, please note that important targets oriented towards Cam 4, 5, 6 and 7 (connected to USB 3.0 ports) were oriented directly towards the windows of the room. These targets were therefore subject to larger lighting variations due to sunlight changes compared to those oriented towards Cam 1, 2, 3 and 8 which were illuminated by the room lights.

The visual cortex study was performed between 2-2:15pm while a temporal left study and temporal right study were performed between 4:40-4:55pm and 5:15-5:30pm, respectively. One set of eight static images were captured by the tracking system at the beginning of each study (2:04pm for visual cortex; 4:42pm for temporal left and 5:17pm for temporal right) for comparative intersection post-processing in VMS. The ambient lighting varied significantly among three studies as well as within each study due to unstable summer weather (02/08/2010) where strong sunlight appeared and disappeared through the window from time to time. Variation of lighting conditions across studies is shown in Figure 5 –11.

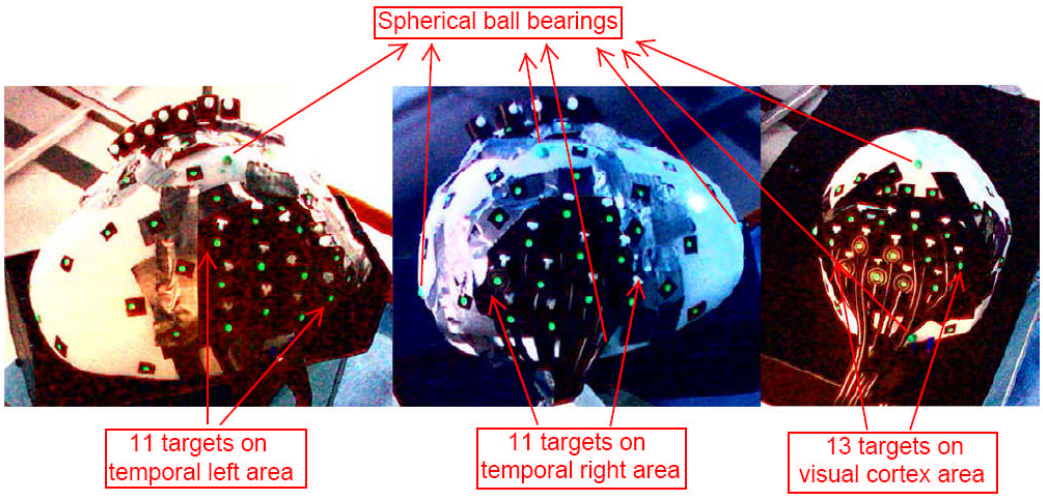


Figure 5 –11. *Areas of interest for three functional phantom studies.*

5.4.2.1.2 Human Subjects

The targets on both the optical sensing pad and the scalp of human subject one are shown in Figure 5 - 12. The areas of interest for this study were the temporal regions located on the upper left and right area on the head. As before, one target was attached on top of each optode (Figure 5 –11) for better characterisation of the sensing pad. Each sensing pad contains eleven such targeted sensors. As shown in Figure 5 - 12, bony landmarks were marked with spherical metal ball bearings painted in green to provide omni-directional targets and minimise view blockage by human hair.

Figure 5 - 12. *Planar and spherical targets used for human subject one.*

Figure 5 - 13 is a screenshot from the tracking system during the functional studies of both subjects. As mentioned in Section 4.5.5.3, appropriate adjustments had been made for cameras 1, 2, 3 and 8 connected to USB 2.0 ports prior to the study through camera driver, while no changes had been made for cameras 4 to 7 connected to USB 3.0 ports. Study of subject one was performed around 11:55am-12:15pm on 04/08/2010 and study of subject two was performed around 3:17-3:37pm on 05/08/2010.

Figure 5 - 13. *Screenshots of views by all eight C500 webcams during the tracking process. (Upper left: Cam 1; Lower right: Cam 8).*

5.4.3 Results

5.4.3.1 Phantom

3D target coordinates for a particular set of frames were regularly output as text files during the tracking process. The only requirement for the operator to perform such task was to press a particular button on the keyboard. Since it was assumed that there was no movement for the phantom study, one set of eight static images with 1280×1024 resolution setting were captured by the system at the beginning of the study. These images were later loaded into VMS software for target intersection processing in order to provide comparative 3D target locations with independent software. All quoted 3D coordinate discrepancies for the phantom studies (temporal left, temporal right and visual area) are between VMS and the tracking solution using the same sets of camera orientation parameters (Table 5 - 6) and target intersection algorithm (Section 3.5.2).

The only difference between these data sets is the image measurements. It should be note that the VMS processing was only based on the eight static images (from eight webcams) extracted from video at the beginning of each study. On the other hand, every set of image measurements of the tracking solution was based on the eight images (from eight webcams) captured at a particular time. The target positions were assumed to be static during the entire tracking process. However the lighting varied continuously such that the lighting condition for image measurements was different from frame to frame. The following 3D coordinate discrepancies compare the fast (fully automatic) image measurement against the off-line (both automatic and manual) image measurement.

5.4.3.1.1 Temporal left

The two different software solutions labeled targets with different ID numbers. In order to be able to compare which targets were measured during the rapid tracking process, target image measurements by each camera from a random frame set were selected and mapped onto the associated images captured at the beginning of the study for VMS processing (Figure 5 - 14 to Figure 5 - 21). It should be noted that while the target positions are assumed to be fixed between the following images and the selected frames, the lighting might vary significantly between these two sets of images.

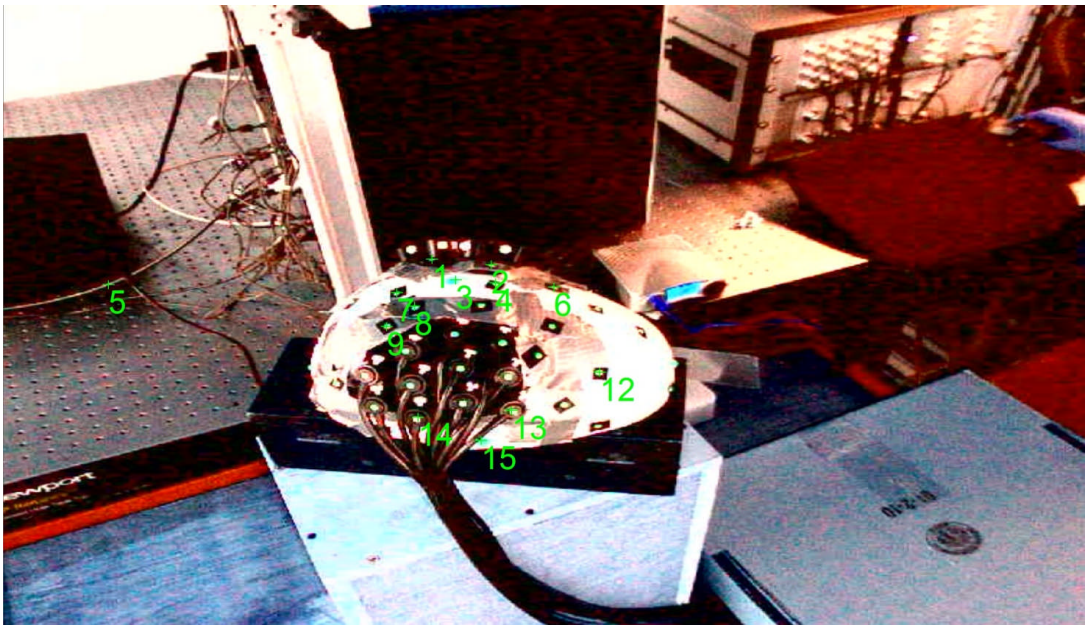


Figure 5 - 14. *Target image measurements by camera one based on a random frame during the live tracking process of temporal left study.*

Figure 5 - 15. *Target image measurements by camera two based on a random frame during the live tracking process of temporal left study.*

Figure 5 - 16. *Target image measurements by camera three based on a random frame during the live tracking process of temporal left study.*

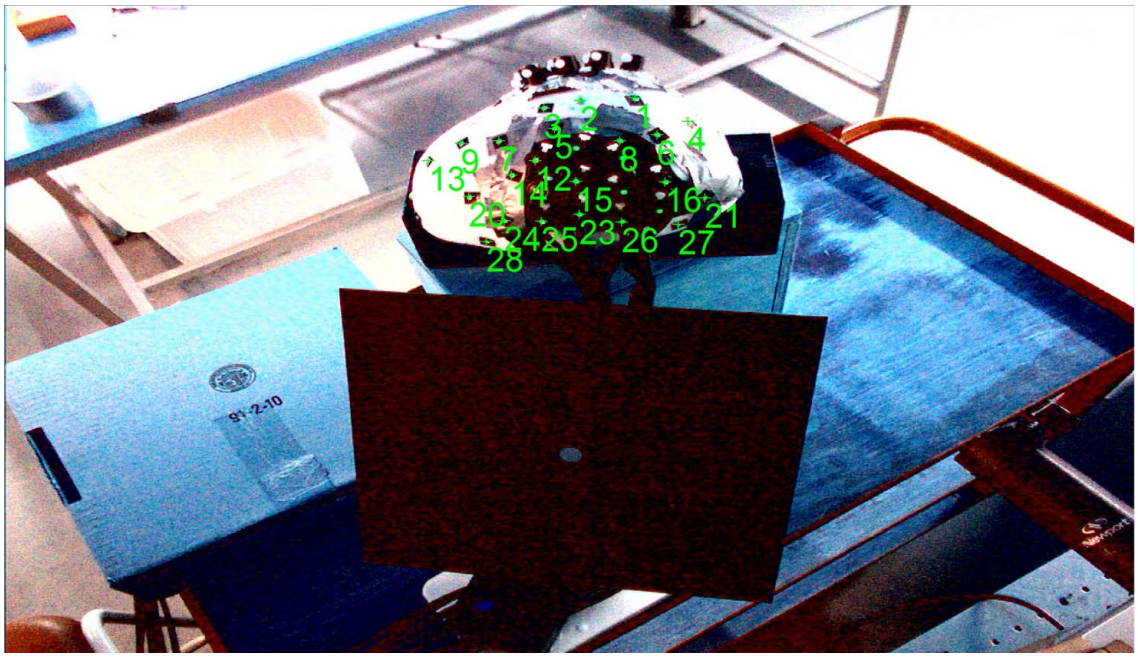


Figure 5 - 17. *Target image measurements by camera four based on a random frame during the live tracking process of temporal left study.*

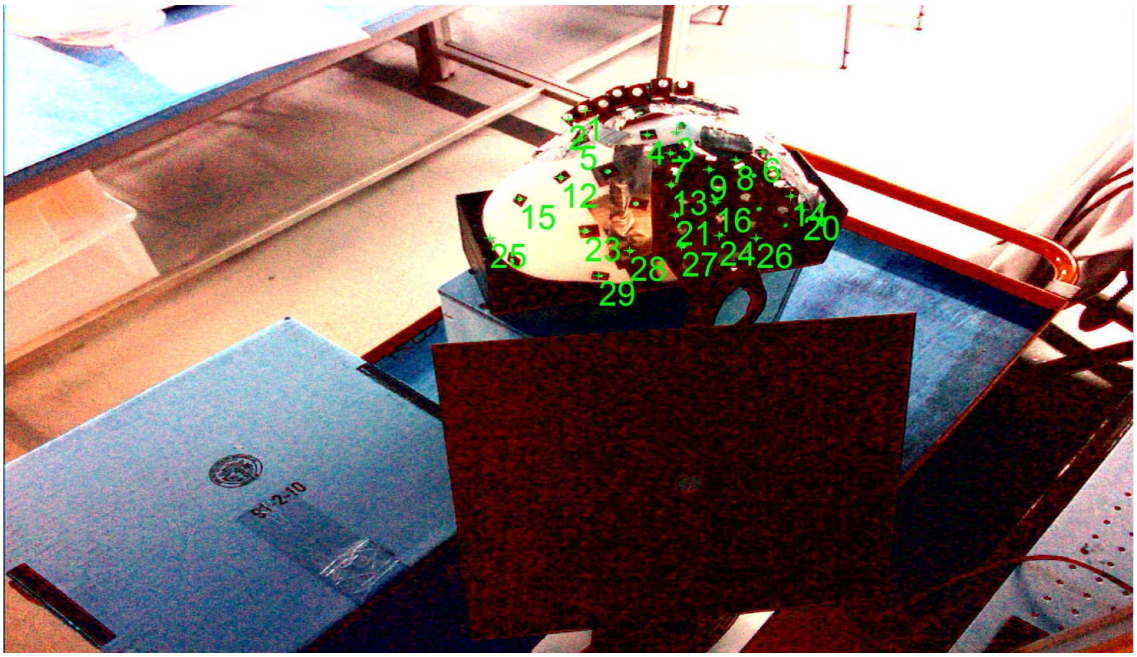


Figure 5 - 18. *Target image measurements by camera five based on a random frame during the live tracking process of temporal left study.*

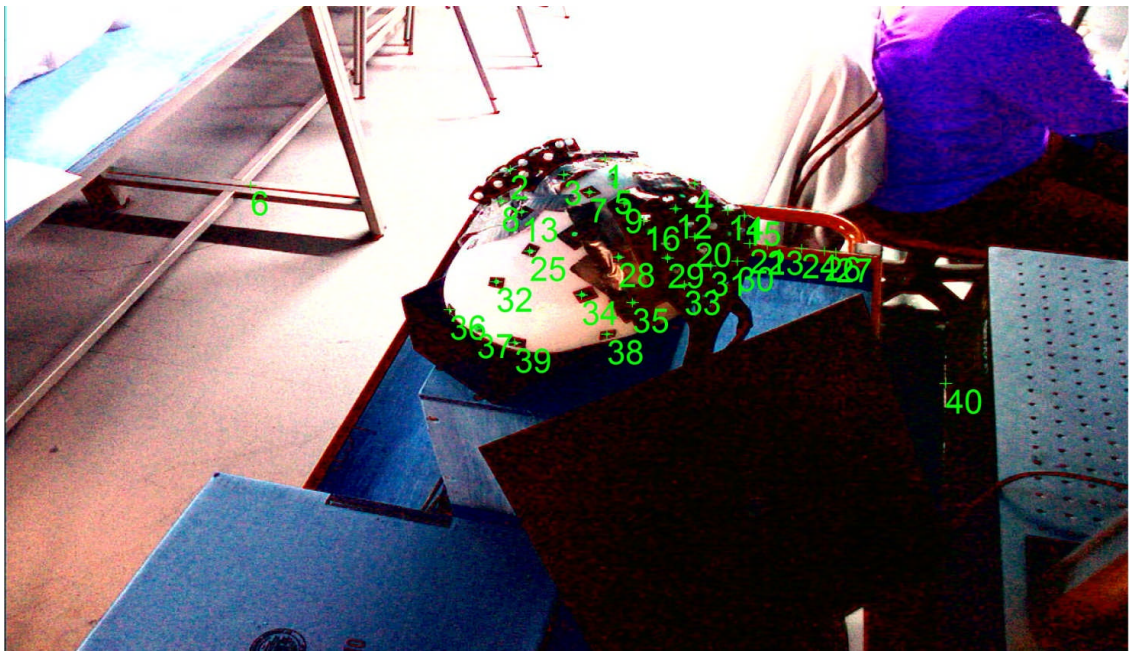


Figure 5 - 19. *Target image measurements by camera six based on a random frame during the live tracking process of temporal left study.*

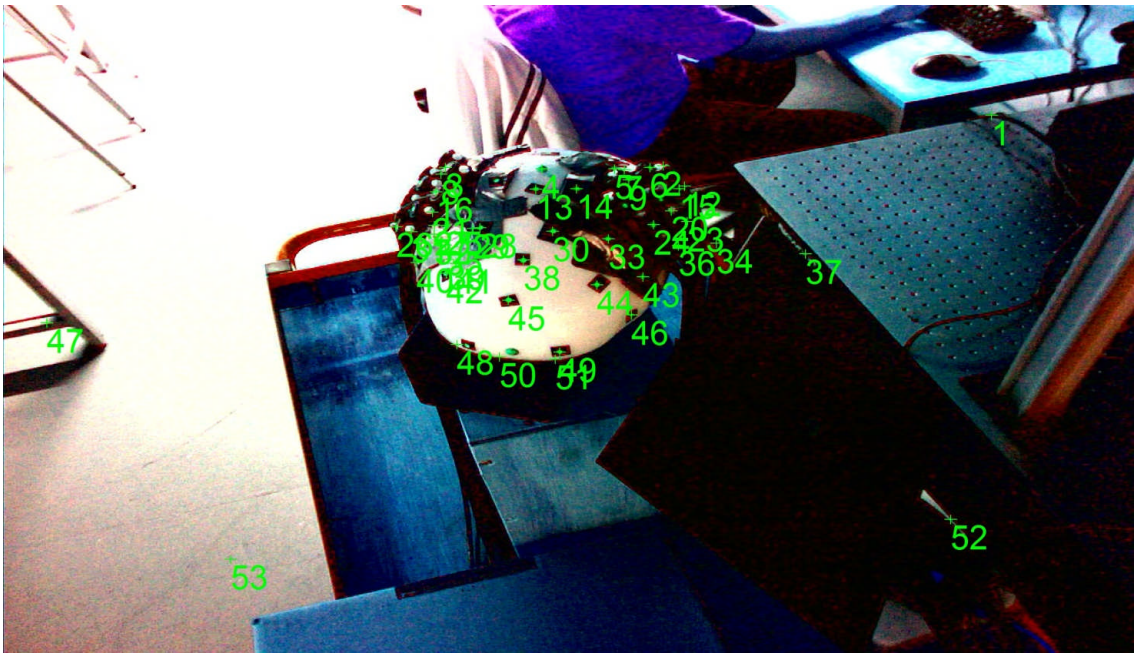


Figure 5 - 20. *Target image measurements by camera seven based on a random frame during the live tracking process of temporal left study.*

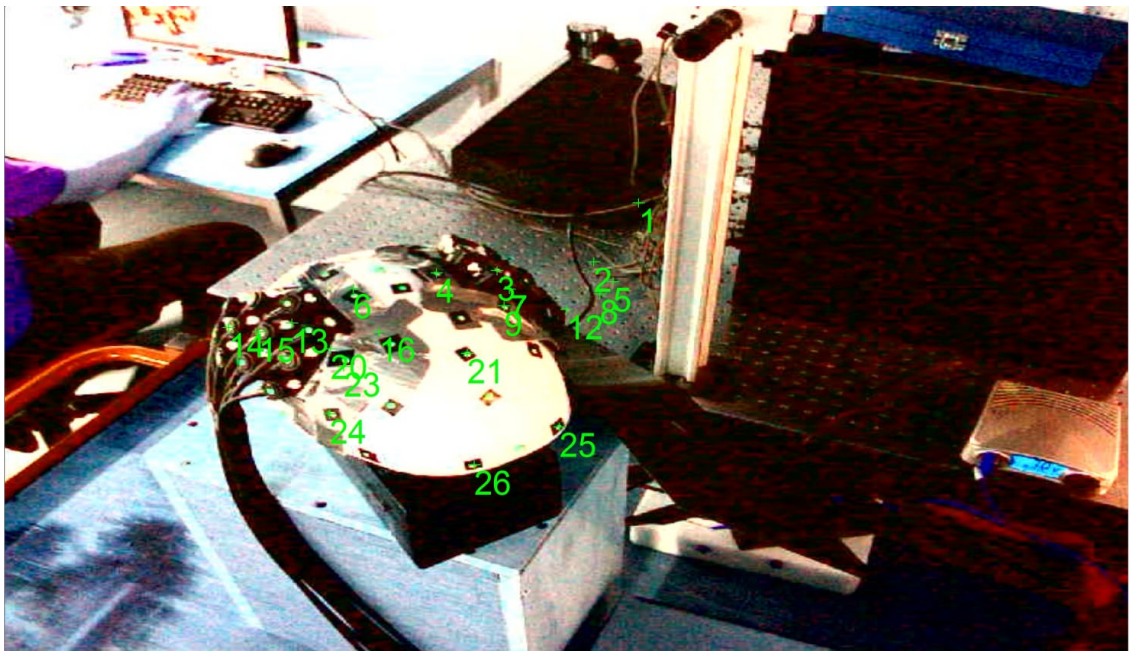


Figure 5 - 21. *Target image measurements by camera eight based on a random frame during the live tracking process of temporal left study.*

Five image sets were randomly selected out of a total of seventeen captured from the live tracking process. Table 5 - 7 summarizes the 3D coordinate discrepancies of all targets computed between data captured with the live system and single sets of snapshot images processed in VMS. The discrepancies are attributable to both target image detection and variations in illumination since the images used in VMS were collected only at the beginning of the study:

Frame no.	1	4	9	13	17
Total no. of targets	52	52	52	52	52
No. of targets coordinated by tracking system	30	31	29	31	28
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.20 Y: 0.31 Z: 0.48	X: 0.31 Y: 0.43 Z: 0.58	X: 0.31 Y: 0.39 Z: 0.53	X: 0.27 Y: 0.41 Z: 0.56	X: 0.29 Y: 0.40 Z: 0.49
Mean 3D discrepancies (mm)	0.60	0.78	0.73	0.74	0.70

Table 5 - 7. *Target 3D coordinate discrepancies of all targets between the live tracking system and fixed individual images with VMS intersection.*

Table 5 - 8 compares number of target image measurements from each individual camera between VMS processing and those successfully used for correspondence

captured from the running tracking system.

Camera	No. of target image measurement
1	27
2	22
3	14
4	26
5	26
6	22
7	24
8	23

	Frame 1	Frame 4	Frame 9	Frame 13	Frame 17
Cam 1	13	12	13	12	12
Cam 2	9	6	8	10	7
Cam 3	18	17	15	11	13
Cam 4	26	27	24	24	22
Cam 5	26	26	27	26	30
Cam 6	26	27	28	27	32
Cam 7	25	29	26	30	28
Cam 8	14	12	13	12	14

Table 5 - 8. *Target image measurements for temporal left study.*

Left: fixed individual images processed with VMS;

Right: those successfully used for correspondence from the live tracking system.

Due to the poor and unreliable image qualities from those cameras connected to USB 2.0 ports, target rejection, controlled by the colour tuning threshold, was lowered from 45 to 35 (Table 5 - 6, together with software enhancement at the time of optical topography tests) to allow target image measurements from all eight cameras rather than just the four connected to USB 3.0 ports (Section 4.5.5.3). A number of image measurements not used in the correspondence from the running tracking system are categorised in Table 5 - 9.

	Frame 1			Frame 4			Frame 9			Frame 13			Frame 17		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Cam 1	2	2	0	0	3	3	4	0	0	1	2	1	1	0	1
Cam 2	0	2	2	1	2	1	2	1	1	0	3	0	3	0	1
Cam 3	0	1	0	1	2	3	3	0	2	0	2	5	1	2	2
Cam 4	0	2	0	0	0	0	2	1	1	0	0	1	1	2	0
Cam 5	0	0	2	1	0	2	0	1	1	0	0	3	1	1	2
Cam 6	0	1	4	2	2	4	3	0	6	0	2	6	1	0	6
Cam 7	0	0	18	1	0	15	2	0	12	0	3	10	2	0	10
Cam 8	1	0	13	0	0	14	0	1	10	0	2	8	0	2	10

Table 5 - 9. *Target image measurements not used in the correspondence from the live tracking frames. A: Correspondence failure; B: Insufficient measurements; C: Not a target.*

For the temporal left study, the targets on the sensing pad covering this area are considered to be the most important in terms of image registration and thereafter functional localisation. As previously described, the pad above the temporal left area

was oriented towards those cameras connected to the USB 3.0 ports for reliable image quality in terms of target detection. Table 5 - 10 summarizes the 3D coordinate discrepancies of these targets computed between data captured with the live system and single sets of snapshot images processed in VMS:

Frame no.	1	4	9	13	17
Total no. of targets on pad	11	11	11	11	11
No. of targets coordinated by tracking system	11	10	9	10	10
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.16 Y: 0.30 Z: 0.35	X: 0.18 Y: 0.44 Z: 0.45	X: 0.23 Y: 0.42 Z: 0.39	X: 0.14 Y: 0.43 Z: 0.40	X: 0.22 Y: 0.43 Z: 0.30
Mean 3D discrepancies (mm)	0.48	0.65	0.62	0.60	0.57

Table 5 - 10. *3D coordinate discrepancies of targets on sensing pad above temporal left area between the live tracking system and fixed individual images with VMS intersection.*

Image enhancement through driver was applied for the four webcams connected to USB 2.0 ports in order to compensate their less reliable image qualities (section 4.5.5.3). The 3D mean discrepancies for the targets on the sensing pad above the temporal left area are compared with the 3D mean discrepancies for all targets (Figure 5 - 22).

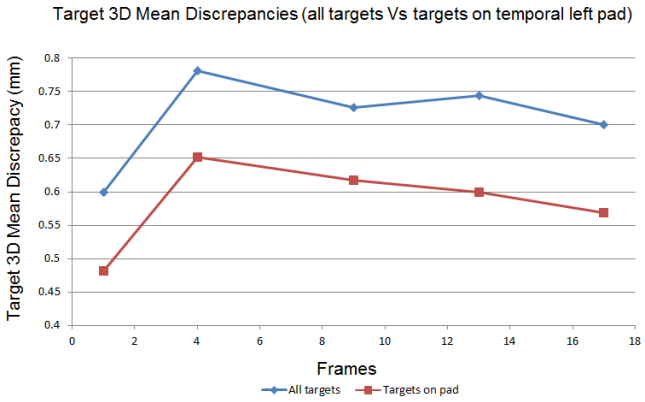


Figure 5 - 22. *Coordination discrepancies between all targets and targets on temporal left pad from the live tracking system.*

In terms of target tracking consistency across frames, Figure 5 - 23 compares the tracking results between more reliable image qualities (delivered from Cam 4, 5, 6, 7 connected to USB 3.0 ports) and those less reliable ones (delivered by Cam 1, 2, 3, 8 connected to USB 2.0 ports).

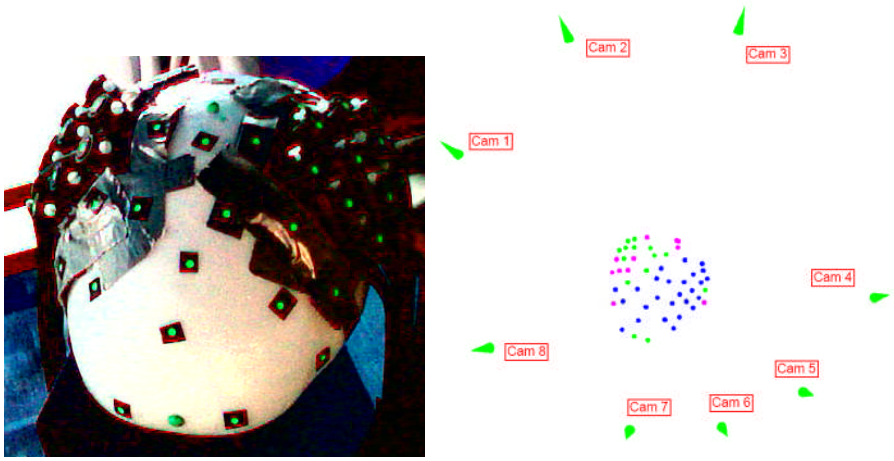


Figure 5 - 23. *Target tracking consistency for temporal left study. Left: Targets on phantom; Right: Relative positions of targets and cameras for tracking solution. (Blue dots: Targets consistently coordinated by the tracking system; Green dots: Targets sometimes disappeared from the coordination; Pink dots: Targets consistently not coordinated by the tracking system)*

For those consistently coordinated targets from the tracking solution (blue dots in Figure 5 - 23), Figure 5 - 24 further compares the number of camera rays used to intersect these targets between VMS processing and the selected frames from tracking solution.

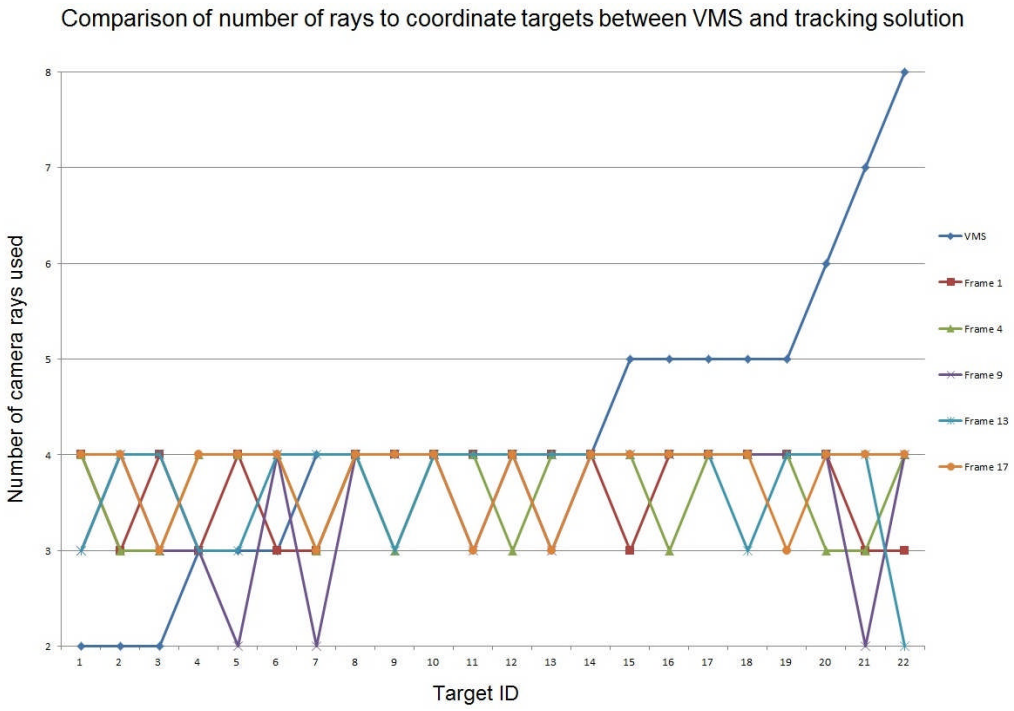


Figure 5 - 24. *Comparison of number of camera rays used to intersect those consistently tracked targets between fixed individual images with VMS intersection and the live tracking frames.*

Since most of the targets coordinated by the tracking system were intersected by either

three or four camera rays, it is worth checking if there are any systematic errors in 3D target coordination by looking into the discrepancies caused by each combination of rays. Figure 5 - 25 shows a typical result of 3D target location discrepancies between the tracking system intersected by three camera rays and VMS intersection while Figure 5 - 26 shows those intersected by four camera rays. In these tests the mean 3D discrepancy for three rays was 0.90mm whilst that for four rays was 0.57mm representing a less extreme example of the 3D discrepancy between using two and six cameras described in Section 5.1.

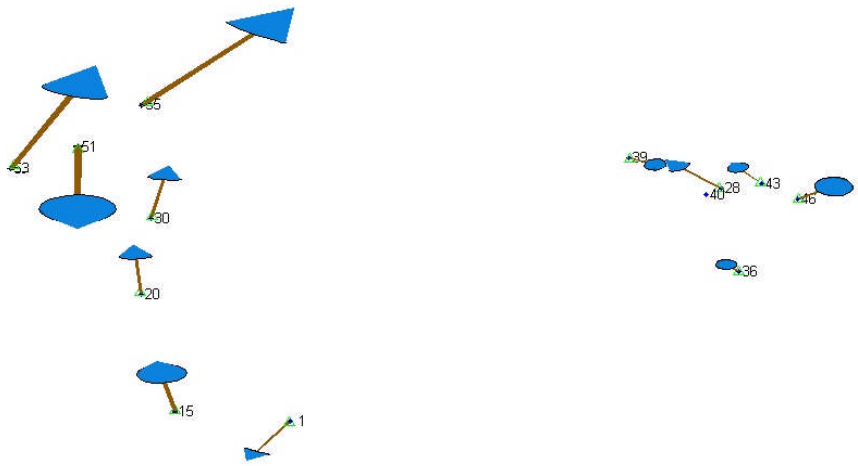


Figure 5 - 25. *Target 3D location discrepancies between tracking system (intersected by three camera rays of the 9th frame) and VMS intersection for temporal left study.*
(Mean discrepancies: X: 0.36mm; Y: 0.43mm; Z: 0.71mm; discrepancy vectors × 50 for visualisation)

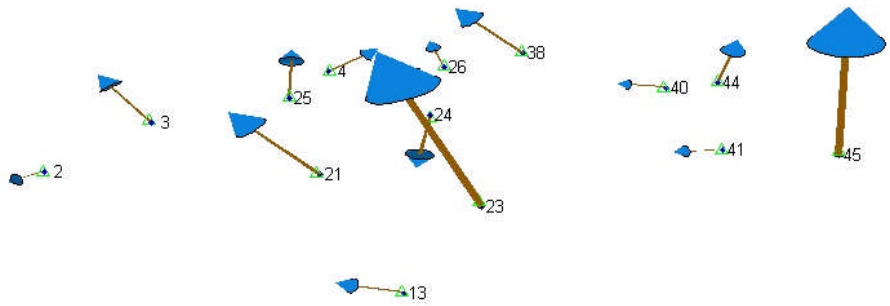


Figure 5 - 26. *Target 3D location discrepancies between tracking system (intersected by four camera rays of the 9th frame) and VMS intersection for temporal left study.*
(Mean discrepancies: X: 0.25mm; Y: 0.36mm; Z: 0.36mm; discrepancy vectors × 50 for visualisation)

5.4.3.1.2 Temporal right

As before, target image measurements from a random frame set captured by each camera during the sequential tracking process were selected. These data were mapped onto the associated images captured at the beginning of the study and used for VMS processing (Figure 5 - 27 to Figure 5 - 34).

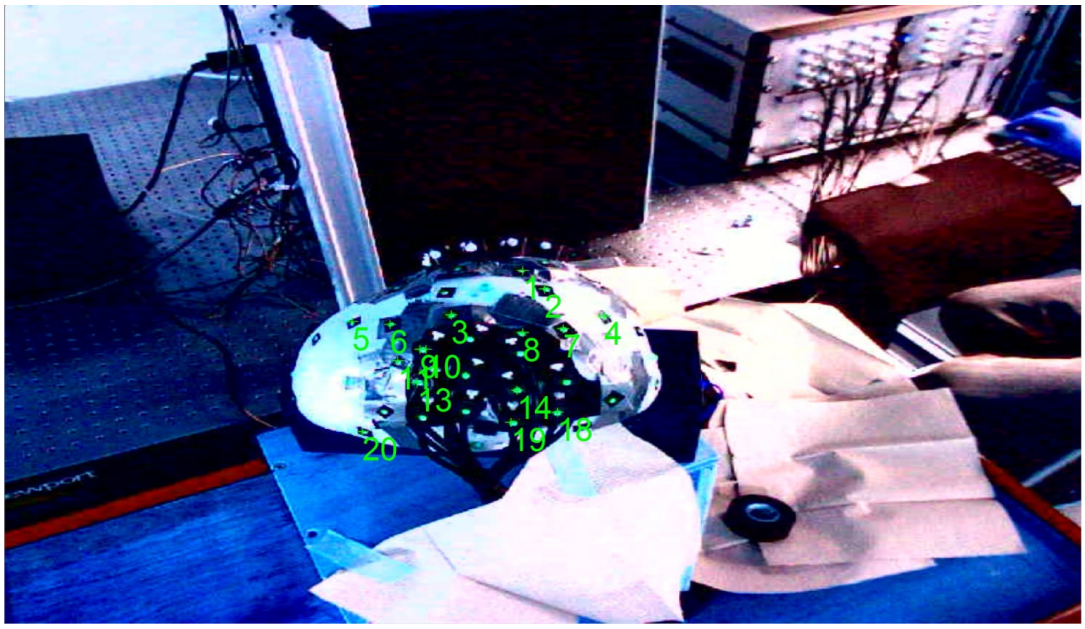


Figure 5 - 27. *Target image measurements by camera one based on a random frame during the live tracking process of temporal right study.*

Figure 5 - 28. *Target image measurements by camera two based on a random frame during the live tracking process of temporal right study.*

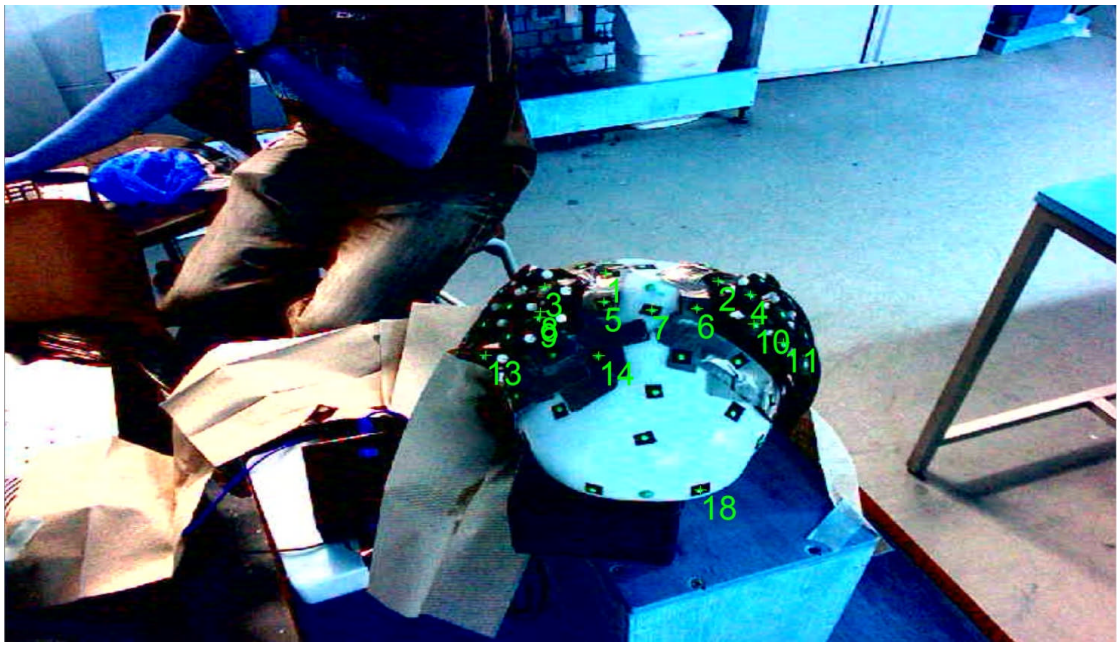


Figure 5 - 29. *Target image measurements by camera three based on a random frame during the live tracking process of temporal right study.*

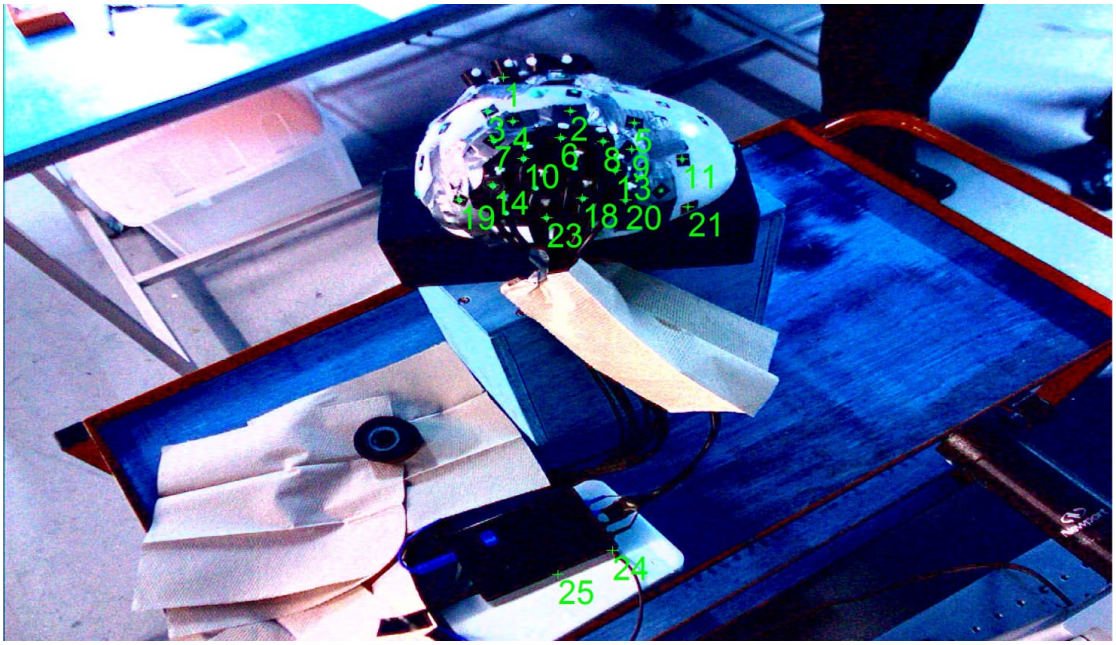


Figure 5 - 30. *Target image measurements by camera four based on a random frame during the live tracking process of temporal right study.*

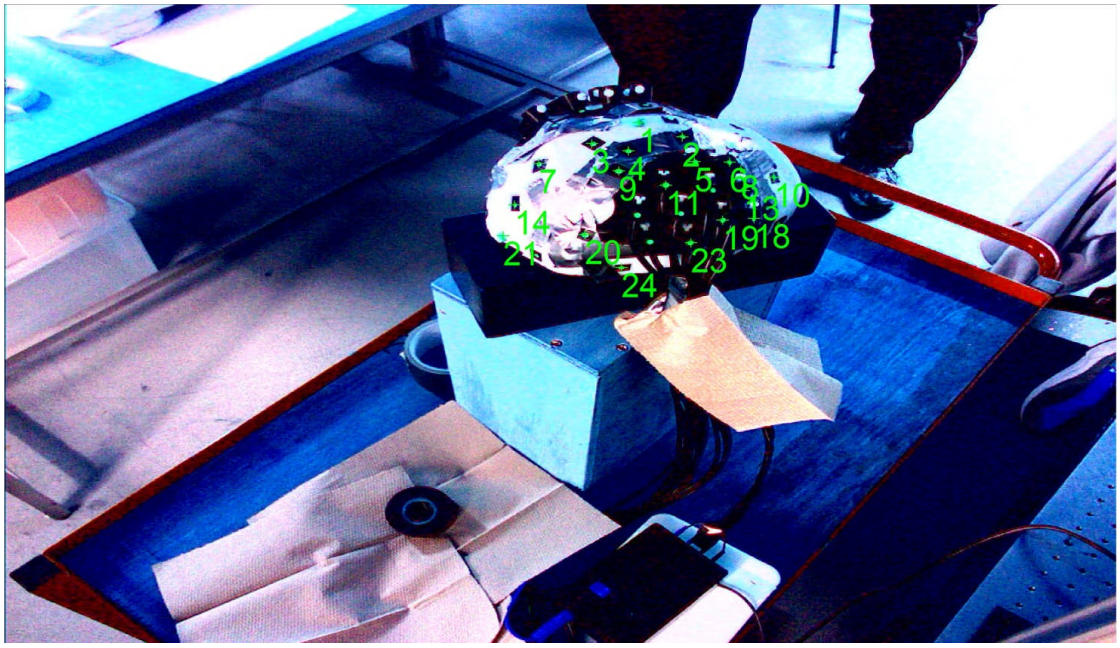


Figure 5 - 31. *Target image measurements by camera five based on a random frame during the live tracking process of temporal right study.*

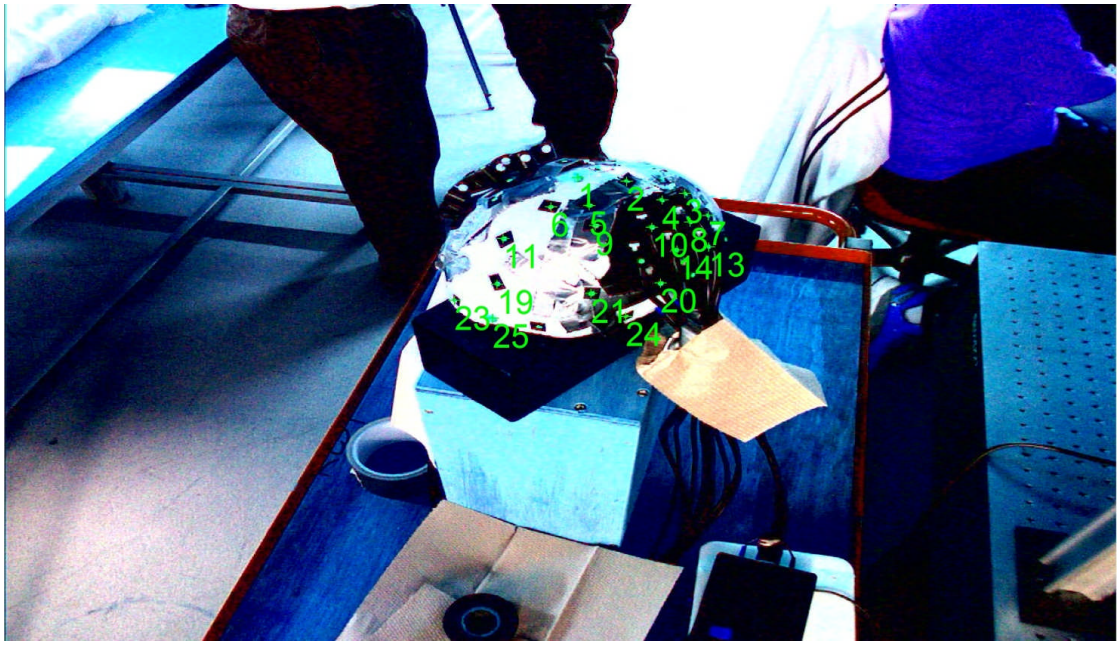


Figure 5 - 32. *Target image measurements by camera six based on a random frame during the live tracking process of temporal right study.*

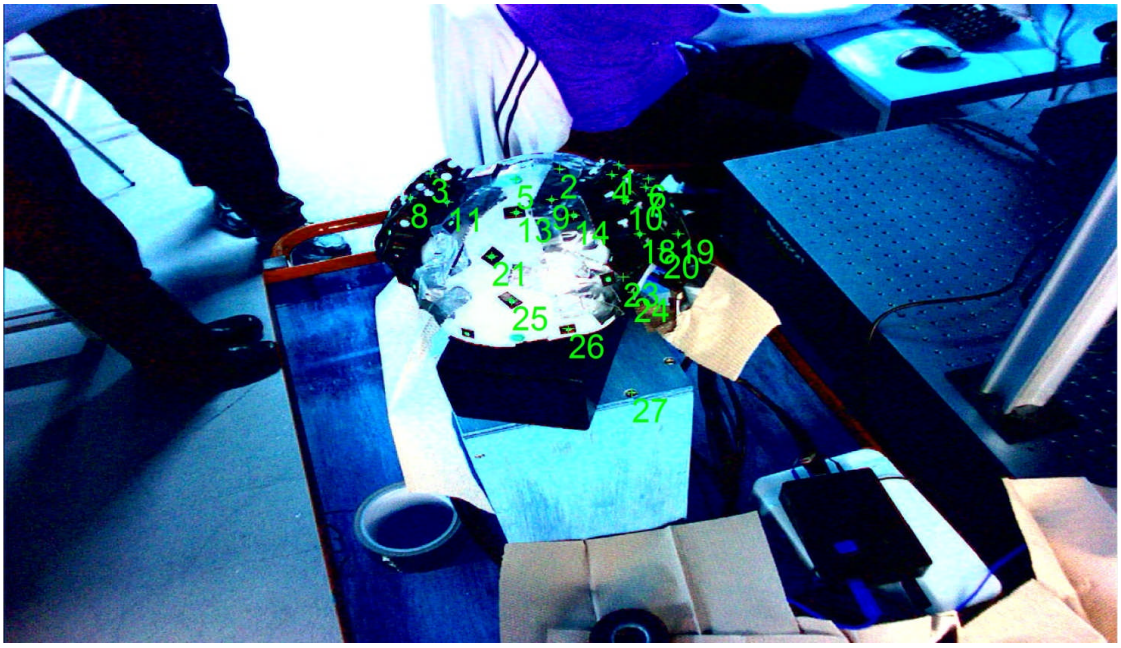


Figure 5 - 33. *Target image measurements by camera seven based on a random frame during the live tracking process of temporal right study.*

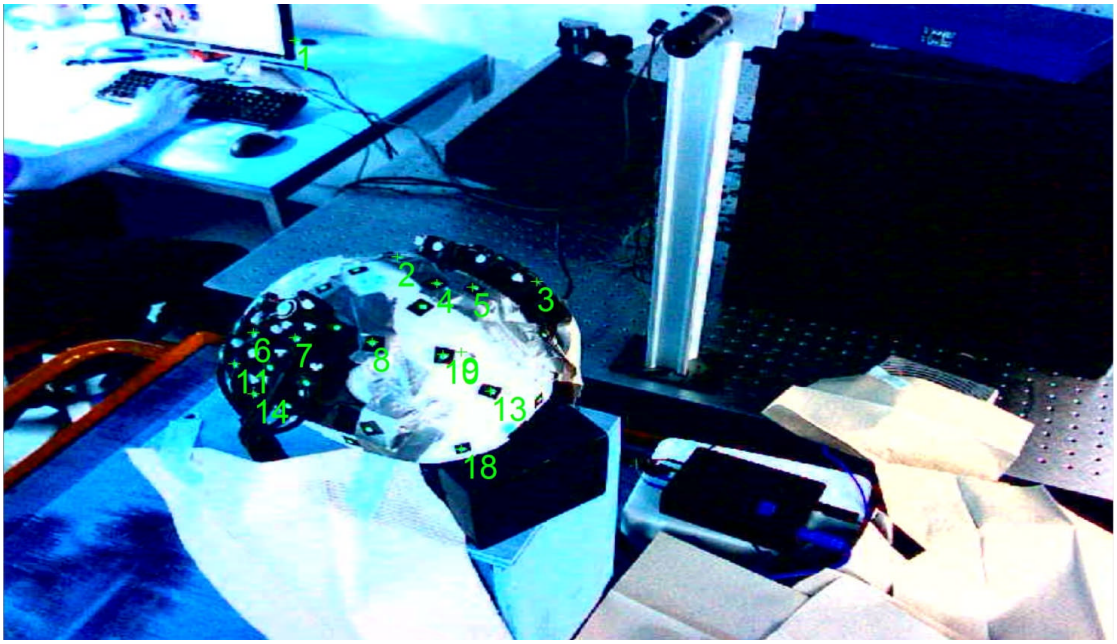


Figure 5 - 34. *Target image measurements by camera eight based on a random frame during the live tracking process of temporal right study.*

As for the temporal right study, five sets of frames were randomly selected out of a total of twelve frames captured from the live tracking process. Table 5 - 11 summarizes the 3D coordinate discrepancies of all targets computed between data captured with the live system and single sets of snapshot images processed in VMS. Note that the discrepancies are attributable to both target image detection and variations in illumination since the images used in VMS were collected only at the beginning of the study:

Frame no.	1	3	6	9	12
Total no. of targets	52	52	52	52	52
No. of targets coordinated by tracking system	29	29	25	32	28
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.29 Y: 0.36 Z: 0.79	X: 0.29 Y: 0.32 Z: 0.69	X: 0.23 Y: 0.33 Z: 0.51	X: 0.25 Y: 0.32 Z: 0.56	X: 0.26 Y: 0.30 Z: 0.60
Mean 3D discrepancies (mm)	0.91	0.81	0.65	0.69	0.72

Table 5 - 11. *Target 3D coordinate discrepancies of all targets between the live tracking system and fixed individual images with VMS intersection.*

To account for the target coordinate discrepancy between VMS processing and the tracking solution, Table 5 - 12 compares number of target image measurements from each individual camera between VMS processing and those successfully used for correspondence captured from the running tracking system.

Camera	No. of target image measurement
1	30
2	23
3	23
4	26
5	25
6	23
7	21
8	18

	Frame 1	Frame 3	Frame 6	Frame 9	Frame 12
Cam 1	10	11	10	11	13
Cam 2	8	9	8	10	9
Cam 3	20	18	14	18	15
Cam 4	17	21	18	23	21
Cam 5	22	25	19	21	23
Cam 6	19	22	19	21	18
Cam 7	20	22	20	22	19
Cam 8	12	9	9	9	10

Table 5 - 12. *Target image measurements for temporal right study.*

Left: fixed individual images processed with VMS;

Right: those successfully used for correspondence from the live tracking system.

Table 5 - 13 categorises image measurements not used in the correspondence from the running tracking system.

	Frame 1			Frame 3			Frame 6			Frame 9			Frame 12		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Cam 1	0	1	0	1	4	0	0	2	0	0	2	0	3	2	0
Cam 2	0	2	1	0	2	6	0	5	4	1	4	0	0	3	2
Cam 3	1	1	1	0	0	1	1	2	2	1	1	2	2	4	1
Cam 4	1	0	0	1	0	1	1	0	1	0	0	0	1	0	2
Cam 5	4	0	1	2	0	3	2	0	1	0	1	0	1	0	3
Cam 6	4	0	1	3	0	4	3	0	1	0	0	0	2	1	1
Cam 7	7	1	4	1	1	3	2	2	6	2	1	3	3	1	2
Cam 8	2	0	2	1	2	4	1	5	1	2	0	5	2	1	3

Table 5 - 13. *Target image measurements not used in the correspondence from the live tracking frames. A: Correspondence failure; B: Insufficient measurements; C: Not a target.*

As previously described, the pad above the temporal right area was oriented towards those cameras connected to the USB 3.0 ports. Table 5 - 14 summarizes the 3D coordinate discrepancies of these targets computed between data captured with the live system and single sets of snapshot images processed in VMS:

Frame no.	1	3	6	9	12
Total no. of targets on pad	11	11	11	11	11
No. of targets coordinated by tracking system	9	9	10	9	11
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.24 Y: 0.46 Z: 0.63	X: 0.15 Y: 0.26 Z: 0.54	X: 0.19 Y: 0.18 Z: 0.40	X: 0.16 Y: 0.16 Z: 0.47	X: 0.13 Y: 0.16 Z: 0.42
Mean 3D discrepancies (mm)	0.82	0.61	0.48	0.52	0.47

Table 5 - 14. *3D coordinate discrepancies of targets on sensing pad above temporal right area between the live tracking system and fixed individual images with VMS intersection.*

The discrepancies for those targets on the sensing pad above the temporal right area are compared with which for all targets, as shown in Figure 5 - 35.

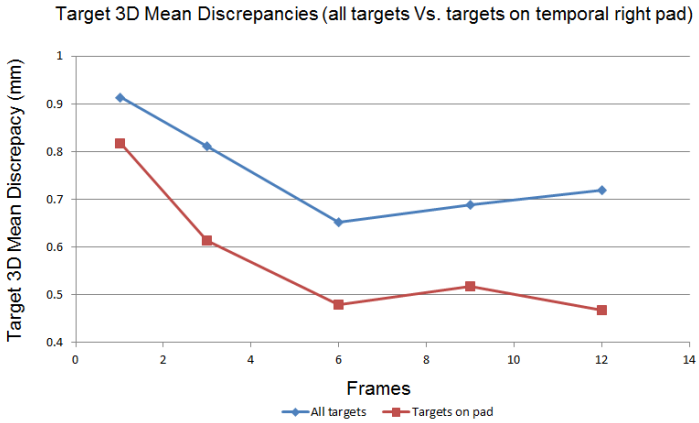


Figure 5 - 35. *Coordination discrepancies of all targets and targets on temporal right pad from the live tracking system.*

In terms of target tracking consistency across frames, Figure 5 - 36 compares the tracking results from the more reliable image quality delivered from Cam 4, 5, 6, 7 connected to USB 3.0 ports, and those less reliable ones delivered by Cam 1, 2, 3, 8 connected to USB 2.0 ports.

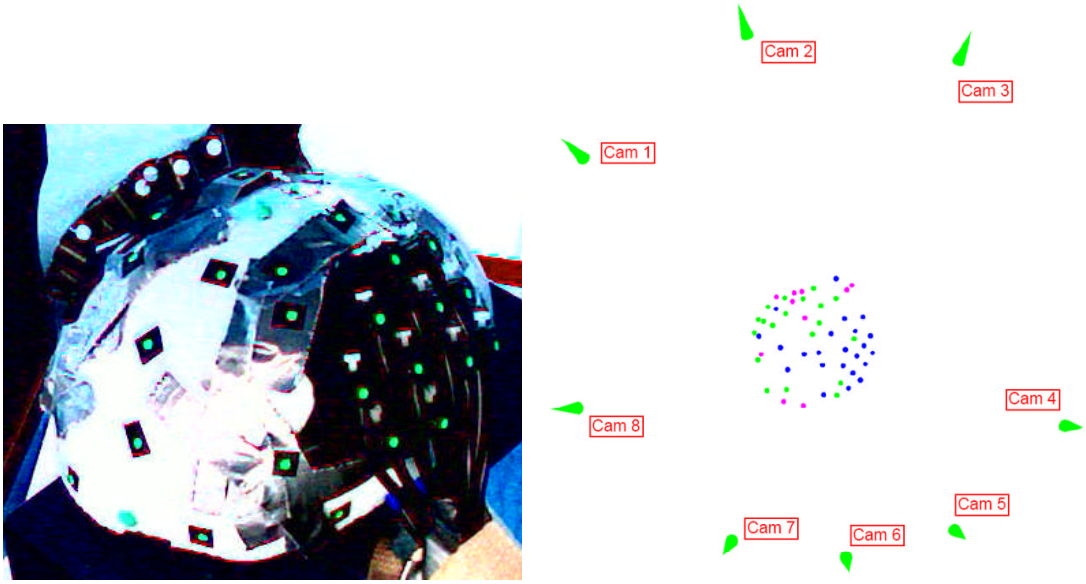


Figure 5 - 36. *Target tracking consistency for temporal right study.*

Left: Targets on phantom; Right: Relative positions of targets and cameras for tracking solution. (Blue dots: Targets consistently coordinated by the tracking system; Green dots: Targets sometimes disappeared from the coordination; Pink dots: Targets consistently not coordinated by the tracking system)

For targets consistently coordinated with the tracking solution (blue dots in Figure 5 - 36), Figure 5 - 37 further compares the number of camera rays used to intersect these targets between VMS processing and the selected frames from tracking solution.

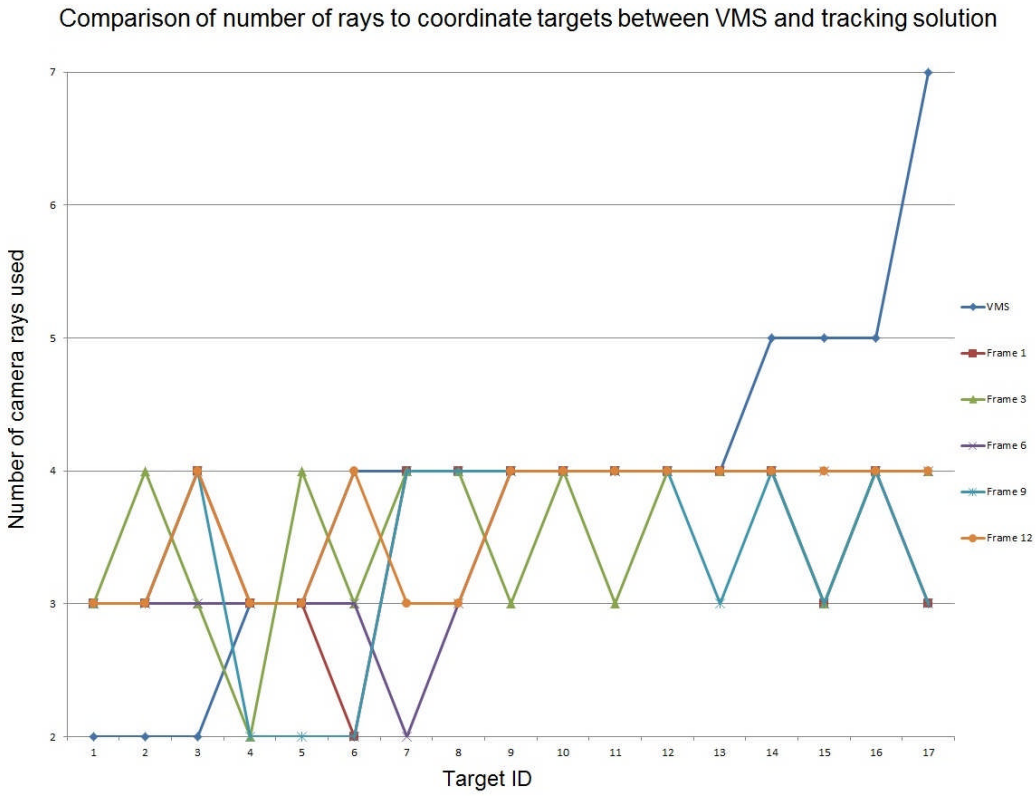


Figure 5 - 37. *Comparison of number of camera rays used to intersect those consistently tracked targets between fixed individual images with VMS intersection and the live tracking frames.*

Figure 5 - 38 shows a typical set of 3D target location discrepancies between the tracking system intersected by three camera rays and VMS intersection. Figure 5 - 39 shows a similar set of data, but for those intersected by four camera rays. In these tests the mean 3D discrepancy for three rays was 0.68mm whilst that for four rays was 0.50mm representing a less extreme example of the 3D discrepancy between using two and six cameras described in Section 5.1.

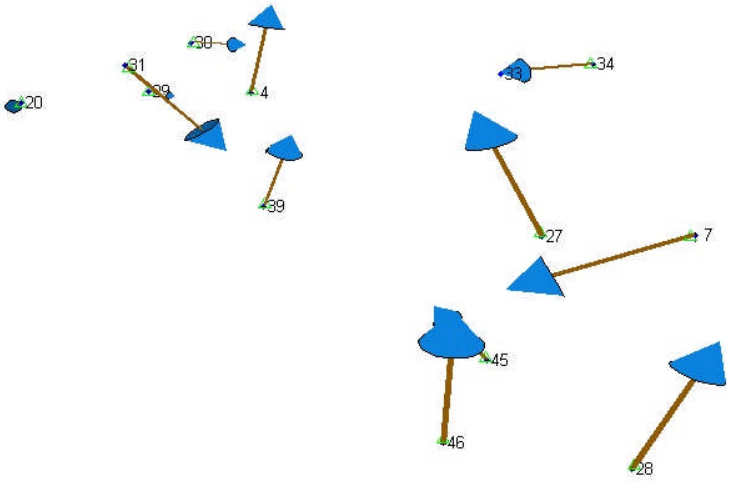


Figure 5 - 38. *Target 3D location discrepancies between tracking system (intersected by three camera rays of the 6th frame) and VMS intersection for temporal right study (Mean discrepancies: X: 0.18mm; Y: 0.35mm; Z: 0.55mm; discrepancy vectors × 50 for visualisation).*

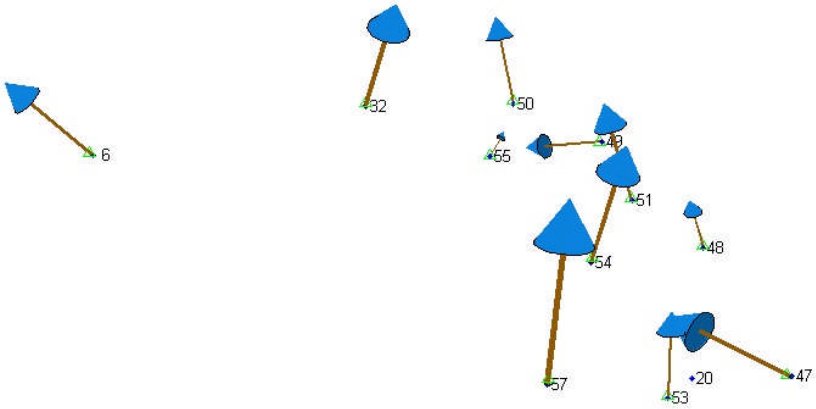


Figure 5 - 39. *Target 3D location discrepancies between tracking system (intersected by four camera rays of the 6th frame) and VMS intersection for temporal right study (Mean discrepancies: X: 0.20mm; Y: 0.20mm; Z: 0.41mm; discrepancy vectors × 50 for visualisation).*

5.4.3.1.3 Visual cortex

As before, target image measurements by each camera from a random frame set during the entire tracking process were selected and mapped onto the associated images captured at the beginning of the study and used for VMS processing (Figure 5 - 40 to Figure 5 - 47).

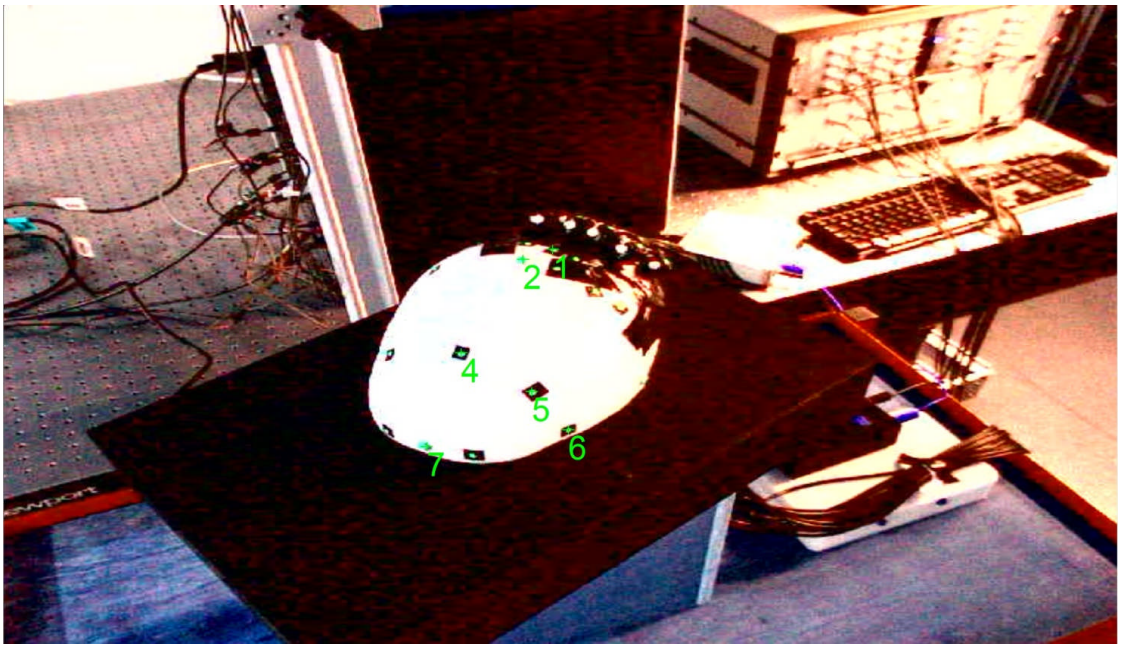


Figure 5 - 40. *Target image measurements by camera one based on a random frame during the live tracking process of visual cortex study.*

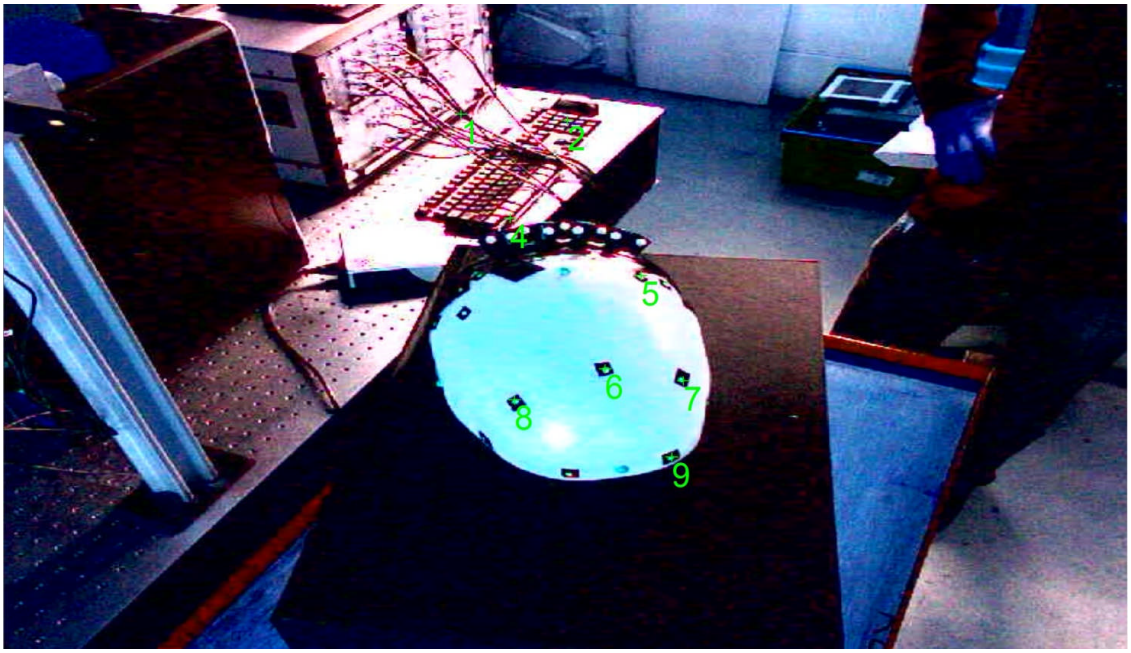


Figure 5 - 41. *Target image measurements by camera two based on a random frame during the live tracking process of visual cortex study.*

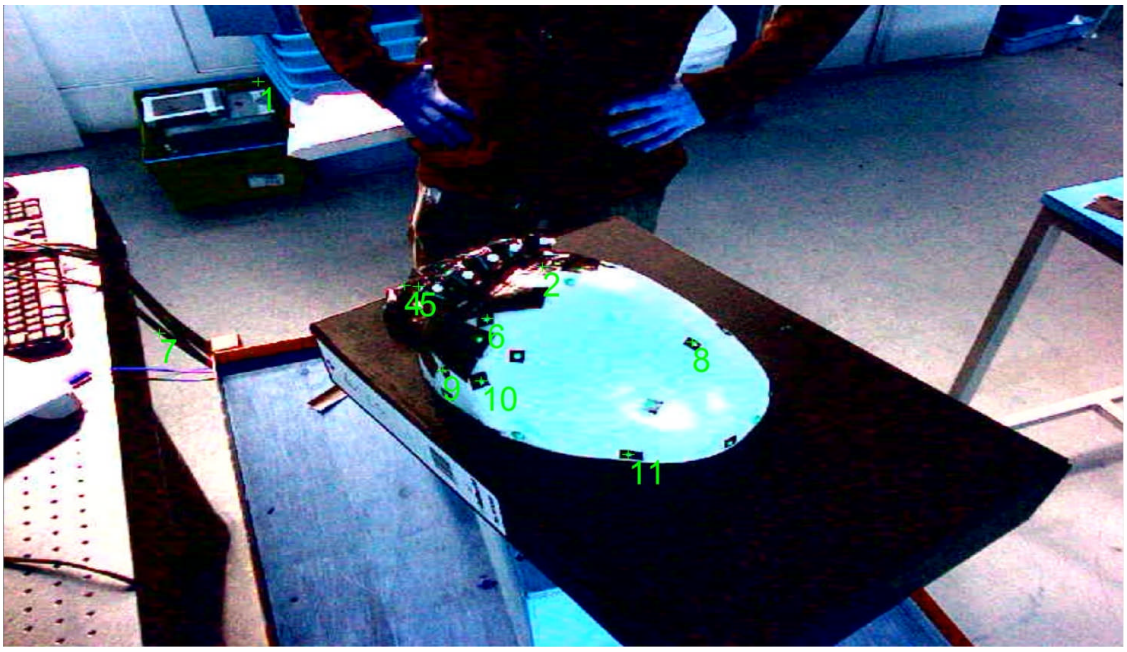


Figure 5 - 42. *Target image measurements by camera three based on a random frame during the live tracking process of visual cortex study.*

Figure 5 - 43. *Target image measurements by camera four based on a random frame during the live tracking process of visual cortex study.*

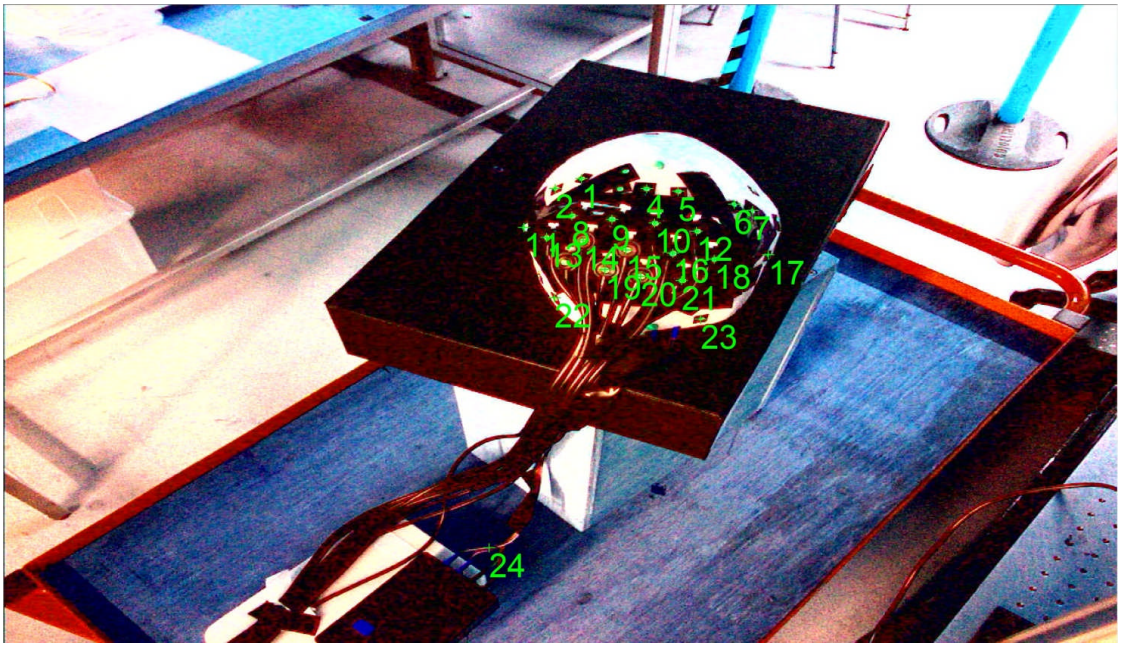


Figure 5 - 44. *Target image measurements by camera five based on a random frame during the live tracking process of visual cortex study.*

Figure 5 - 45. *Target image measurements by camera six based on a random frame during the live tracking process of visual cortex study.*

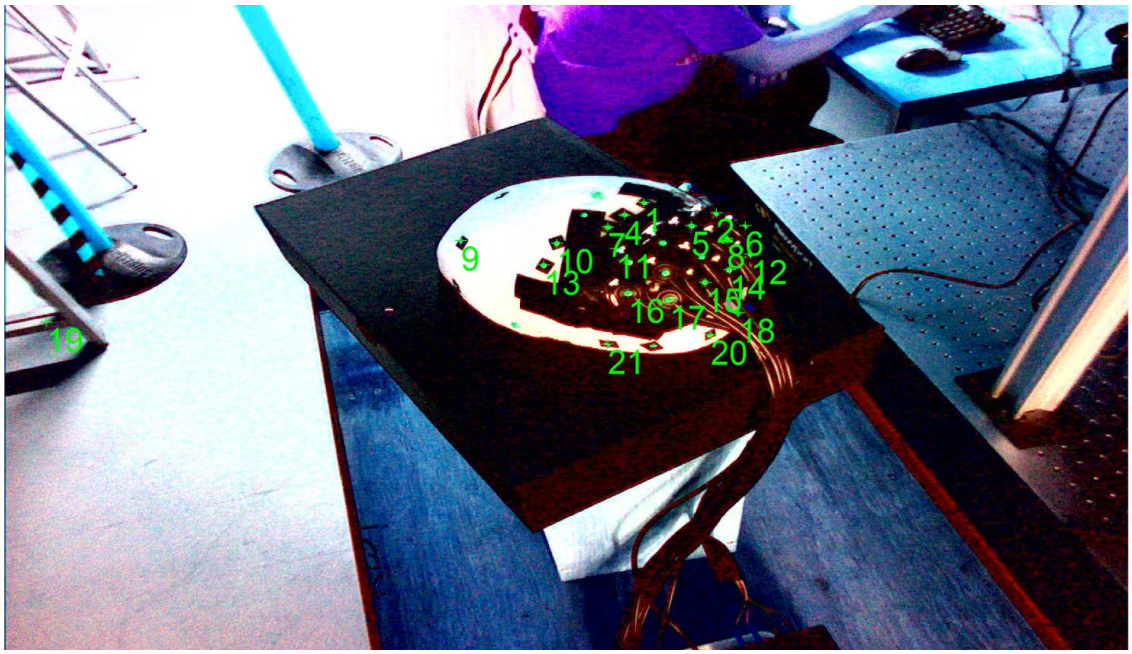


Figure 5 - 46. *Target image measurements by camera seven based on a random frame during the live tracking process of visual cortex study.*

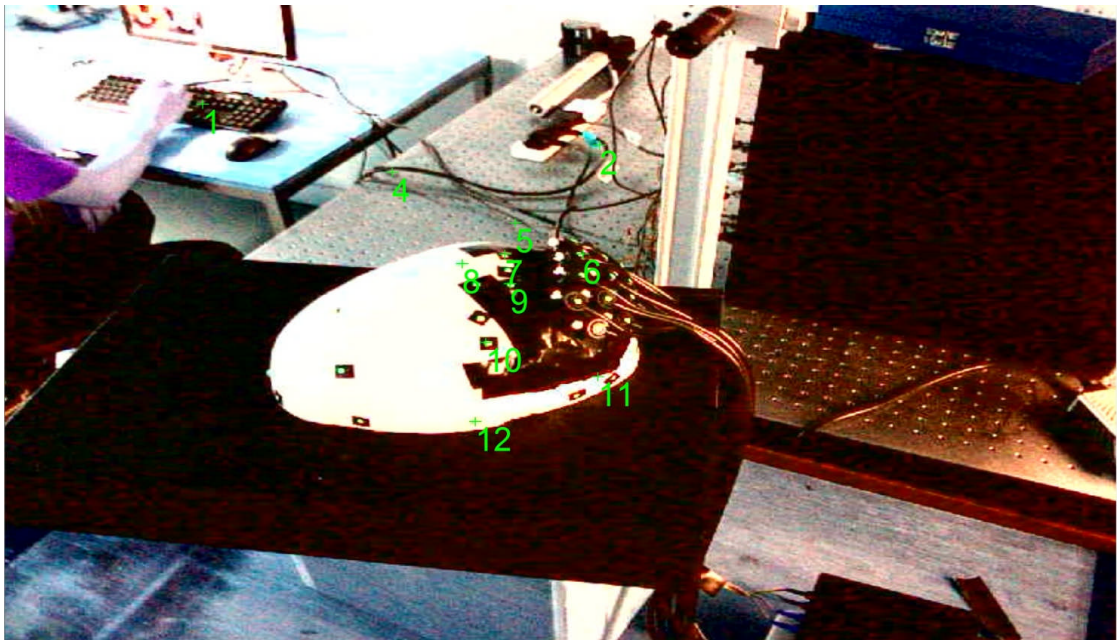


Figure 5 - 47. *Target image measurements by camera eight based on a random frame during the live tracking process of visual cortex study.*

Five frames were also randomly selected out of the total of eight frames captured from the live tracking process. Table 5 - 15 summarizes the 3D coordinate discrepancies of all targets computed between data captured with the live system and single sets of snapshot images processed in VMS:

Frame no.	1	3	4	7	8
Total no. of targets	39	39	39	39	39
No. of targets coordinated by tracking system	24	18	22	19	24
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.19 Y: 0.29 Z: 0.50	X: 0.19 Y: 0.35 Z: 0.37	X: 0.15 Y: 0.22 Z: 0.54	X: 0.19 Y: 0.27 Z: 0.46	X: 0.19 Y: 0.31 Z: 0.52
Mean 3D discrepancies (mm)	0.61	0.54	0.60	0.57	0.63

Table 5 - 15. *Target 3D coordinate discrepancies of all targets between the live tracking system and fixed individual images with VMS intersection.*

To account for the target coordinate discrepancy between VMS processing and the tracking solution, Table 5 - 16 compares number of target image measurements from each individual camera between VMS processing and those successfully used for correspondence captured from the running tracking system.

Camera	No. of target image measurement
1	16
2	12
3	10
4	28
5	26
6	25
7	27
8	18

	Frame 1	Frame 3	Frame 4	Frame 7	Frame 8
Cam 1	8	5	7	6	8
Cam 2	6	6	6	5	3
Cam 3	4	8	8	6	8
Cam 4	20	19	19	21	23
Cam 5	23	18	20	20	24
Cam 6	22	21	20	17	20
Cam 7	18	14	17	19	15
Cam 8	10	7	8	6	9

Table 5 - 16. *Target image measurements for visual cortex study.*

Left: fixed individual images processed with VMS;

Right: those successfully used for correspondence from the live tracking system.

Table 5 - 17 categorises image measurements not used in the correspondence from the running tracking system.

	Frame 1			Frame 3			Frame 4			Frame 7			Frame 8		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Cam 1	0	1	0	0	2	0	0	2	0	0	1	1	1	2	1
Cam 2	0	1	2	0	0	2	0	0	2	0	2	1	0	1	2
Cam 3	0	1	0	0	0	2	1	1	0	0	0	2	0	0	1
Cam 4	1	1	0	1	1	1	3	1	0	0	2	0	0	1	2
Cam 5	1	1	2	2	3	0	3	0	0	2	0	1	0	0	2
Cam 6	2	0	0	3	1	0	3	1	0	3	1	1	2	1	0
Cam 7	1	3	4	3	2	1	2	2	2	2	4	1	2	3	3
Cam 8	0	1	3	0	0	4	0	0	2	0	0	3	1	0	0

Table 5 - 17. *Target image measurements not used in the correspondence from the live tracking frames. A: Correspondence failure; B: Insufficient measurements; C: Not a target.*

For this test, the pad above the visual area was oriented towards those cameras connected to the USB 3.0 ports. Table 5 - 18 summarizes the 3D coordinate discrepancies of these targets computed between data captured with the live system and single sets of snapshot images processed in VMS:

Frame no.	1	3	4	7	8
Total no. of targets on pad	13	13	13	13	13
No. of targets coordinated by tracking system	13	8	10	10	12
Absolute mean discrepancies of coordinated targets (mm) between VMS processing and tracking solution	X: 0.16 Y: 0.27 Z: 0.45	X: 0.17 Y: 0.29 Z: 0.39	X: 0.08 Y: 0.11 Z: 0.51	X: 0.16 Y: 0.23 Z: 0.47	X: 0.14 Y: 0.17 Z: 0.31
Mean 3D discrepancies (mm)	0.55	0.52	0.53	0.54	0.38

Table 5 - 18. *3D coordinate discrepancies of targets on sensing pad above visual cortex area between the live tracking system and fixed individual images with VMS intersection.*

The discrepancies for those targets on the sensing pad above the visual cortex area are compared with which for all targets, as shown in Figure 5 - 48.

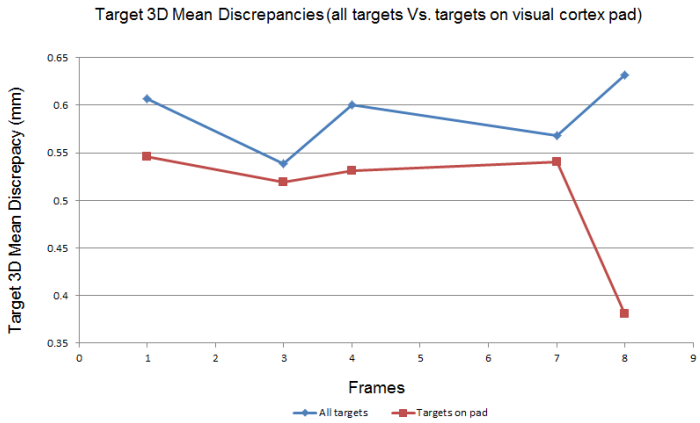


Figure 5 - 48. *Coordination discrepancies of all targets and targets on visual cortex pad from the live tracking system.*

In terms of target tracking consistency across frames, Figure 5 - 49 compares the tracking results between more reliable image qualities delivered from Cam 4, 5, 6, 7 connected to USB 3.0 ports and those less reliable ones delivered by Cam 1, 2, 3, 8 connected to USB 2.0 ports.

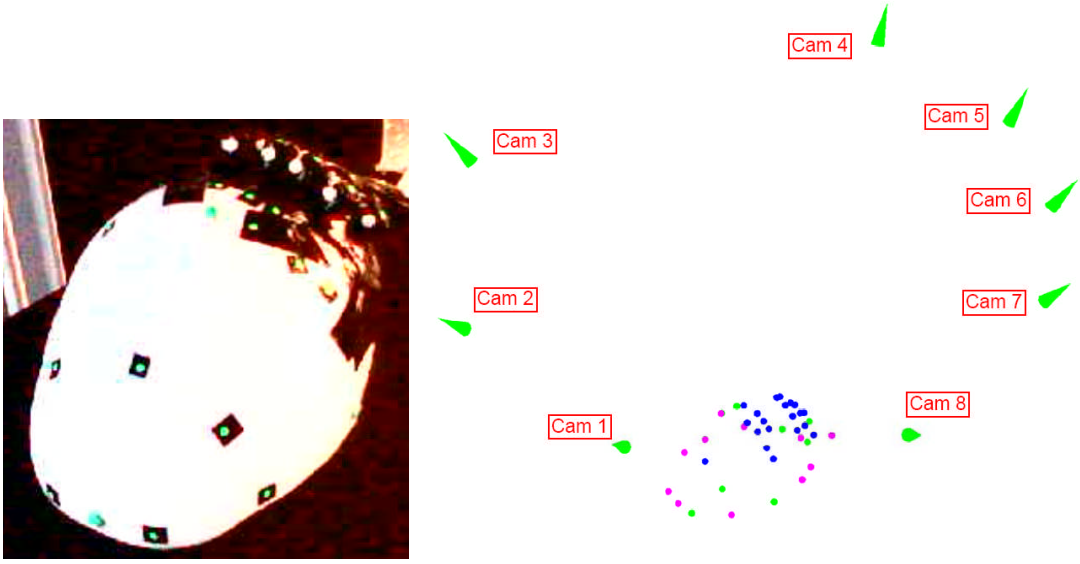


Figure 5 - 49. *Target tracking consistency for visual cortex study.*

Left: Targets on phantom; Right: Relative positions of targets and cameras for tracking solution. (Blue dots: Targets consistently coordinated by the tracking system; Green dots: Targets sometimes disappeared from the coordination; Pink dots: Targets consistently not coordinated by the tracking system)

For targets consistently coordinated with the tracking solution (blue dots in Figure 5 - 49), Figure 5 - 50 further compares the number of camera rays used to intersect these targets between VMS processing and the selected frames from tracking solution.

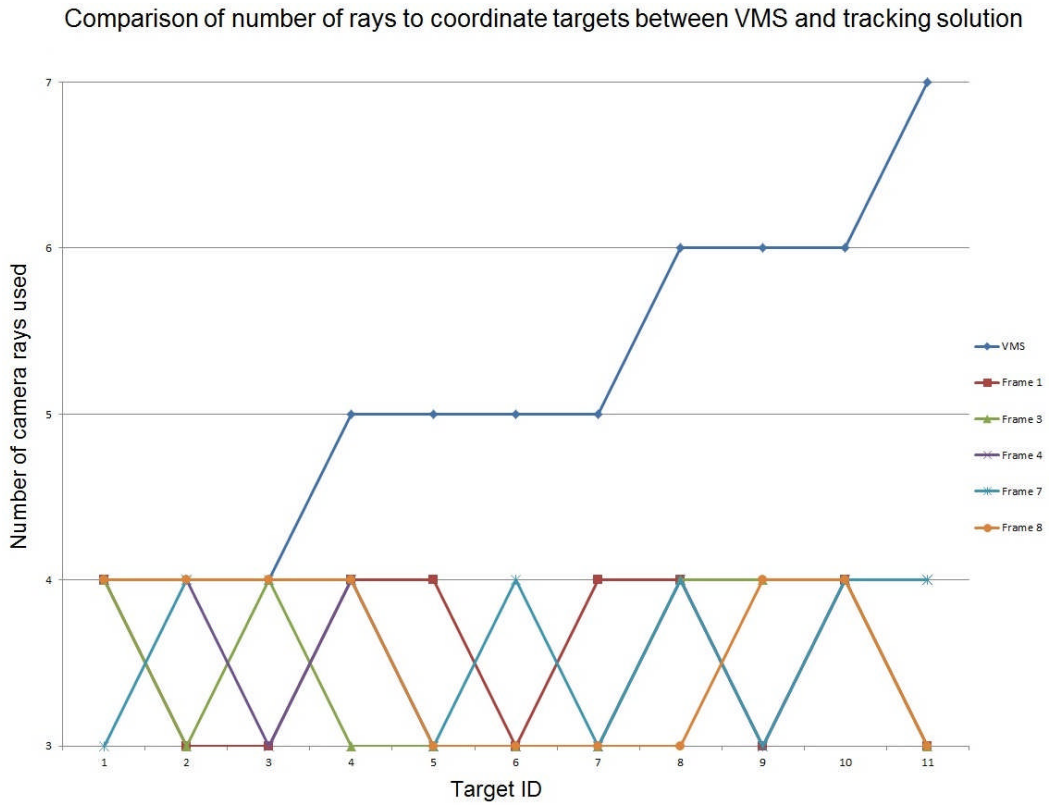


Figure 5 - 50. *Comparison of number of camera rays used to intersect those consistently tracked targets between fixed individual images with VMS intersection and the live tracking frames.*

Figure 5 - 51 shows a typical set of 3D target coordinate discrepancies between the tracking system intersected by three camera rays and VMS intersection while Figure 5 - 52 shows those intersected by four camera rays. In these tests the mean 3D discrepancy for three rays was 0.77mm whilst that for four rays was 0.40mm representing a less extreme example of the 3D discrepancy between using two and six cameras described in Section 5.1.

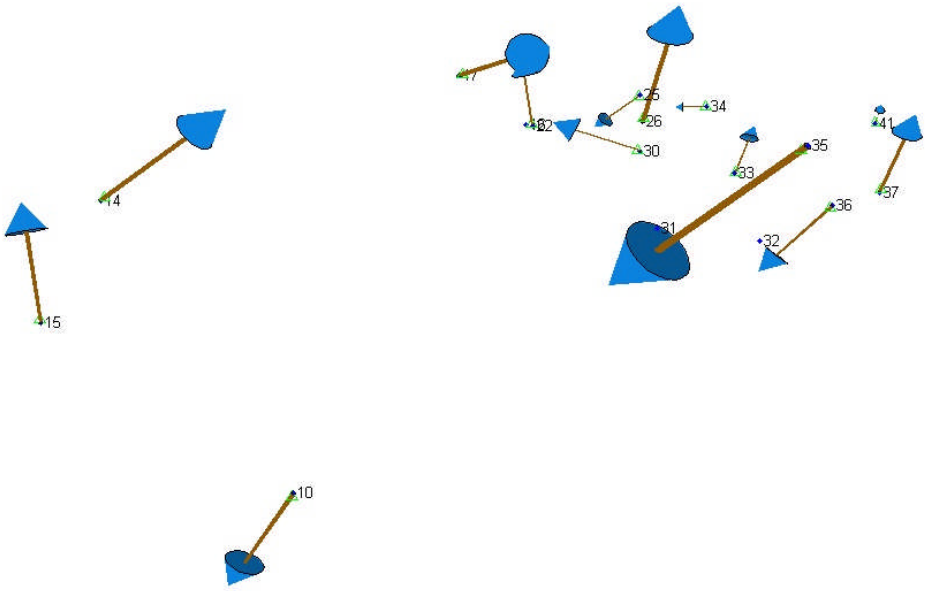


Figure 5 - 51. *Target 3D location discrepancies between tracking system (intersected by three camera rays of the 5th frame) and VMS intersection for visual cortex study (Mean discrepancies: X: 0.22mm; Y: 0.45mm; Z: 0.58mm; discrepancy vectors × 50 for visualisation).*

Figure 5 - 52. *Target 3D location discrepancies between tracking system (intersected by four camera rays of the 5th frame) and VMS intersection for visual cortex study (Mean discrepancies: X: 0.14mm; Y: 0.25mm; Z: 0.29mm; discrepancy vectors × 50 for visualisation).*

5.4.3.1.4 Image reconstruction

Gibson *et al.* (2005) presents detailed methodology of image reconstruction for optical topography. The diffusion of light across the head was modelled by Arridge *et al.* (2000) using a finite element method, which requires a head surface mesh which matches the optodes' positions and meanwhile align with the pad contours.

Figure 5 - 53. *Scanned data (left) and head mesh generated for image reconstruction (right). Left: 669379 surface cloud points; Right: 64952 volume cloud points.*

Conventionally optodes' positions were first measured by manually pointing a 3D digitizing arm (Microscribe 3D, Immersion Co., USA) to each optode in contact with the scalp. A few more positions of the optical sensing pad were then taken with the same device to generate the pad contours. This follows by warping a generic head surface mesh using NETGEN (an automatic 3D mesh generator) to align the mesh with the measured positions to generate a individual finite element mesh [Gibson *et al.*, 2003]. However, such positioning method is subject to manual measurement errors. Therefore in this phantom study, the phantom head was scanned with an Arius3D foundation system 3D scanner and the head mesh was then generated directly from the scanned point cloud (Figure 5 - 53) by Department of Medical Physics and Bioengineering of UCL. Since the light field was very even, the mesh density did not need to be high and therefore the cloud points were reduced from the original scanned data for reconstruction purpose (Figure 5 - 53). This low mesh density requirement is also in agreement with the relatively low spatial resolution of the optical imaging (Section 2.1.3.3). Rather than aligning the mesh to match the optode's position, a project method based on triangle normal in close proximity to the surface mesh (search for a surface normal coincided with the 3D optode point) was used to shift the optode's positions derived from the fast tracking system onto the surface of the head mesh (Figure 5 - 54).

The finite element generation and the optode reprojection process were carried out as part of the optical topography reconstruction by the Department of Medical Physics.

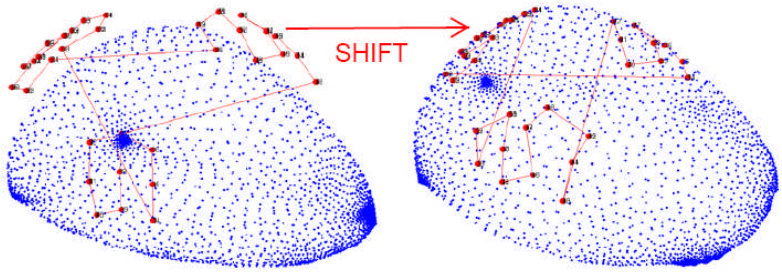


Figure 5 - 54. *Alignment of the optodes' positions with the head mesh.*

A three dimensional (3D) image representing the absorption change below the optical sensing pad has been reconstructed for each study (Figure 5 - 55), with a limited spatial resolution (Section 2.1.3.3). Details regarding the image reconstruction procedures were described in [Hebden *et al.*, 2008].

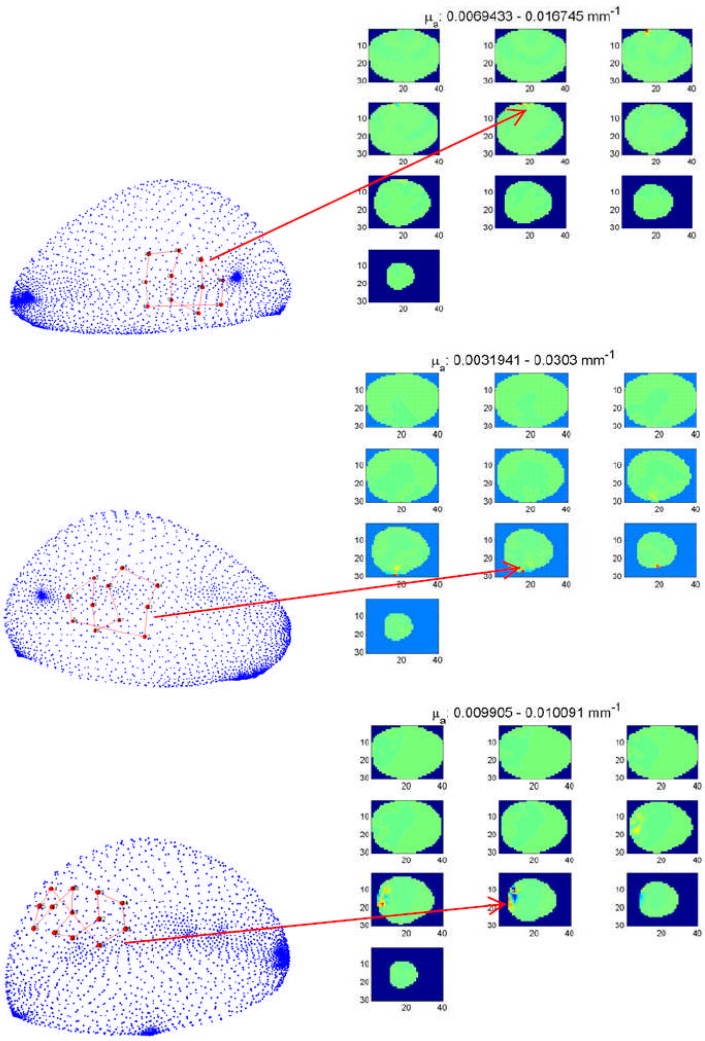
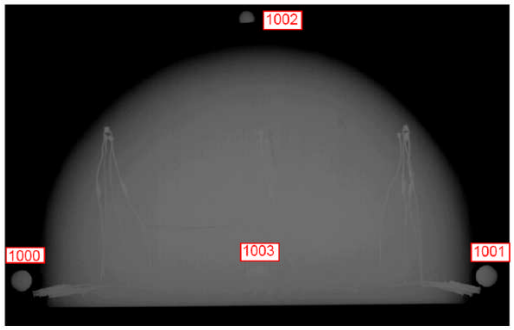


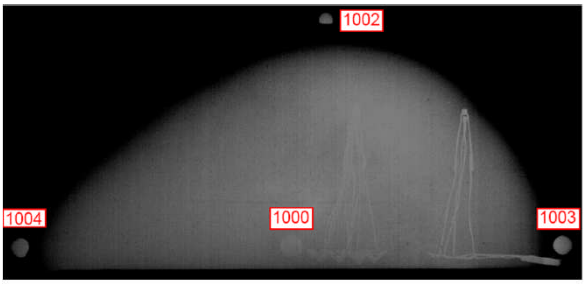
Figure 5 - 55. *Image reconstruction of temporal left (1st row), right (2nd row) and visual (3rd row).*

5.4.3.1.5 Co-registration between optodes and anatomy

Three X-ray images of the phantom were taken in order to coordinate the active locations within the phantom in co-registration with the surface mesh. Common to the two data sets are five ball bearings representing five bony landmarks on the head which were fixed in position on the surface of the phantom before X-ray and laser scanning. Since the five ball bearings could be identified in both the 3D scanning through sphere fitting and the X-ray images of the phantom by centroiding, they could be used as common points for co-registration of the three heating locations (temporal left, temporal right and visual) inside the phantom.



Distance	Scale factor (pixel/mm)
1002-1003	6.582
1000-1001	9.618
1000-1002	9.303
1001-1002	9.994
1000-1003	5.858
1001-1003	5.541



Distance	Scale factor (pixel/mm)
1003-1004	9.694
1000-1002	7.227
1002-1003	9.670
1002-1004	9.721
1000-1004	7.786
1000-1003	7.886

Distance	Scale factor (pixel/mm)
1003-1004	8.621
1000-1001	8.649
1000-1004	8.681
1000-1002	5.398
1000-1003	8.684
1002-1003	5.256
1001-1003	8.472
1001-1002	6.877
1001-1004	8.685
1002-1004	7.814

Figure 5 - 56. *Scale factors computed for three X-ray images.*

Since no information about the X-ray system (e.g. principal distance and aperture) were known at the time when this experiment was conducted, a physically based photogrammetric model could not be established to derive parameters for the X-ray camera(s). Under these circumstances a first trial would be to assume that the X-ray approximated a parallel projection. One way to test this is to compute scale factors of various distances for the X-ray images (Figure 5 - 56) and to compare them with the known separations between spheres from the laser scanned data. The scale factor is defined as:

$$\text{Scale factor} = \frac{\text{Slope between two points in image space}}{\text{slope between the two points in object space}} \text{ (pixel/mm)} \quad 5 - 1$$

Figure 5 - 56 suggests that scale factor varies across landmarks for each X-ray image. If parallel projection was used for x-ray images, the worst case uncertainty in 3D coordination can be considered to be represented by the longest inter-sphere separation (1003 to 1004). In this case the variation in scale factor yields a maximum discrepancy of 176.5mm compared to the 3D separation measured from the 3D scanned data of the phantom, which is much larger than the 1mm tolerance (Section 2.2.3).

Scale factor (pixel/mm)	3D discrepancy(mm)
Average (8.001)	42.7
Minimum (5.256)	170.4
Maximum (9.994)	-6.1
Maximum discrepancy	176.5

Table 5 - 19. *Largest 3D discrepancy resulted from variation of scale factors. Distance of 1003-1004: X-Ray image measurement: 1956.7 pixel; Actual 3D distance: 201.8mm.*

The above results demonstrate that parallel projection is unsuitable for estimation of sphere coordinates from the X-ray images. As an alternative Goktepe and Kocaman (2011) describe the collinearity based Direct Linear Transformation (DLT) to calculate camera parameters belonging to a picture pair of X-ray images. The main advantage of DLT method is that it does not require a calibrated camera. Mathematical model of DLT method is determined by calculation of image coordinates (u, v) of any point by using

object coordinates (X,Y,Z) of that point:

$$u = \frac{L_1X + L_2Y + L_3Z + L_4}{L_9X + L_{10}Y + L_{11}Z + 1}$$
$$v = \frac{L_5X + L_6Y + L_7Z + L_8}{L_9X + L_{10}Y + L_{11}Z + 1} \qquad 5 - 2$$

Transformation parameters $(L_1, L_2, ..., L_{11})$ used in (5-2) can be calculated by means of at least six points at both image and object coordinate systems where individual coordinates are known [Abdel-Aziz and Karara, 1971]. Since only five ball bearings are available, electrical connectors located on the edge of the phantom and visible in the x-ray images and in the surface scan were selected as additional control points, where their 3D and 2D coordinates were obtained by manually measuring the centres of holes of these connectors from the 3D scanned data and X-ray images respectively noting that these manual measurements will be significantly less accurate than the sphere measurements. The DLT solution was implemented using DGAP [DGAP, 2005]. Table 5 - 20 shows the parameters for the DLT solution:

```
Input data
==========
Number of ...
  - images            :   3
  - cameras           :   1
  - image points      :   41
  - object points     :   13

Least-squares parameters/options
================================
Standard deviation of unit weight  :      4.00 um
Convergence limit                  :  1.00e-016
Unknowns used for iteration criterion
  - object point coords.           :  on
Convergence limits
  - object point coords.           :    0.0010 m
  - perspective centre coords.     :    0.0010 m
  - rotation parameters            :  1.57e-006
Weight for suppression of unknowns :  1.00e+010
Weight for constraints             :  1.00e+010


Photogrammetric network design parameters/options
================================================
Minimal number of rays
  - per object point               :  2
  - per image                      :  6

Cameras
=======
Coord. system type :  pixel
Principal distance :  c  =   500.0000 mm
Image format       :  sx =    23.0400 mm      sy =    32.0000 mm
Number of pixel    :  nc =       2304 pix     nr =       3200 pix
Pixel spacing      :  sc =     10.000 um/pix  sr =     10.000 um/pix

Number of observations, unknowns and redundancy
-----------------------------------------------
Observations:
   Image point coordinates      :      37
   Total number of observations :      74
Redundancy                      :      62
```

Table 5 - 20. *Parameters for DLT solution.*

However, the derived positions of the three heating locations from DLT iterations are unreliable, where these positions were significantly different from those expected (Figure 5 - 57):

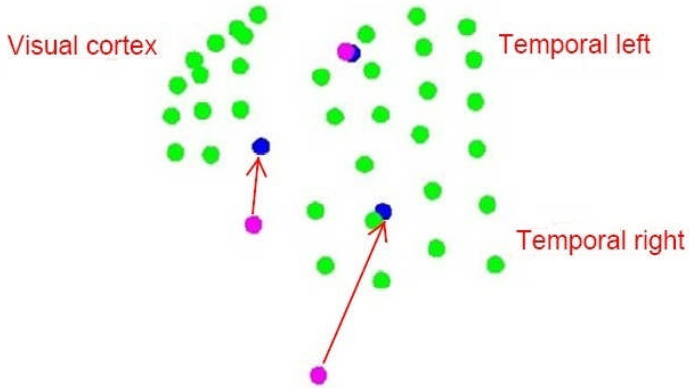


Figure 5 - 57. *3D positions of heating locations derived from DLT solution (red) and their expected positions (blue) below the OT optode's array. (green dots represent optode positions)*

One criterion for reliably estimating the eleven parameters for each image is that the distribution of control points needs to be homogenous, i.e. control points need to be placed on different planes [Abdel-Aziz and Karara, 1971]. In this study, the common feature points which can be distinguished in three X-ray images are the five ball bearings and the eight connectors (four on left and four on back side), as indicated by pink dots in Figure 5 - 58. By formulating equations in (5-2) as linear equations with the DLT coefficients as unknowns we get:

$$\begin{aligned} X \times L_1 + Y \times L_2 + Z \times L_3 + L_4 - u \times X \times L_9 - u \times Y \times L_{10} - u \times Z \times L_{11} - u = 0 \\ X \times L_5 + Y \times L_6 + Z \times L_7 + L_8 - v \times X \times L_9 - v \times Y \times L_{10} - v \times Z \times L_{11} - v = 0 \end{aligned} \qquad 5 - 3$$

When control points are coplanar, the computation of the DLT coefficients becomes impossible since Z becomes a linear function of X and Y, i.e. for each control point:

$$Z = aX + bY + c \qquad 5 - 4$$

where a, b and c are arbitrary variables. Under this circumstance, the system of equations of (5-3) will only solve eight independent coefficients out of a total of eleven no matter how many coplanar control points are used. This is because the rank of matrix containing a collection of equations of (5-3) will be at most 11-3=8, where this loss of rank is caused by columns corresponding to the unknowns L_3, L_7 and L_{11}

with coefficients containing Z. Therefore the 11×11 matrix containing the system of normal equations will be singular. If there is one control point outside the common plane of the others (such as the top pink dot shown in Figure 5 - 58), the rank can be increased from eight to ten making it still a singular normal matrix.

Figure 5 - 58. *Back view of phantom head showing the control points used for DLT (pink) and the expected heating locations (blue) to be estimated.*

As shown in Figure 5 - 58, the control points are nearly coplanar resulting in an ill-posed system. As a consequence small errors in control point coordinates (or their image measurements) can cause large variations in the DLT coefficients. These in turn will propagate to the positions of projection centre, which will be used in inverse DLT (from 2D to 3D). Even if the control points were error free, DLT is still unreliable for points (blue dots in Figure 5 - 58) away from the control points (pink dots) in case of near coplanarity [Kwon, 1989]. Kwon (1989) further suggested using control points which convex hull cover the complete region (also with respect to Z coordinates (heights)) if one wants to make sure that DLT is reliable for the whole region of interest.

Based on the geometry of the feature (control) points located on the phantom in this study, DLT is not an appropriate solution to estimate the 3D coordinates of the heating blobs inside. Compared to DLT which estimates eleven parameters for each image, VMS (Geometric Software) estimates six parameters (three for position and three for rotation) allowing a more stable estimation of the entire network geometry. Note that VMS based on collinearity requires estimation for principal distance and principal point. Table 5 - 21 shows the input parameters for VMS to estimate the orientations of three X-ray images and then intersect to estimate the 3D heating positions:

Program Control Variables	Values
Network datum definition type	External constraints on targets
Maximum iterations for a solution	10
Default target image precision by camera (μm)	80.00
Minimum images for a network target	2
Rejection criterion for image errors	50μm
Input Summary	**Values**
Number of camera calibration sets	1
Number of target image observations	37
Total number of exposures	3
Number of exposures in the network	3
Total number of targets	16
Number of targets in the network	16
Results for the calibration solution	**Values**
Unit weight estimate (sigma zero)	1.29
RMS image residual (μm)	81.40
Number of observables in the network	117
Number of unknowns in the network	70
Number of redundancies in the network	47
Mean number of images per target	2.3
Target Precision Summary	**Values**
Mean precision of target coordinates (μm)	292.03
Relative precision for the network (1:X)	1000

Table 5 - 21. *Summary of parameters of estimating 3D positions of heating blobs.*

In order to use a conventional bundle adjustment to achieve a best fit, the X-ray imagery was taken to be close to a parallel projection by setting the principal distance of the X-ray camera to 500mm. The principal distance was constrained along with the principal point coordinates with a standard deviation of 0.1mm so that the adjustment solution would converge. The RMS image residual of 81.4μm is equivalent to 8 pixels provided with the difficulty in manually measuring the locations of the X-ray image features, noting that the perimeter of a ball bearing in X-ray images is ~60 pixels. A radial and an affine distortion parameter were included in the adjustment as unknowns (Figure 5 - 59).

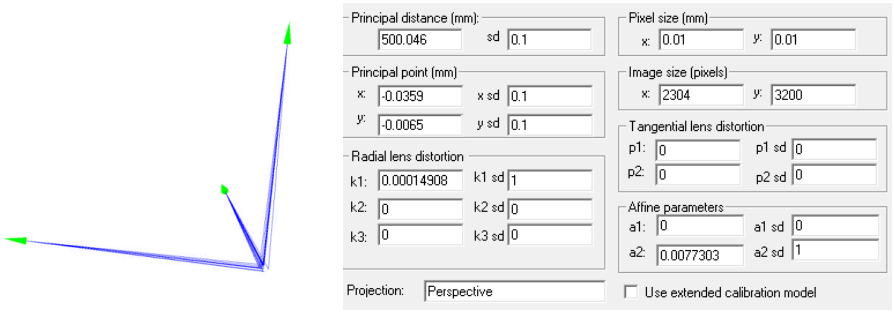


Figure 5 - 59. *Relative positions between X-ray cameras and phantom (left) and parameters for bundle adjustment (right).*

The above setting allowed VMS to use the best fit from bundle adjustment to estimate the 3D coordinates of the three heating blobs (Table 5 - 22). In the results, the larger uncertainty for the right blob was due to the fact that the right blob can only be measured in two out of three X-ray images (Figure 5 - 56).

Blobs	X (mm)	Y (mm)	Z (mm)	sX (μm)	sY (μm)	sZ (μm)	#Images
left	-454.7	-117.3	-35.6	704.9	739.8	695.8	3
right	-449.6	-215.8	-40.3	1113.4	748.6	1012.6	2
back	-493.5	-169.8	-40.4	756.4	702.6	702.2	3

Table 5 - 22. *3D coordinates estimated for the heating blobs.*

Figure 5 - 60 illustrates the methodology used to co-register the optodes from optical topography with the functional activated area (heating blobs) inside brain (within the same coordinate system) utilising the bony landmarks (defined by the international 10-20 system, Section 2.2.3) as common points.

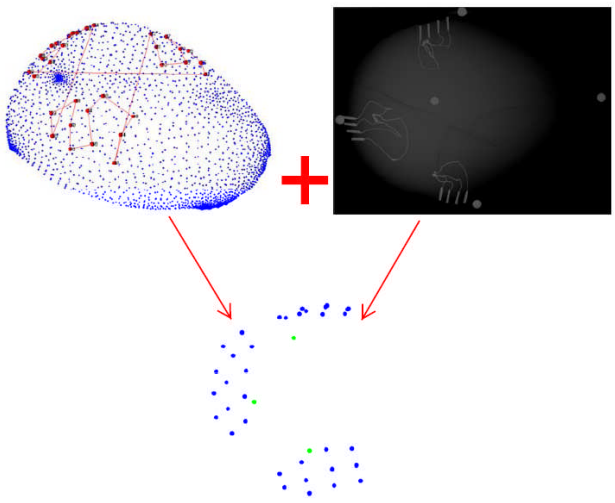


Figure 5 - 60. *Co-registration of heating blobs inside phantom (green) and OT optodes (blue).*

A peak position (Table 5 - 23) representing the highest optical change for each area (Figure 5 - 55), was manually searched through the reconstruction by the Department of Medical Physics. Since the image reconstruction was based on the surface mesh, these peak positions were under the same coordinate system as the optodes allowing a direct comparison of the three peak positions estimated between optical imaging and photogrammetric co-registration (Table 5 - 22) without the need of further coordinate transformations (Table 5 - 23).

Peak in mesh grid	X (mm)	Y (mm)	Z (mm)	3D discrepancy (mm)
Left	-459	-91	-53	31.9
Right	-452	-229	-40	13.5
back	-502	-177	-35	12.4

Table 5 - 23. *Peak positions from reconstructed optical images and 3D coordinate discrepancy from the coordinates estimated from bundle adjustment (Table 5 - 22).*

Peak positions from optical images (pink dots in Figure 5 - 61) appeared more closer to the phantom surface (or optodes) for both temporal right and visual area, with respect to the positions derived from bundle adjustment (blue dots in Figure 5 - 61). Y-coordinate of the peak position from optical image data for temporal left was beyond those of optodes (pointed by arrows in Figure 5 - 61).

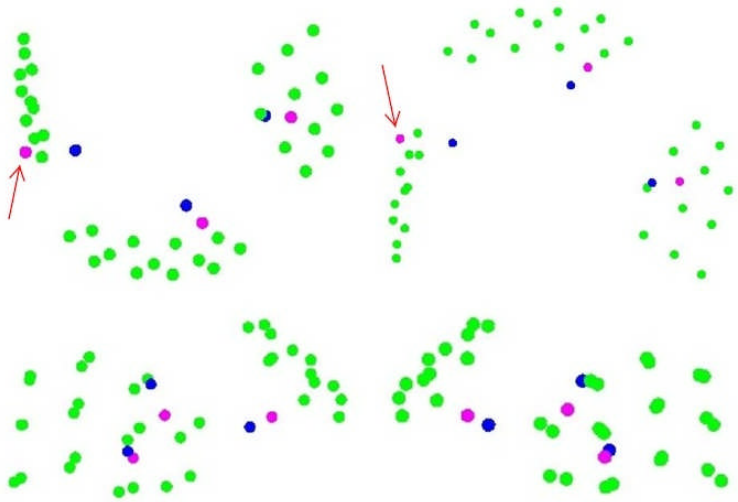


Figure 5 - 61. *Visual comparison of three heating locations estimated between bundle adjustment and optical imaging. Upper left: top view; Upper right: bottom view; Lower left: left-sided view; Lower right: right-sided view. Green dots: Optode positions; Blue dots: Heating centre locations estimated by bundle adjustment; Pink dots: Peak positions from optical images.*

5.4.3.2 Human subject one

During the optical topography study, 3D target coordinates computed from sequential sets of frames were regularly output to a text files during the tracking process in order to reconstruct the overall target movement. Five frames were randomly selected from the twenty-minute tracking sequence captured during the functional study. These data represented approximately equal time intervals along the sequence, i.e. the locations of the head and the pad every four minutes. The task for this functional study was repeating of finger tapping to see if a particular temporal area would be activated. However, it was observed that subject one fell asleep after a while since the study began and this can be clearly reflected from the reconstructed target movement during the study as shown in Figure 5 - 62.

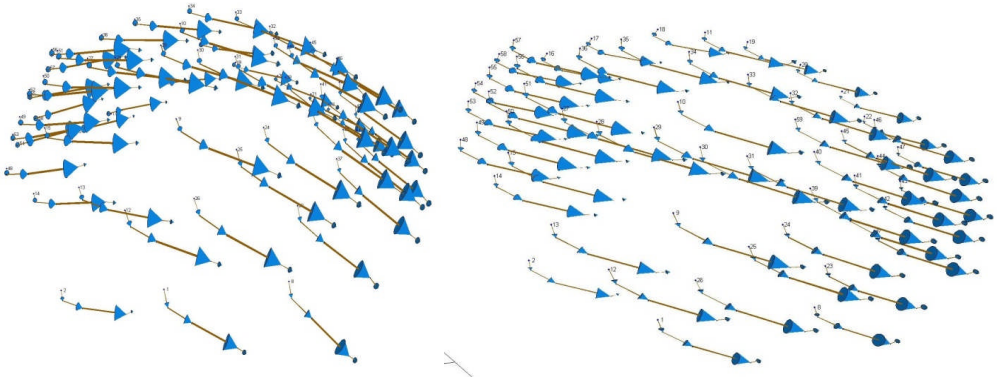


Figure 5 - 62. *Front (left) and top (right) views of target movement of subject one during the 20-minute study.*

The newly designed optical sensing pad was firmly attached onto the scalp by several stickers (right picture of Figure 5 - 63). Please note that this is an advance on the original pad design which used a bandage to attach the pad onto the scalp (left picture of Figure 5 - 63). Despite the change in design, position misalignment between the sensing pad and the scalp underneath can still make the optical topography data invalid.

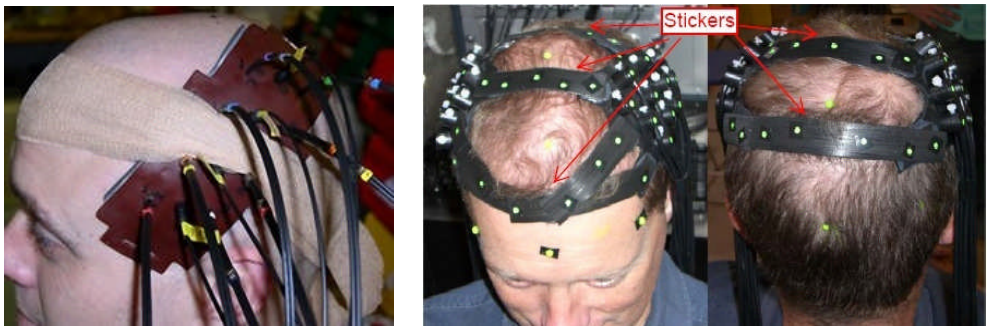


Figure 5 - 63. *Old design with bandage (left) and new design with stickers (right) of optical sensing pad.*

To investigate the presence of differential movement, centroids computed from the four bony landmarks (Section 2.2.3) represented the head position, while centroids computed from the eleven targets on the left (right) pad represented the position of the left (right) pad. Sequential centroids from frames 2, 3, 4 and 5 were then co-registered to their associated centroids computed from frame 1 by 3D similarity transformation [Dewitt, 1996]. Coordinate discrepancies of these centroids after transformation from the reference are shown in Table 5 - 24.

Centroid (mm)	Head			Left pad			Right pad		
Frame	x	y	z	x	y	z	x	y	z
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0.02	0.21	0.03	0.06	0.30	-0.84
3	0	0	0	0.04	0.02	-0.31	-0.20	0.26	-0.26
4	0	0	0	-0.35	0.55	0.27	-0.27	-0.30	-0.44
5	0	0	0	0.16	1.04	0.23	-0.72	0.44	-2.32

Table 5 - 24. *Discrepancies of centroid coordinates of head, left and right pads after co-registration to coordinate system of the 1st frame.*

The co-registered 3D discrepancies of these centroids across frames (Figure 5 - 64) demonstrated how the head and sensing pad had moved differentially during the 20 minute study period. Clearly identifiable are a movement up to 2.5mm in the case of the right pad. In this case it is clear that the pads and head had moved with respect to each other making optical imaging reconstruction invalid since the reconstruction is based on the difference between two sets of data where the only change allowed is the optical change (brain activity).

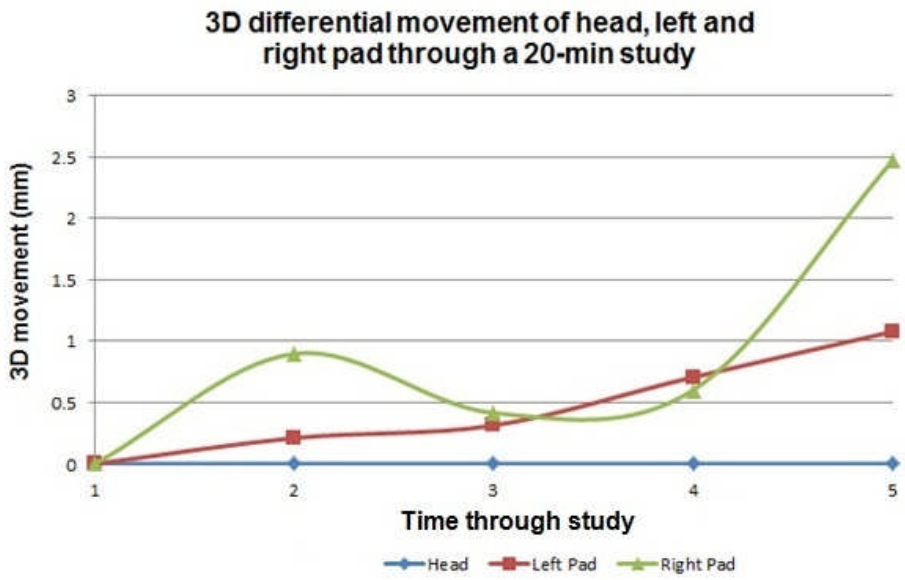


Figure 5 - 64. *Differential movement between the sensing pad and scalp during the study of subject one.*

5.4.3.3 Human subject two

The method used for investigation of differential movement for subject two is the same as that for the first subject. It was observed that subject two did not fall asleep during the study such that the head movements are significantly smaller (Figure 5 - 65).

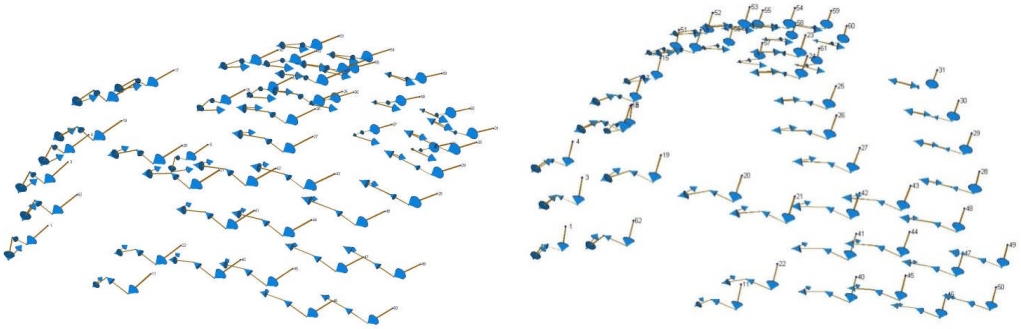


Figure 5 - 65. *Front (left) and top (right) views of target movement of subject two during the 20-minute study.*

Again sequential centroids from frames 2, 3, 4 and 5 were co-registered to their associated centroids computed from frame 1 by 3D similarity transformation [Dewitt, 1996]. Coordinate discrepancies of the targets located on the head and pads after transformation from the reference are shown in Table 5 - 25.

Centroid (mm)	Head			Left pad			Right pad		
Frame	x	y	z	x	y	z	x	y	z
1	0	0	0	0	0	0	0	0	0
2	-0.07	0.03	-0.03	0.06	0.03	0.01	0.01	-0.06	0.02
3	-0.03	0.07	-0.02	0.06	-0.02	0.02	-0.03	-0.05	0.01
4	-0.07	0.00	-0.02	0.04	0.05	0.00	0.04	-0.05	0.02
5	-0.14	0.03	-0.04	0.09	0.07	0.01	0.05	-0.10	0.03

Table 5 - 25. *Discrepancies of centroid coordinates of head, left and right pads after co-registration to coordinate system of the 1st frame.*

The co-registered 3D discrepancies between these target centroids across frames (Figure 5 - 66) demonstrated how the head and sensing pad move differentially during the 20 minute study period. In this case the developed system confirmed that the pad and head had moved together such that the topography can be effectively computed.

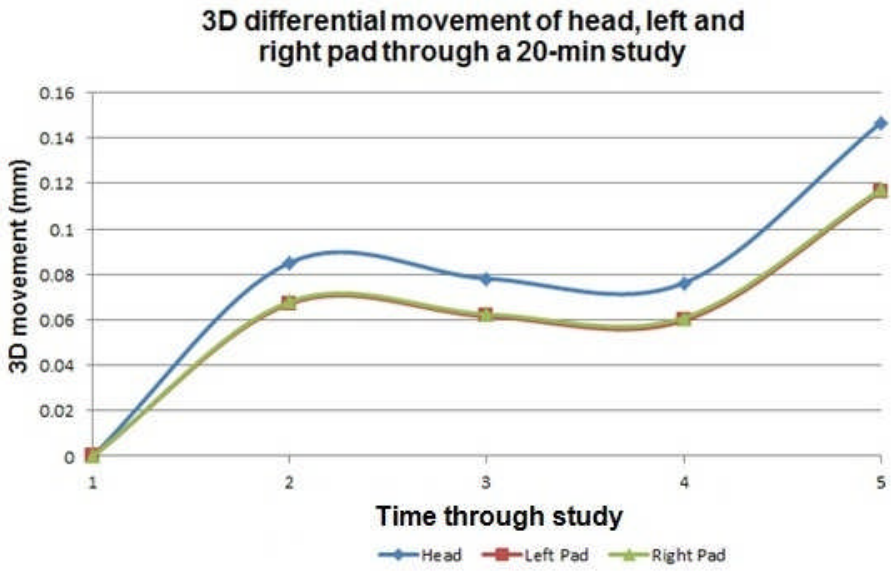


Figure 5 - 66. *Differential movement between the sensing pad and scalp during the study of subject two.*

5.4.4 Discussion

5.4.4.1 3D coordination – targets used for correspondence

Compared with the results of 0.35mm (ten targets on the polystyrene head in Section 5.2.2.2) and that of 0.18mm (two targets on the caliper in Section 5.2.2.3), the 3D mean target coordinate discrepancies for all phantom studies (with VMS intersection results as the reference) are somewhat higher (between 0.54mm and 0.91mm) though still well below the 1mm accuracy required for reliable reconstruction (Section 2.2.3). The difference was mainly caused by the variation of both target locations and target orientations on the phantom. The targets on the polystyrene head and caliper were situated on nearly 2D planes which were almost perpendicular to the camera rays (Figure 5 - 5 and Figure 5 - 6). In case of phantom the curvature of upper sphere increases the difficulty of 3D target intersection. The lower the horizontal level, the more difficult cameras can gain a clear view of the entire target. Such a situation needs to be identified automatically if the tracking process is going to function without any human interaction.

The target coordinate discrepancies for temporal left, right and visual cortex (Section 5.3.3.1) reveal that for the coordinated targets, the Z-coordinate generally shows a larger discrepancy than X- and Y-coordinates accounting for most of the 3D target coordinate discrepancy. Discrepancy dominated by the Z-axis is clearly shown from 3D plots of target 3D location discrepancies for all three area studies (maximum Z-discrepancy of 0.58mm/0.79mm/0.54mm for temporal left/right/visual cortex) on the phantom (Figure 5 - 25, Figure 5 - 26, Figure 5 - 38, Figure 5 - 39, Figure 5 - 51 and Figure 5 - 52). This phenomenon is attributable to the network geometry with the mounting of all eight cameras on the same horizontal level producing lines of sight that converge more narrowly in the direction of the Z-axis. Figure 5 - 67 shows a typical example, where this target had the largest coordinate discrepancies among all coordinated targets from a worse case set of frames. The lowest horizontal level of target location has resulted in the Z-coordinate discrepancy nearly eight times as big as the X- or Y-coordinate discrepancy.

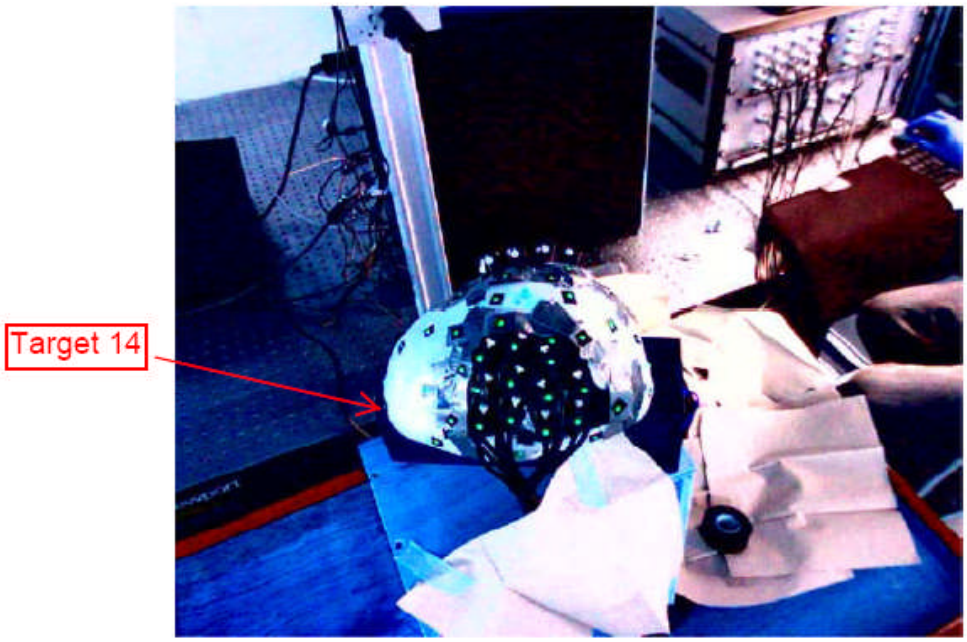


Figure 5 - 67. *The target with the largest discrepancy on Frame 1 of temporal right.*

(with the largest discrepancy of 0.8mm)

(Target coordinate discrepancies: x: 0.46mm; y: 0.344mm; z: 2.268mm)

The above factors have some photogrammetric implications in terms of the future design of the tracking system:

1. A stronger multi-camera convergent geometry can be achieved with cameras located at different horizontal levels for a better determination of target Z-coordinates (Section 3.2.3.1). For example a two-level wheel ring system where each ring contains eight cameras would be needed to make sure that each target on a head sized object can be seen by at least four camera rays (two per level);
2. The camera-to-camera separation should be reduced if accuracies for all targets on the tested object are equally important;
3. The standard 60cm wheel diameter, whilst convenient, is too small for an eight-camera system in terms of tracking targets distributed over a head sized object. The issue of target occlusion/ambiguity can be largely improved if the wheel diameter were increased to allow cameras to simultaneously have a better view of targets on lower horizontal levels of a curved object (Section 3.2.3.1).

5.4.4.2 3D coordination – targets not used for correspondence

Some of the measured target images were not used for 3D correspondence in the running camera system. Causes of this were insufficient target image measurements, incorrectly recognized targets and target images failing correspondence. Each of these issues is discussed as follows:

Insufficient target image measurements

Those image measurements not used for correspondence are identified in Table 5 - 9 for temporal left, Table 5 - 13 for temporal right and Table 5 - 17 for visual cortex. Insufficient measurements imply that there was only one line of sight to the target. Missing lines of sight are attributable to physical target occlusions and target detection inconsistencies due to the unreliable image quality of the cameras connected to the USB 2.0 ports.

In terms of physical target occlusion, there were some targets which only visible to one camera and meanwhile were oriented towards cameras sparsely mounted on the wheel ring (Figure 5 - 68). Moreover, the inconsistency of target image detection by Cam 1, 2, 3 and 8 left some targets (oriented towards these cameras) only detectable by one camera within one frame set again causing insufficient measurements. Regarding this the target detection consistency across frames is considered to be critical to make sure that sufficient camera rays can intersect targets (Section 3.2.3.1) for every single frame in terms of optical topography applications requiring continuous 3D coordination.

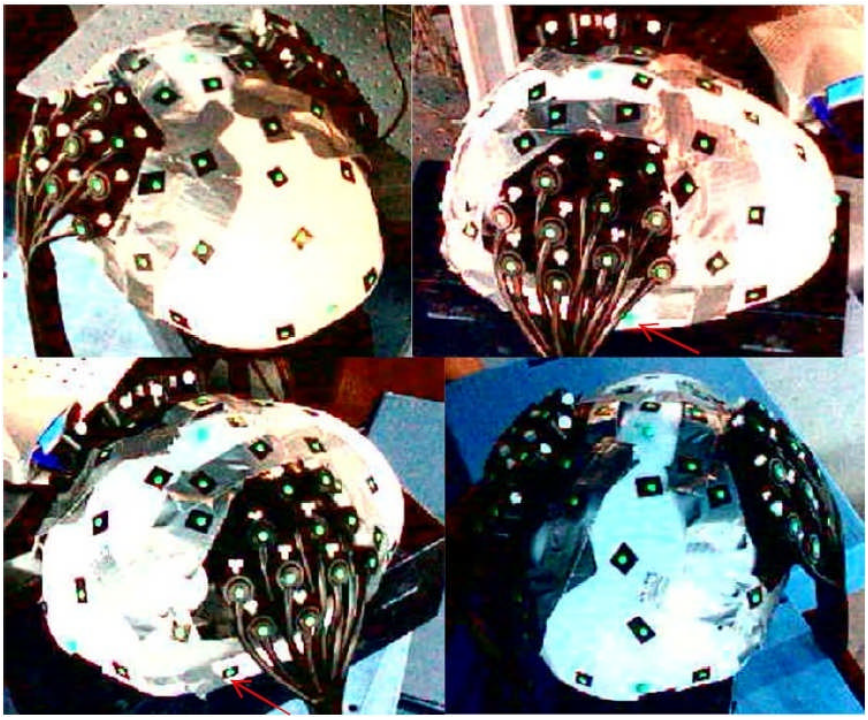


Figure 5 - 68. *Views of Cam 8, 1, 2, 3 (from upper left to lower right) demonstrating some targets (with arrows) only visible by one camera causing insufficient measurement for correspondence.*

Incorrectly recognized targets

Incorrectly recognized targets are highlighted in Table 5 - 9 for temporal left, Table 5 - 13 for temporal right and Table 5 - 17 for visual cortex. These data are the result of adjusting colour tuning thresholds to permit the four cameras connected to USB 2.0 ports to detect some target images. The selected colour tuning threshold was lowered from 45 to 35 for all eight cameras ensuring uniform parameter setting for the system, but resulting in a significant increase in the number of unnecessary/incorrectly recognized targets. Incorrect recognition is particularly obvious in over-exposed image areas especially for Cam 4, 5, 6, 7 which effectively eliminated unnecessary background information with the threshold of 45. Figure 5 - 14 to Figure 5 - 21, Figure 5 - 27 to Figure 5 - 34 and Figure 5 - 40 to Figure 5 - 47 in Section 5.2 highlights some examples of the live target image measurements by each camera in each study. In the worst case where targets were oriented towards the direct sunlight, up to eighteen incorrectly recognized target images (Table 5 - 9) were found (Figure 5 - 69). Even under this circumstance, most of actual (real) targets were correctly recognized and used for correspondence demonstrating the robustness of the correspondence solution in cluttered environments.

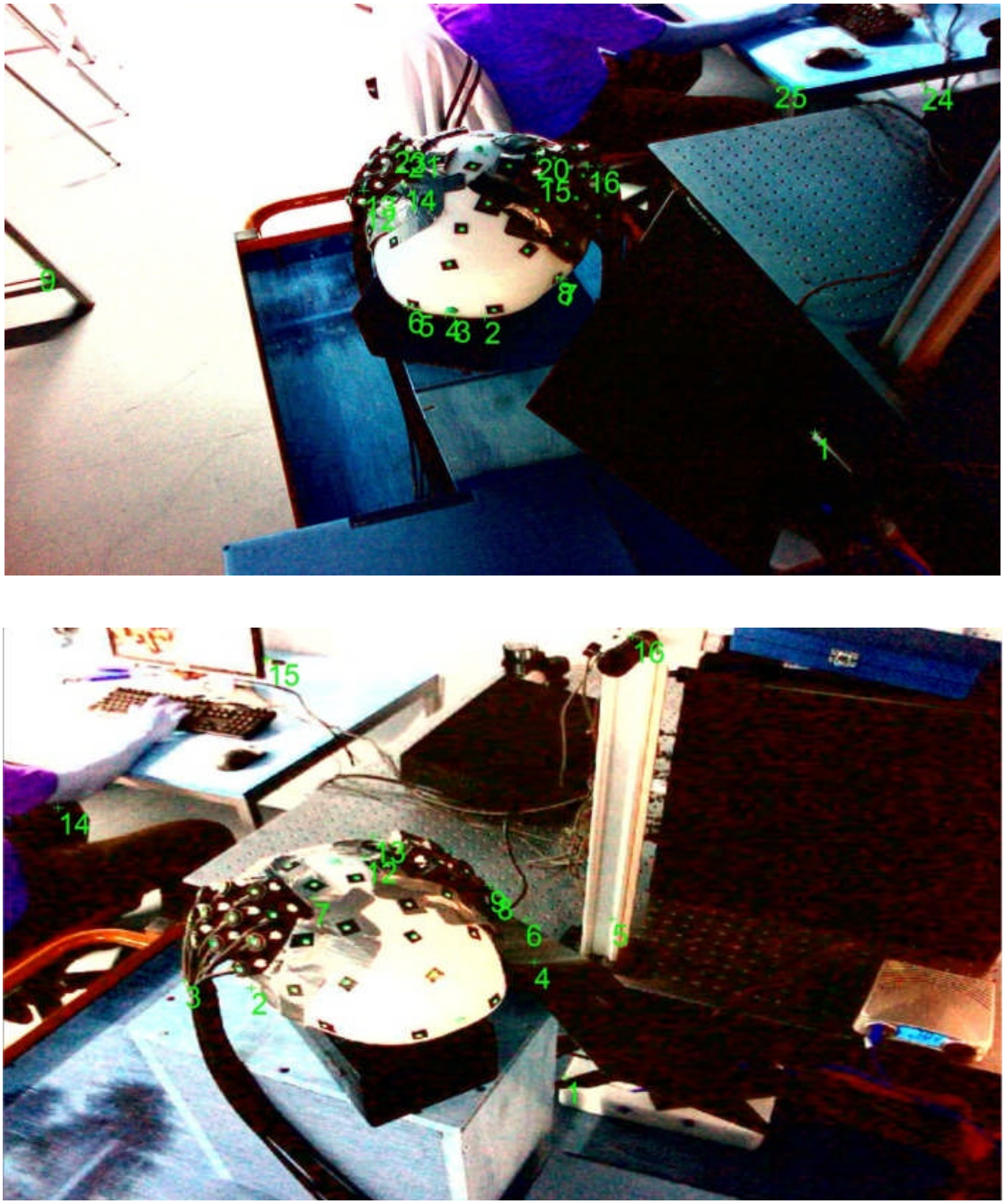


Figure 5 - 69. *18 incorrectly recognized target images (not used for correspondence) by Cam 7 (upper image), and 13 incorrectly recognized target images (not used for correspondence) by Cam 8 (lower image) in Frame 1 of temporal left study.*

Note that the lighting of Frame 1 might be significantly different from that shown from the above images (static images captured at the beginning of the study).

Regarding target image ambiguity for correspondence, it is suggested that incorrectly recognized target images should be minimised in future tests. Whilst incorrectly recognized targets outside the area of interest should not be a concern, targets

incorrectly recognized within the area of interest are important. As an example incorrectly recognized targets on both the sensing pad and the phantom can be seen in Figure 5 - 69 and will increase the possibility of 3D correspondence errors. Both reliable image qualities across all the cameras in the tracking system and an appropriate rejection tuning threshold are required in order to minimize incorrectly recognized target images within all camera views in future studies.

Target images failing correspondence

Target images failing the 3D correspondence solution are identified in Table 5 - 9 for temporal left, Table 5 - 13 for temporal right and Table 5 - 17 for visual cortex. These represent target images corresponding to the same target that are present in two or more images but are not used for correspondence. The main cause of those failing the test appear linked to the different extents of chromatic aberration (Section 3.3.1.4) of one target on different camera viewpoints. Since the phantom was located at different positions within the eight viewpoints in the tracking system, the extent of transverse chromatic aberration around a target image for one image can be different from the transverse chromatic aberration in a second image. Transverse chromatic aberration progressively worsens towards image corners (Section 3.3.1.4). Therefore the displacement among the three colour channels observed for target images towards image corners was much larger as compared with which of target images towards the image centre (Figure 5 - 70). How much the aberration distorts the imaged position in the two images will give rise to reprojection error and hence correspondence failure.

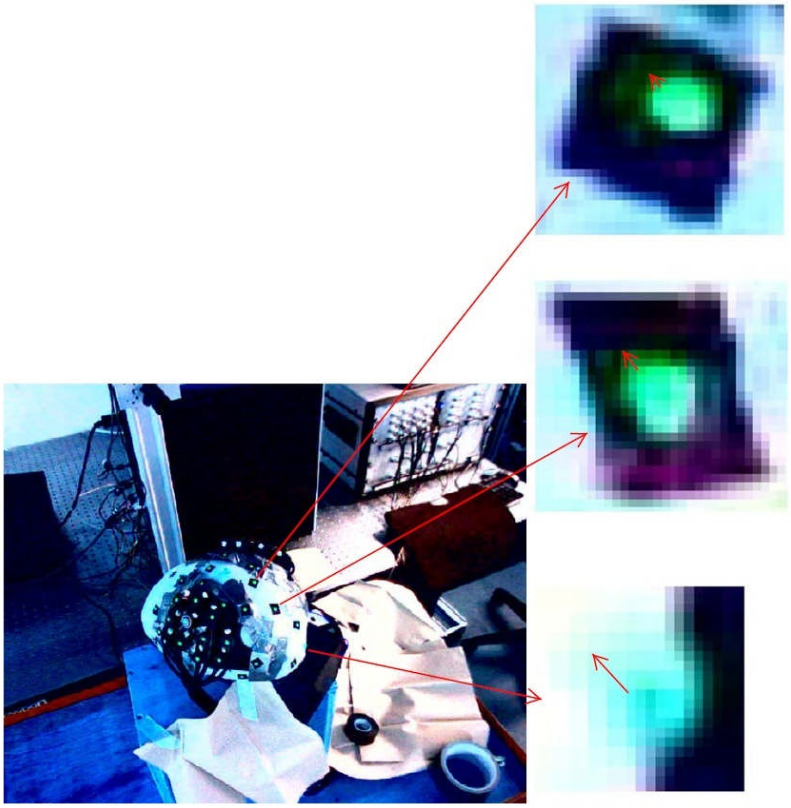


Figure 5 - 70. *Transverse chromatic aberration progressively worsened towards image corners.*

The lens in the C500 webcams has a simple four blade aperture, in combination with fixed focus the result is that many of the target images were not actually in sharp focus (Section 3.3.1.4). At the worst case target images appear as squares giving rise to errors in the colour target image detection capability to search for centroids of these squares (Figure 5 - 70).

Since a calibration object with coloured targets was not available during this research, the self-calibrating bundle adjustment for the camera system (Section 4.3) used a calibration object with retro-reflective targets and a different target image measurement method (in VMS). As a result the calibration might not hold for largely distorted coloured target images towards image corners giving rise to reprojection errors larger than those permitted for a successful correspondence (Section 3.5.3). An example is given in Figure 5 - 71, which also demonstrates, along with Figure 5 - 23, Figure 5 - 36 and Figure 5 - 49, that most targets consistently not coordinated by the running system were located mostly towards image corners. Note that for webcams (1, 2, 3, 8) delivering less reliable images, image compression is also a factor accounting for correspondence failure (Figure 5 - 71).

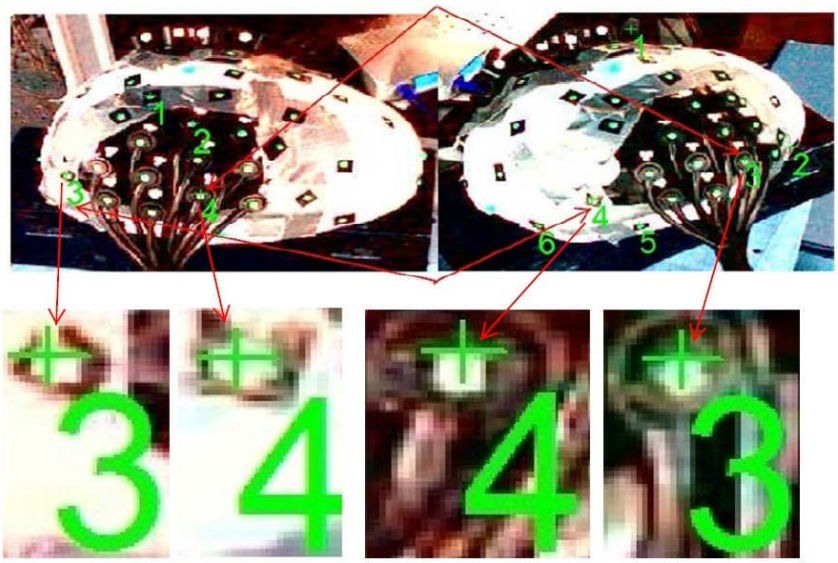


Figure 5 - 71. *Example of target images detected but not used for correspondence.*

Left: Cam 8; Right: Cam 1.

The size of the phantom is much larger than that of the small sensing pad and calipers used in studies in Section 5.2 (Figure 5 - 72). As a result targets on the phantom extend into the image corners compared with target images on the small pad and calipers which are more central. The larger target image distortion towards image corners also explains the overall larger discrepancy compared to auto-centroiding with the window based threshold target image detection method and manual centroid used within VMS. As a result in the phantom studies discrepancies of the order of ~0.6-0.8mm are seen, compared to results of 0.35mm for ten targets on the polystyrene head in Section 5.2.2.2 and; that of 0.18mm for two targets on the caliper in Section 5.2.2.3.

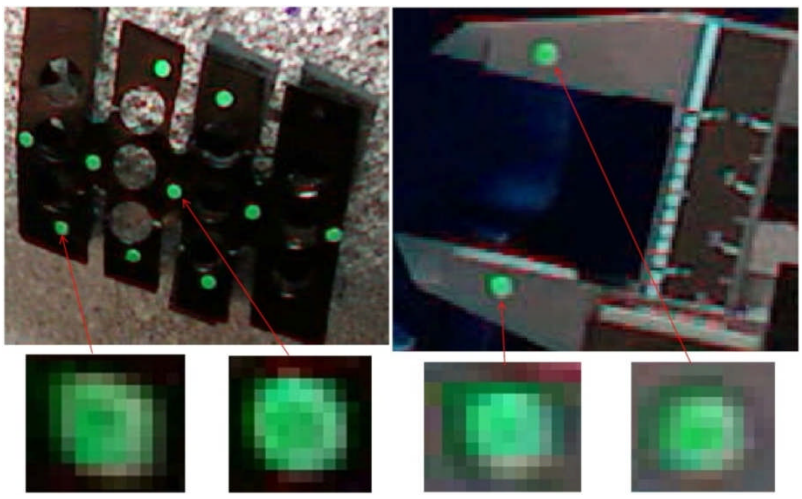


Figure 5 - 72. *Relatively small distortion occurred for target images in studies of Section 5.2.*

Chromatic aberrations were also observed on target images from the optical sensing pad, although these targets were oriented towards Cam 4, 5, 6 and 7 delivering reliable image qualities. The smaller image distortion accounts for the slightly lower 3D discrepancies of these targets as compared with those of all targets (over phantom and pad) where the data includes target images in the image corners (Figure 5 - 22, Figure 5 - 35 and Figure 5 - 48). Interestingly it was also observed that target images close to saturating caused by direct sunlight, were much more distorted as compared with those without saturating sensors (Figure 5 - 73).

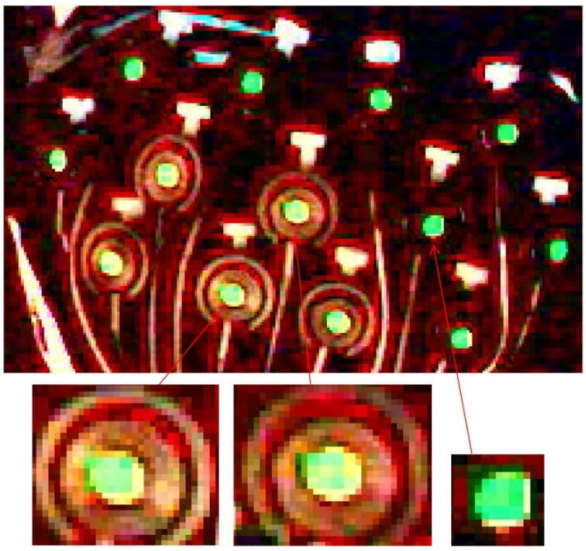


Figure 5 - 73. *More distortion of target image was caused by strong sunlight.*

In order to include chromatic target image distortions in the error budget, a head-sized calibration object equipped with coloured targets suited to the developed colour target-tuning detection method is necessary. Such an object would support calibration of the complete camera system ensuring the full validity of calibration parameters for the target coordination applications. The corrected lens distortion parameters in this way should account for the chromatic aberration in images which will allow more target images to pass 3D correspondence and increase the overall coordination accuracy.

Another approach is to increase the tolerance for target image matching so that targets between images can be matched with each other even if they are slightly distorted. However such an approach would increase the possibility of target mismatches and result in lower 3D coordination accuracy as compared with the previous approach where the computed lens distortion would include the chromatic error.

5.4.4.3 Consistency

As described in Chapter 2, the coordinates of optodes need to be registered onto the brain anatomy for brain functional localisation, which means these locations need to be accurately coordinated in optical topography studies. Results have demonstrated that most of these key targets were consistently detected and coordinated across image frames during the optical topography studies. However, there are one or two targets which were not coordinated in some frames (Table 5 - 10, Table 5 - 14 and Table 5 - 18). By tracing back the actual images during the studies it is found that since there was no light control for optical topography studies in this research and these studies were all conducted in the noon or afternoon during summer, there were inevitably some targets facing directly into the strong sunlight entering from the windows of the testing room. These targets were imaged at near sensor saturation resulting in the green colour no longer standing out against the other colour components, i.e. the difference in intensity between green and red, and between green and blue, are lower than the target detection tuning threshold, i.e. 35 (Table 5 - 6).

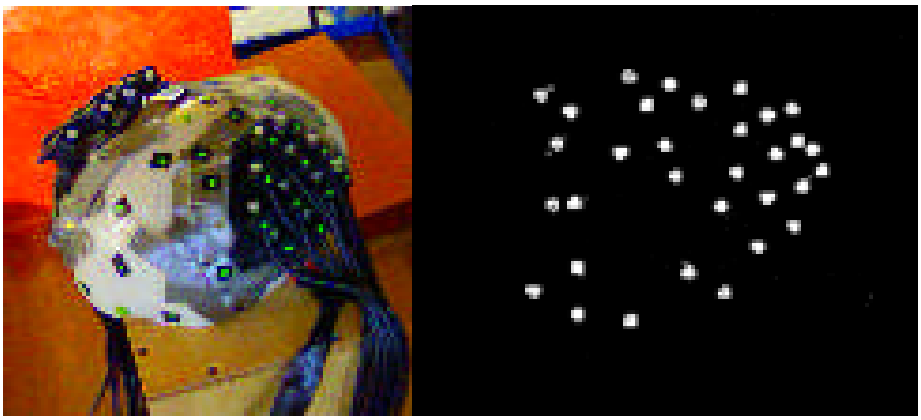


Figure 5 - 74. *Full target detection with no lighting control and no interference from strong sunlight. (Left: actual camera views; Right: corresponding target tracking)*

Figure 5 - 74 shows a scenario of full target detection while Figure 5 - 75 gives an example of large lighting variation (caused by strong sunlight) during a study. The influence from strong sunlight from time to time during the fifteen-minute period within a study not only caused loss of track of one or two targets on optodes across different frame sequences within a study (Table 5 - 10, Table 5 - 14 and Table 5 - 18), but also partly accounted for the variation in terms of the total number of targets being

coordinated by the camera system across a study (Table 5 - 7, Table 5 - 11 and Table 5 - 15). Such variations were also caused by the inconsistency of target detection resulting from the lower image quality attainable from Cam 1, 2, 3 and 8.

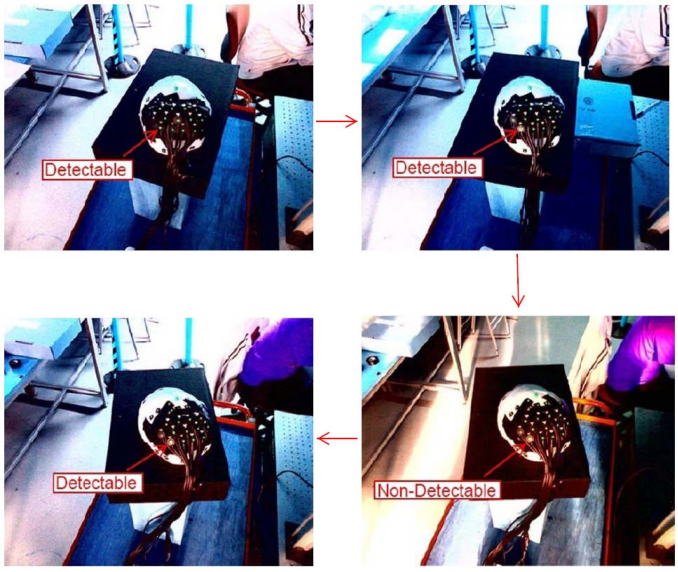


Figure 5 - 75. *Large variation of lighting occurred during the visual cortex study.*

Arrows between images represent the time sequence of the study.

Where sunlight caused an overexposure of a large area in the image (Figure 5 - 75), the colour target detection method can be particularly advantageous in comparison to the normal target detection method. This advantage is due to the fact that the image pixel values in the over-exposed area could be treated as irrelevant and therefore filtered out prior to target image detection even if they have high intensities.

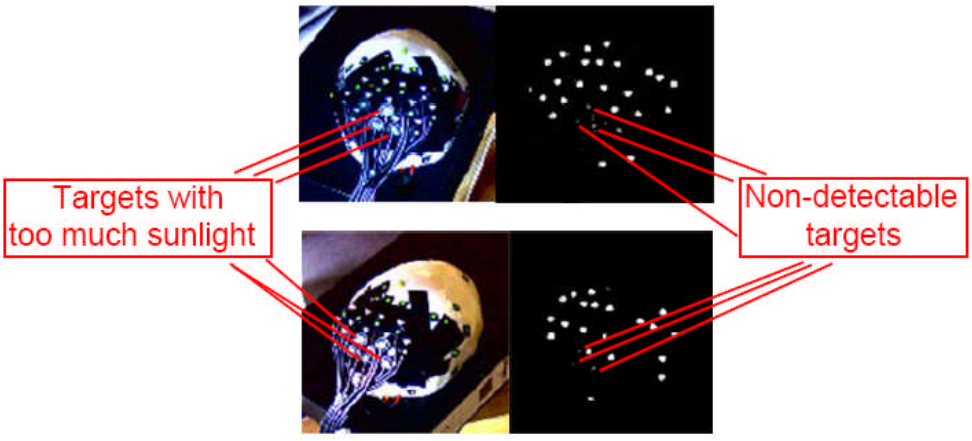


Figure 5 - 76. *Effect of strong sunlight on targets.*

(those recognized as targets are marked with white circles)

(Left: actual camera views; Right: corresponding target tracking)

The imaged colour distribution inside the real targets altered due to strong sunlight, as shown in the left image in Figure 5 - 76, and has resulted in the green component being less apparent. It is apparent from the right image in Figure 5 - 76 that some of the pixels inside these target images were recognised as target pixels. However, valid target image pixels with less apparent green component had been filtered out as background giving rise to edge disconnections. Since more than 50% of these pixels were under the influence of strong sunlight, the morphological operation (Section 3.4.4.3) was unable to fully reconnect the disconnected edges. These disconnected edges failed the edge tracing process (Section 3.4.4.4) and therefore the target image detection.

Despite this issue, target tracking consistency was found to remain consistent under room lighting provided that the target image distortion due to chromatic aberration is within tolerance of correspondence matching. This situation held even under high light intensities provided that a specular light reflection did not occur close to or at a target surface. Since most clinical environments contain room light whereas sunlight does not normally appear in such environments, the effect from sunlight should not be a concern for future studies.

As expected Cam 4, 5, 6 and 7 delivered much more reliable image qualities as compared with those from Cam 1, 2, 3 and 8. Image measurements processed with VMS represent the maximum number of target image measurements within each viewpoint, based on the lighting condition of the first set of static images captured. Image measurements successfully used for correspondence from the running camera tracking system for all temporal left (Table 5 - 8), temporal right (Table 5 - 12) and visual cortex (Table 5 - 16) studies are based on the lighting conditions at different moments during the studies. Results from these tables demonstrate that the reliable image qualities delivered by Cam 4, 5, 6 and 7 enabled these cameras to measure nearly all target images within viewpoints for correspondence. However, even under the tuning threshold adjustment and software image enhancement, the less reliable image qualities delivered by Cam 1, 2, 3, and 8 only allowed about half or less of the total number of target images being measurable within viewpoints for correspondence (highlighted as red in Table 5 - 8, Table 5 - 12 and Table 5 - 16).

Not only did unreliable image quality result in poor target image measurement performance, the target tracking consistency from these images was also much lower. Most targets oriented towards Cam 4, 5, 6 and 7 were consistently coordinated across frames (blue dots). However, most targets including those which sometimes were not coordinated (green dots) and those consistently not coordinated (pink dots) were oriented towards Cam 1, 2, 3 and 8 (Figure 5 - 23, Figure 5 - 36 and Figure 5 - 49). Furthermore, while some targets could be intersected with up to eight camera rays from VMS processing, the fact that all the consistently tracked targets could only be intersected with up to four camera rays (mainly from Cam 4, 5, 6 and 7) from the running tracking system (Figure 5 - 24, Figure 5 - 37 and Figure 5 - 50) once again demonstrates the inconsistency and unreliability in terms of target detection from Cam 1, 2, 3 and 8. For detailed explanations regarding the generation of these less reliable images, please refer to Section 4.4.3.3.

In contrast some targets were only measurable by a maximum of two camera rays from VMS processing based on the first set of images captured at the beginning of each study, whilst the same targets were measurable from more than two rays in certain frames from the running tracking system (Figure 5 - 24, Figure 5 - 37 and Figure 5 - 50). This situation once again demonstrates the large variation of lighting within these studies.

5.4.4.4 Speed

During this research, the developed system included descriptive text giving on-screen feedback during processing as well as OpenGL frame by frame rendering of the captured images frames to the computer screen. The resulting refresh rate of the tracking system (Section 4.6) was two FPS for a four-camera system and a half FPS for the eight-camera system. A test with four Logitech C500 webcams, where two were connected to USB 2.0 ports and the other two connected to USB 3.0 ports, was performed after the phantom and volunteer studies. The aim is to roughly demonstrate the speedup of the system by eliminating the output of unnecessary information on screen, as shown in Figure 5 - 77.

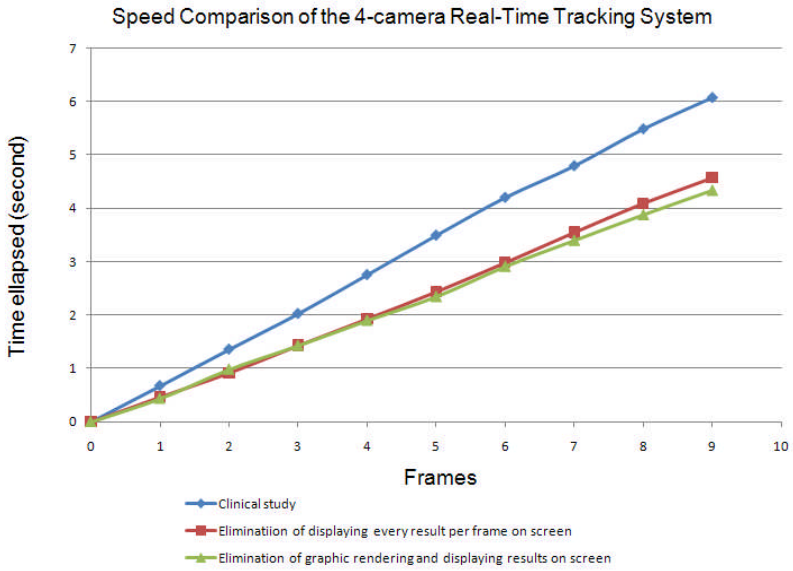


Figure 5 - 77. *Speedup of the live tracking after eliminating unnecessary screen output.*

Results demonstrate that about 25% of the processing time was dedicated to the on-screen output of results. After eliminating this information from the screen, about 1.5-second's time can be saved for every 10 consecutive frames, provided that in this independent test, 3D target coordinates were output to a text file for every single frame. By further eliminating the OpenGL graphic rendering of each frame from each camera onto the screen, the tracking system is able to perform even faster with a single desktop computer (Dell Precision 490 ~2008). With the current Intel Core i7 CPU technology, the tracking system can be easily brought to a single laptop making it more flexible and portable.

Another test with four Logitech C500 webcams, where two were connected to USB 2.0 ports and the other two connected to USB 3.0 ports, was also performed after the phantom and volunteer studies. In order to find out the percentage of time occupied by each main process for one 3D coordination loop, system counters were placed before and after each main process. Different number of targets was then used to find out the change of time percentage with the target number for each process (Figure 5 - 78). In terms of percentage of the total time elapsed for one loop, it was observed that the percentage for correspondence matching increased with the number of targets whilst slight decreases were observed for other processes. This latency should provide a reference for the future design and the associated coding optimization in terms of delivering faster 3D coordination for optical topography studies.

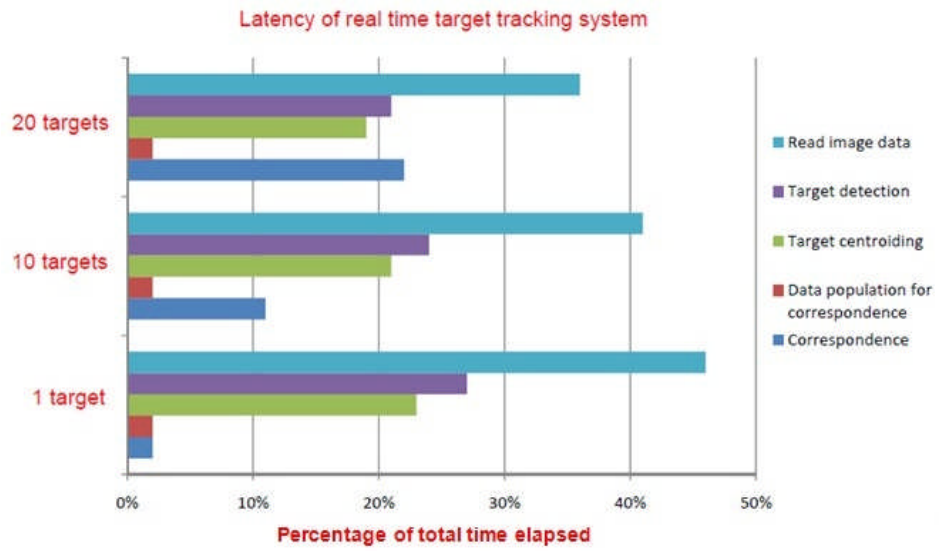


Figure 5 - 78. *Latency of the 4-Cam tracking system.*

5.4.4.5 Optical topography activation and localisation

Blasi *et al.* (2007) and Correia *et al.* (2009) emphasised that the low spatial accuracy is the main obstacle for the current state of the optical topography, i.e. poor ability to accurately locate the optical absoption change occurred, particularly in the depth direction. This agreed with all three positions (Figure 5 - 61) where the peak positions from the phantom activation map were found to be shallower (closer to the pad surface) than the positions estimated from photogrammetric co-registration. Results from optical topographic experiments with 4 month old infants showed that the activation occurred in the pad surface but without depth discrimination [Blasi *et al.*, 2007]. Changes in optical absorption over the left motor cortex at a depth averaged between 10 and 15mm was reported by Cooper *et al.* (2009). A quantitative assessment of the depth sensitivity of the same optical topography system carried out by Correia *et al.* (2009) using a solid dynamic tissue phantom containing seven discrete targets (similar to the phantom used in this research but with 4.5mm diameter and 41mm length) demonstrated that the spatial resolution (contrast) of the reconstructed image decreased almost linearly (exponentially) with depth, i.e. the distance from scalp surface. The sensitivity to depth is considered to be reasonably accurate within the 20mm depth range. The reconstructed image can easily loose its depth sensitivity if the activation occurs at a depth more than 20mm, in which this fact takes account of the lower image quality with depth [Correia *et al.*, 2009]. Moreover, the simultaneous use (i.e. near-infrared light

illumination) of multiple sources (modulated at different frequencies, Section 2.1.3.2) causes noise in the detected signal, where the larger the source-detector separation, the worse the signal [Blasi *et al.*, 2007]. Depth sensitivity of optical topography systems will continuously improve in near future with optimizing source-detector separations and sequentially illuminating sources [Blasi *et al.*, 2007]. Where further comparisons and verifications of activation localisation between optical imaging and photogrammetric approaches are necessary as future work, photogrammetric approaches will continuously support optical imaging to localize the activation area.

Unfortunately in the volunteer studies conducted for this thesis no useful optical data were collected for either subject. Movement artefacts (Figure 5 - 64) were reported as being a main cause for the high drop-out rate of optical topography studies [Blasi *et al.*, 2007]. The drop-out rate could be very high as reflected from the 60% drop-out rate reported by Taga *et al.* (2003), where they examined 20 infants foor visual activation, where 12 out of 20 were rejected because of either movement artefacts or poor contact due to hair. Therefore the invalid optical topography result from the two human studies in this research might also be attributable to the hair thickness of adults, which is much higher than that of infants. Besides the movement artefacts and contact issue due to hair thickness, other factors such as physiological noise from heart rate, respiration or vasomotion (the spontaneous oscillation in tone of blood vessels, independent of heart beat or respiration [Haddock and Hill, 2005]) can also interfere with the optical topography signals. These effects become more evident with higher skin thickness for adults as compared with infants [Goodwin *et al.*, 2004]. Moreover, the protocol used in human studies in this research was finger-tapping. Although activations on sensor-motor cortex from finger-tapping were reported by several researchers [Maki *et al.*, 1995; Boas *et al.*, 2001; Zhao *et al.*, 2006], optical imaging techniques are particularly well suited to imaging infants as demonstrated from the first report of optical topography on premature babies [Chance *et al.*, 1998]. Optical topography studies carried out by Kusaka *et al.* (2004) demonstrated different responses between adults and infants. While the optical sensing pad used in this research was primarily designed and optimized for infants by the Department of Medical Physics, the source-detector separation might not be well comparable for adults, with thicker skull and scalp.

5.4.4.6 Differential movement

For human subject studies, target coordinates derived from VMS intersection with an accuracy level of better than 0.2mm (Table 5 - 2) were used to better demonstrate the differential movement between the sensing pad and head.

For subject one, whilst the sequence of head centroids were co-registered onto the original position of the reference coordinate system (blue line), the left (red line) and right (green line) pad show different amounts of movement relative to the head after co-registration (Figure 5 - 64). The relative distance between the right pad and head increased to about 2.5mm towards the end of the study when subject one completely fell asleep (Figure 5 - 62). This differential movement of 2.5mm has exceeded the acceptable ~1mm accuracy for an optical topography study. As shown in Figure 5 - 64, the relative distance between the right pad and the head was over 1mm after the 4^{th} frame, and therefore for this study any data collected after this time interval must be disregarded since the assumption of OT is based on the temporal difference being due to brain activity alone.

For subject two, the maximum relative distance between head and left/right pad is below 0.04mm after co-registration (Figure 5 - 66). Since it was observed that subject two did not fall asleep at all during the whole study, the overall head movement of subject two is forward, right followed by backwards all with a small amount of distance by tracing back to the overall motion reconstruction (Figure 5 - 65). Theoretically all the centroid coordinates should have no discrepancies compared with those initial positions (as reference) after co-registration. However, a maximum discrepancy of 0.15mm away from origin was observed (Figure 5 - 66) and this magnitude of discrepancy in 3D space matches with that of VMS intersection shown in Table 5 - 2. Since the curves of all head centroid, left and right pads show the same shape across frames, it can be concluded that there is no differential movement between the pad and head for subject two during the functional study. Although the discrepancy away from the origin is attributable to the discrepancy of target 3D coordination from VMS intersection, this level of disagreement is well within the comparison data achieved with the phantom.

Blasi et al. (2007) addressed that the high drop-out rate due to movement artefacts is a major concern for optical topography experiments, where results became invalid due to the movement artefacts present in the OT signals. Moreover, the new design of optical sensing pad has made the optical fibres (optodes) approach the head at a 90 angle (Figure 5 –11), making the area of contact between the fibre pad and scalp small [Blasi *et al.*, 2007]. The susceptibility to movement artefacts of this design also addresses the need of live movement detection for future optical topography experiments.

5.4.5 Summary

In summary, the developed tracking system with a total imaging component cost of less than 400GBP has been successfully applied in optical topography studies. The following points can be made:

1. 3D target coordination accuracy delivered by the fast tracking system is much better than 1mm. If required 3D coordination capability could be further improved by adding additional cameras at different horizontal levels to have a more accurate coordination on Z-axis and fewer occlusions;
2. Important targets for optical topography studies have been consistently tracked and coordinated to sub-millimetre accuracy across image frames with no specific lighting control making the system highly suitable for the clinic;
3. Due to the necessity of checking of results and image frames during the optical topography studies performed in this research, a lot of CPU workload was dedicated to the output of results and image rendering onto the computer screen. As the system matures its speed can be enhanced by a better-written program as well as eliminating unnecessary display and checking;
4. In the optical topography studies carried out for this research, surface differential movement to sub-millimetre sensitivity has been detected. When this work was conducted the 3D similarity transformation between pad and head was performed after the tracking has completed. In future, this part of co-registration can be included in the automated tracking system making the detection of any differential movement between surfaces nearly real time. A warning could therefore be generated whenever the detected differential movement exceeds

the maximum range allowed. Not only this can provide a continuous validation for medical studies, since to the author's best knowledge there is currently no equipment to validate medical studies at patient's bedside such as optical topography, but it can also perform a high-accuracy monitoring for industrial manufacturing process making the production more efficient since any over-ranged movement can be immediately reported and corrected.

5. The allowance of free movement together with the live detection of differential movement between medical devices and underneath area, are advantageous over the MRI which strictly requires the tested subject to remain still during the study. Compared with the conventional brain functional studies such as watching a screen inside the MRI scanner, this has explored a new direction for the study design since brain functional studies involving body movements will not be limited any more.

Chapter 6 Conclusion and Future Work

This chapter first presents key results from experimental work which explicitly satisfy the research goals (Section 6.1), followed by a critical assessment of the limitations of the experimental method (Section 6.2) and discussions of possible future research directions resulting from this work (Section 6.3). Section 6.4 concludes this thesis.

6.1 Research results and conclusions

This section is divided into eight areas that address the key results from this research in response to the research aims (section 1.3) and objectives (Section 1.4) outlined at the commencement of this research.

1. **Suitability of webcams for optical topography**

   Webcam imagery from both webcam models (Logitech QuickCam Pro 4000 and C500) are capable of coordinating well defined targets to significantly better than the 1mm accuracy requirement. The advances of webcam technology have enabled clearer boundary between target image and background in favour of target image detection and measurement.

2. **Webcam stability**

   Webcam internal calibration parameters are sufficiently stable to reliably deliver 3D coordination over an extended period of time. Efforts regarding individual camera calibrations for each optical imaging study can therefore be minimised. Such stability has also demonstrated the consistency of target coordination accuracy for continuous 3D measurement during optical studies.

3. **Effective camera system calibration**

   It has been demonstrated that a self-calibrating bundle adjustment process is capable of effectively calibrating a system of multiple webcams, where only principal point, principal distance and 3rd power term of radial distortion are considered significant and therefore need to be included in the adjustment process. Taking the automatic changes of image quality (hidden from user resulting from different USB connections) into consideration, webcams should be calibrated as a system but not as individual sensors.

4. **Minimisation of infrastrature and cost**

   A camera system has been built under a total cost of £500 (excluding the host computer), which consists of eight Logitech C500 webcams, a bicycle wheel rim for mounting webcams, in-house built aluminium stand and rods ensuring lightness but stable exterior orientations of the camera system. Experimental results have demonstrated the construction is capable of rigidly defining the imaging geometry for an extended period of time.

5. **Target image detection**

   Recognition and measurement of target images have been proven as the key process for accurate 3D measurements. Two target image detection methods have been developed aiming at effectively extracting features from webcam imagery, under testing scenes with background clutter without specific lighting control. It has been demonstrated that both detection methods allow 3D target coordination significantly better than 1mm requirement. However a sub-millimetre (<0.2mm) 3D discrepancy between the two detection methods, resulted from a slightly different definition of the target centroid locations was found. Both calibration and measurement should therefore use a common target image measurement process.

6. **Target coordination accuracy**

   It has been demonstrated that for webcam imagery, that higher accuracy photogrammetric measurements can be attained with a better hardware resolution setting. Most important is to use a resolution that most closely matches the number of pixels on the image sensor. Studies have also demonstrated that with appropriate parameter recovery from a photogrammetric bundle adjustment process, a system of webcams is capable of 3D measurements with an absolute accuracy significantly better than 1mm. In particular, 3D measurements made between the webcam system and a calibrated pair of electronic calipers (standard deviation of 20μm) demonstrate an absolute 3D measurement accuracy of 0.18mm. Coordination results of phantom studies, as compared with VMS processing based on the static image set captured at the beginning of each study, are presented in Table 6 - 1.

Mean 3D coordinate discrepancy	Temporal left	Temporal right	Visual cortex
All coordinated targets (mm)	0.6 ~ 0.8	0.6 ~ 0.9	0.5 ~ 0.6
Key targets on optodes (mm)	0.4 ~ 0.6	0.5 ~ 0.8	0.4 ~ 0.5

Table 6 - 1. *Mean absolute 3D coordinate discrepancy for phantom studies.*

Factors affecting coordination accuracy have also been identified, including camera geometry, target locations on the spherical object, uncorrected target distortion and image compression. Following 3D measurements of optode positions and head surface landmarks, the individual optode positions can be effectively co-registered with the internal anatomical positions representing activation area by using 10-20 landmarks as common points.

7. **Coordinating speed and automation**

   Results have demonstrated that given appropriate threshold settings for target image detection, a fully automated 3D target coordination process is achievable with webcams. The multi-webcam system delivered 3D target coordination on per frame basis at 2FPS with 4 webcams (~0.5FPS with 8 webcams), provided with ~50 targets and a single host computer. With code optimisation and hardware advances, a faster coordination can be expected in near future.

8. **Detection of differential movement**

   Differential movements of 2.5mm and 0.15mm between optodes and scalp during optical imaging studies have been detected for two volunteer studies. These results demonstrate the capability of the camera system in terms of continuous target coordination and detection of differential movement which, undetected, would invalidate the topography experiments.

6.2 Critical assessment

Following the research conclusions presented in Section 6.1, a critical evaluation of the limitations of the experimental methods is demonstrated as follows:

- **Automated webcam calibration:** The main obstacle in bundle adjustment with low cost webcams in this research is the target image qualities for commonly

used photogrammetric coded targets, which are unreliable or must be made disproportionally large. With the implementation of the target image detection methods developed in this research into a static image measurement process, the VStars type codes (spatially distributed circular targets) should work automatically with webcam imagery.

- **Accuracy of 3D target coordination:** The developed multi-webcam tracking system successfully delivers fast 3D coordination to better than 1mm accuracy. However, webcams in the developed system were all mounted at the same horizontal level. As expected Z-coordinates, corresponding to depth, have a higher discrepancy compared with X- and Y-coordinates. To achieve higher 3D coordination accuracies, webcams would need to be located at two horizontal levels providing a stronger convergent geometry as well as increasing the number of camera rays intersecting targets (distributed over spherical objects). Please note that with the PCI express interface card (Figure 6 - 1) released towards the end of this research, tests have confirmed that all eight C500 webcams are able to deliver the same reliable image qualities solving problems with image qualities at the time of this research. With 4 such cards in a single host computer, a system consisting of 16 webcams (8 per horizontal level) can be built to achieve full target coverage and higher coordination accuracies.

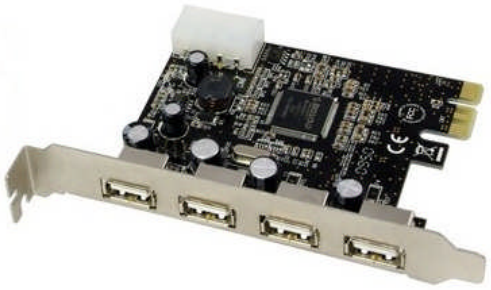


Figure 6 - 1. *Logilink PCI Express to USB2.0 host controller card [Logilink].*

(4 individual controllers per card)

Moreover, calibration of Logitech C510 webcam (released 2010) has demonstrated promising results of the current webcam technology. The C510 series delivers HD imagery at the same price as C500 series used in this research and delivers RMS Image Residuals of 1/13th pixel (240nm) to a retro-reflective target. While comparing to Nikon D100 measurement of the

calibration object, it achieves mean absolute discrepancies of 0.014mm (X), 0.014mm (Y) and 0.019mm (Z), with a discrepancy standard deviation of 0.019mm (X), 0.019mm (Y) and 0.023mm (Z), which demonstrates a nearly 10 times higher accuracy of 3D coordination than C500 webcams. Eight images of calibrated C510 webcam are able to reconstruct a head mesh with 27040 point clouds (Figure 6 - 2) under ~15 minutes with MeshLab [MeshLab] utilising PMVS [Furukawa and Ponce, 2010], Bundler [Snavely, 2006] and CMVS [Furukawa *et al.*, 2010] from computer vision community. However, the accuracy of image measurements and 3D coordinates of the generated point clouds has not been fully assessed and verified. Therefore future research investigating the 3D accuracy of these methods is necessary in order to reliably reconstruct a head surface mesh for optical topographic image reconstruction.

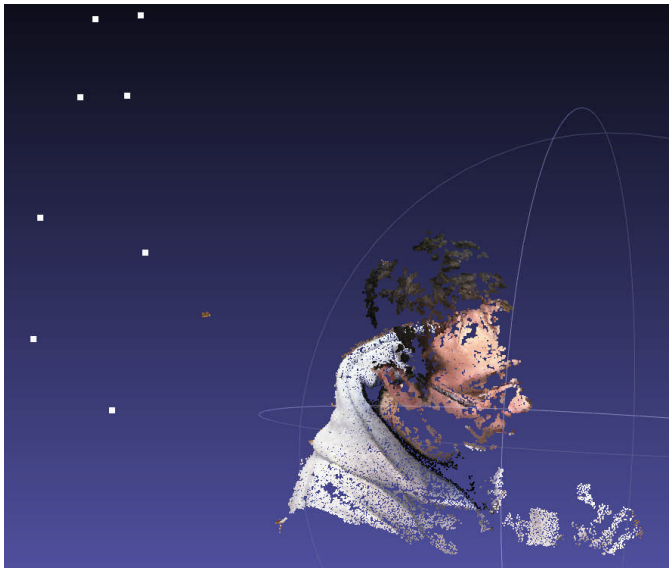


Figure 6 - 2. *27040 mesh points reconstructed from 8 images Logitech C510 webcams. (image resolution: 1600×1200, pixel size: 3.2μm) White dots show 8 positions where images were taken.*

- **Movement of tracked object:** In this research it is assumed that movement of the coordinated object is small compared to the averaged time delay (~0.005-0.007s) between image frames (captured by different webcams) within one frame set. However, completely synchronised camera images would be necessary if its application is extended to coordinate a fast moving object in future. Such a system would require a gunlock signal which is not possible on webcam systems, but is available on board level camera modules.

- **Range of webcams:** This research mainly investigated two generations of Logitech webcams representing the mainstream webcam technologies at the commencement and towards the end of this research respectively. Therefore results should only apply for the webcams investigated in this research, with either manual focus or fixed focus. Please note that results from this research do not apply for other webcams especially those with automated zoom and autofocus which are unlikely to be stable. In the reverse, the results and conclusions drawn from this research could probably be extended to board cameras which are similar in construction to the investigated webcams. Full tests and comparisons of accuracy, precision and reliability across different webcam brands/models (with manual or fixed focus) should be conducted drawing a more representing conclusion about the webcam suitability for accurate close range photogrammetric measurements.

6.3 Future research

This thesis has presented an investigation of the suitability of webcam photogrammetry for optical topography studies followed by the system development of multiple webcams and its associated applications in optical topography studies. The presented webcam photogrammetric system can be extended to any close range measurement tasks without specific lighting control demanding sub-millimetre accuracy. Slow object motion is allowed and any target(s) can be removed and added from the continuous coordination process (such as industrial manufacturing process, continuous monitoring of cracks inside pipes, continuous medical monitoring at patient's bedside). Suggestions for future research in terms of optimisation of the current method that would improve the performance as well as extend the current framework in the context of method application are presented as follows:

6.3.1 Method extension

- **Featurization of object:** Different coloured targets can be used together with the developed tuning target detection method (subject to correction of chromatic distortion). This can be implemented through either passive targets (used in this

research) or active targets (e.g. illuminated diodes manufactured by NDI [NDI]). For example, green dots can be used for optodes and red dots can be used for bony landmarks in which their computed 3D coordinates can be distinguished from each other during the continuous coordination. Centroids of optodes and scalp can be computed and transformed running each frame set during the monitoring process. This would eliminate any post-processing work by technicians dedicated to find out any differential movement between medical device and human skin, or any kinds of movement which should be prevented or minimised.

- **Auto-tuning process:** An automated tuning process for target image detection can be designed and implemented in the coordination system. This is done by gradually adjusting the tuning parameters (thresholds for target image detection based on specific light condition and background clutter) until an optimal number of targets are found within camera views. Though the coordination process might not be able to start immediately with this auto-tuning process, the target coordination can however work in different environments with much more flexibility.
- **Intersection of only new or moved targets:** It has been shown that the percentage time for 3D correspondence increases with the number of coordinated targets. The correspondence solution can be extended so that the tracking system only intersects targets which are newly added or moved (larger than a pre-set threshold). This includes the correspondence solution for all detected target images for the first frame set, followed by back drive of 3D positions derived from previous frame (temporarily stored in memory) to 2D image coordinates of the following frame set to check if there are any newly detected target images. While newly coordinated targets are added to and those failing the back drive are removed from the memory (which stores the most updated 3D target coordinates), 3D correspondence will only be computed for those newly detected target images after the first frame set. The frame rate of the tracking system in this way might be significantly increased subject to further tests and verification.
- **Functional localisation for human subjects**: The newly generated Logitech C510 webcams demonstrates a nearly 10 times higher accuracy of 3D

coordination as compared with the C500 webcams. Such a significant increase should allow even higher webcam-based coordination accuracy in optical topography studies. Co-registration accuracies between the optodes and the internal anatomical locations representing brain activation areas should also benefit from such a big jump. Furthermore, since optical topographic reconstruction is based on optodes' positions relative to head mesh, accuracy of functional localisation with the mesh generated with webcams (Figure 6 - 2) could be compared with the current technique used, where a generic head mesh is wrapped to align the optodes in order to create an individual head mesh.

The above considerations are critical in terms of the implementation, robustness, efficiency and practical usability of the low cost webcam photogrammetry.

6.3.2 Application extension

Following the results presented in Section 6.2, 3D coordination to sub-millimetre from a system built with multiple low cost webcams is appropriate for applications as follows:

- **Removal of movement restrictions in medical studies:** Taking advantage of the compactness, portability of low cost webcams, the multi-webcam system can easily be transported and fitted to any medical environment. Where promising improvements are shown over webcam generations, coordination discrepancies in microns (or tens of microns) between webcams and high end cameras such as the D100 is achievable in near future with the same or lower price point. Allowing movement in medical studies should allow the development of more robust medical devices (e.g. mobile optical topographic sensing pad) and more experiment protocols to be designed. For example, a brain functional study can be designed in the sense that a mobile pad is placed on various scalp locations to examine if more than one region of cerebral cortex are simultaneously activated by certain stimulation(s). Different pad positions coordinated by the system during the study can directly be used for the following registration process. It is also possible to build a low cost tracking system for MRI imaging solving the movement problem during MRI scans.
- **Pattern recognition:** It is expected that a multi-layer webcam-based

coordination system can be built to cover targets distributed all over a head shaped object. The fast and accurate coordination all around the head would support pattern recognition applications such as facial expression categorization, in which the movements of scalp and/or neck (besides face) can all be taken into account for analysis.

- **Long-term monitoring:** Provided with the webcam stability lasting for an extended period, a multi-webcam system can be used for studies such as the growth rate of flowers or certain species (biology field) in which object features can be coordinated in certain time interval, e.g. once per week. Data collected at the end of studies should directly draw a conclusion of the growth rate. The continuous monitoring should allow the system to be used in studies such as object deformation over time or under a certain temperature (engineering field), where deformation curves over time can be constituted in a similar sense.

6.4 A final point

At the outset of this research it was unknown whether low cost cameras would offer accurate 3D measurements suited to optical topography studies. This research has successfully proven the suitability of webcams for optical topography studies in the sense that a series of webcams are able to deliver 3D coordination of locations on the skin and optical sensing pads to better than 1mm accuracy and maintain calibration for extended periods. A fast multi-webcam photogrammetric coordination system which works in cluttered environments and without specific light control has been developed. In terms of medical applications, the system allows a low cost and simple setup within minimal infrastructure, operates in room light without interfering with laser/infrared illumination, and delivers a fast process to monitor initial pad locations, any differential movement (between skin and pad) and in future tracking of a mobile sensing pad. Where image quality combined with robust target detection methods are key to the coordination performance, low cost cameras are capable of providing a suitable platform for fast photogrammetric coordination given attention to camera and host computer design. Initial tests with the next generation C510 webcams delivering HD imagery are extremely promising and more applications of low cost webcam photogrammetry can be foreseen in the research community.

References

Abdel-Aziz, Y.I. & Karara, H.M., 1971. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. Proceedings of the Symposium on Close-Range Photogrammetry. Falls Church, VA: American Society of Photogrammetry, pp. 1-18.

Abraham, S. & Hau, T., 1997. Towards autonomous high precision calibration of digital cameras. Proceedings of SPIE Videometrics V, El-Hakim (Ed.), Vol. 3174, pp. 82-93.

Ahmad, Anuar, Setan, Halim, Majid, Zulkepli, Chong & Albert, 2004. Calibration of digital camera for medical photogrammetry. In: International Symposium and Exhibition on Geoinformation (ISG), Kuala Lumpur, Malaysia, pp. 21-23.

Ariyawansa, D.D.A.P. & Clarke, T.A., 1997. High speed correspondence for object recognition and tracking. Videometrics V, San Diego, USA, 30-31 July, Proc. SPIE 3174, pp. 70-79.

Arridge, S.R., 1999. Optical tomography in medical imaging. Inverse Problems 15 R41–R93.

Arridge, S.R., Dehghani, H., Schweiger, M. & Okada, E., 2000. The finite element model for the propagation of light in scattering media: a direct method for domains with nonscattering regions. *Med. Phys.* 27, pp. 252–64.

Arridge, S.R. & Schweiger, M., 1995. Photon measurement density functions: II. Finite element method calculations. *Appl. Opt.* 3480, pp. 26–37.

Arridge, S.R., Schweiger, M., Hiraoka, M. & Delpy, D.T., 1993. Finite element approach for modeling photon transport in tissue. *Med. Phys.* 20, pp. 299–309.

Atkinson, K.B., 1996. Close range photogrammetry and machine vision. Whittles Publishing, Caithness, Scotland, UK.

Atkinson, K.B., 2001. Close Range Photogrammetry and Machine Vision. Whittles Publishing, ISBN: 1-870325-73-7.

Aw, Y.B. & Koo, T.B., 1993. Phototriangulation using CCD Camera in Close-range Environment. Jounal of Surveying Engineering. Vol. 119. No 2., pp. 52-57.

Azuma, R. & Bishop, G., 1994. Improving static and dynamic registration in an optical see-through hmd. 21st International SIGGRAPH Conference, 24-29 July, Orlando, FL, USA, pp.197–204.

Azuma, R. & Bishop, G., 1995. A frequency-domain analysis of head-motion prediction. In R. Cook, editor, SIGGRAPH ’95, 6-11 Aug., Los Angeles, CA, USA, pp. 401–408.

Baltsavias, E.P., 1991. Multiphoto geometrically constrained matching. PhD Thesis, pp. 1-40.

Barnett, G.H., Kormos, D.W., Steiner, C.P. & Morris, H., 1993. Registration of EEG electrodes with three-dimensional neuroimaging using a frameless, armless stereotactic wand. Stereotactic and Functional Neurosurgery 61, pp. 32-38.

Bauer, H., Lamm, C., Holzreiter, S., Holländer, I., Leodolter, U. & Leodolter, M., 2000. Measurement of 3D electrode coordinates by means of a 3D photogrammetric head digitizer. *Neuroimage* 11: S461.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Communication of the ACM 18 (9), pp. 509-517.

Beyer, H., 1993. Determination of Radiometric and Geometric Characterisitics of Frame Grabbers. SPIE Vol. 2067. Videometrics II., pp. 93-103.

Blasi, A., Fox, S., Everdell, N., Volein, A., Tucker, L., Csibra, G., Gibson, A.P., Hebden, J.C., Johnson, M.H. & Elwell, C.E., 2007. Investigation of depth dependent changes in cerebral haemodynamics during face perception in infants. Physics in Medicine and Biology 52, pp. 6849-6864.

Boas, D.A., Gaudette, T., Strangman, G., Cheng, X., Marota, J.J.A. & Mandeville, J.B., 2001. The Accuracy of Near Infrared Spectroscopy and Imaging during Focal Changes in Cerebral Hemodynamic. NeuroImage 13, pp. 76-90.

Böcker, K.B.E., van Avermaete, J.A.G. & van den Berg-Lennsen, M.M.C., 1994. The international 10-20 system revisited: cartesian and spherical co-ordinates. Brain Topography 6, pp. 231-235.

Brett, M., Johnsrude, I.S. & Owen, A.M., 2002. The problem of functional localization in the human brain. Nat. Rev., Neurosci. 3, pp. 243–249.

Brinkmann, B.H., O'Brien, T.J., Dresner, M.A., Lagerlund, T.D., Sharbrough, F.W. & Robb, R.A., 1998. Scalp-recorded EEG localization in MRI volume data. Brain Topography 10, pp. 245-253.

Brodmann, K., 1905. Beiträge zur histologischen Lokalisation der Grosshirnrinde: dritte Mitteilung: Die Rindenfelder der niederen Affen. Journal für Psychologie und Neurologie 4, pp. 177-226.

Brodmann, K., 1909. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. J.A. Barth, Leipzig (In German).

Brodmann, K., 1912. Neue Ergebnisse u¨ber die vergleichende histologische Lokalisation der Grosshirnrinde. Anat. Anz. 41, pp. 157– 216 (In German).

Brown, D., 1984. Tools of the trade. Close range photogrammetry and surveying - state of the art. Pub. The American Society of Photogrammetry. 941 pages, pp. 83-252.

Brown, L.G., 1992. A survey of image registration techniques. ACM Computing Surveys, 24(4), pp. 325-376.

Brown, D., 1971. Close-Range Camera Calibration. Photogrammetric Engineering, 37(8), pp. 855-866.

Brown, D., 1976. The bundle adjustment - progress and prospects. Invited paper, Commission III, Proc. Of XIIIth ISP Congress. International Archives of Photogrammetry 21(3), 33 p.

Bruce, H.W., 1978. Understanding secondary color. Optical Spectra 12, pp.44-46.

Buchsbaum, W.H., 1975. Color TV Servicing, third edition. Englewood Cliffs, NJ: Prentice Hall.

Burner, A.W., Snow, W.L., Shortis, M.R. & Goad, W.K., 1990. Laboratory calibration and characterization of video cameras. SPIE. Vol. 1395. Close Range Photogrammetry Meets Machine Vision. pp. 664-671.

Calder, A.J., Lawrence, A.D. & Young, A.W., 2001. Neuropsychology of Fear and Loathing. Nature Reviews Neuroscience, Vol.2 No.5, pp. 352-363.

Canny, J.F., 1986. A computational approach to edge detection. IEEE Trans pattern Analysis and machine Intelligence, 8, pp. 679-698.

Canon, 2009. Technology explained - digital SLR: http://www.canon.co.uk/Images/EOS-1D_Mark_IV_Technologies_Explained_tcm14-687344.pdf

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D. & Thomas, R., 1998. A novel method for fast imaging of brain function, non-invasively, with light. Optics Express 2, pp. 411-423.

Chen, J. & Clarke, T.A., 1992. The automatic location and identification of targets in digital photogrammetric engineering measurement. Int. Archives of Photogrammetry and Remote Sensing, XVII(5), pp. 686-693.

Chen, J., Clarke, T.A., Cooper, M.A.R. & Grattan, K.T.V.G., 1995. An optimised target matching based on a 3-D space intersection and a constrained search for multiple camera views, Videometrics III. SPIE Vol. 2350, pp. 324 - 335.

Chen, J., Clarke, T.A. & Robson, S., 1993. An alternative to the epipolar method for automatic target matching matching in multiple images for 3-D measurement. Optical 3-D measurements techniques II, Pub. Wichmann, Karlsruhe, pp. 197-204.

Chen, Y. & Medioni, G., 1992. Object modelling by registration of multiple range images. Image and Vision Computing 10(3), pp. 145-155.

Cheong, W.F., Prahl, S.A. & Welch, A.J., 1990. A review of the optical properties of biological tissues. IEEE Journal of Quantum Electron. 26, pp. 2166-2185.

Chetverikov, D., 1991. Fast neighborhood search in planar point sets. Pattern Recognition Letters 12(7), pp. 409-412.

Clarke, T.A., Cooper, M.A.R. & Fryer, J.C., 1993. An estimator for the random error in subpixel target location and its use in the bundle adjustment. Optical 3-D Measurement Techniques II, Ed. A. Gruen & H. Kahmen, Pub. Wichmann, Karlsruhe, pp. 161-168.

Clarke, T.A., 1994. An analysis of the properties of targets uses in digital close range photogrammetric measurement. Videometrics III. Boston. SPIE Vol. 2350, pp. 251- 262.

Clarke, T.A., Cooper, M.A.R., Chen, J. & Robson, S. 1994. Automated 3-D measurement using multiple CCD camera views. Photogrammetric Record. Vol. XV. No 86, pp. 315-322.

Clarke, T.A. & Fryer, J.G., 1998. The development of camera calibration methods and models. The Photogrammetric Record, Vol. 16(91), pp. 51-66.

Clarke, T.A., Robson, S., Qu, D.N., Wang, X., Cooper, M.A.R. & Taylor, R.N., 1995. The sequential tracking of targets in a remote experimental environment. International Archives of Photogrammetry and Remote Sensing, 30(5W1), pp. 80–85.

CNN review, 2011: http://reviews.cnet.com/webcams/logitech-quickcam-pro-4000/4505-6502_7-20276742.html?tag=rnav#reviewPage1

Collins, D.L., Neelin, P., Peters, T.M. & Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, pp. 192–205.

Cooper, R.J., Everdell, N.L., Enfield, L.C., Gibson, A.P., Worley, A. & Hebden, J.C., 2009. Design and evaluation of a probe for simultaneous EEG and near-infrared imaging of cortical activation. Physics in Medicine and Biology 54, pp. 2093-2102.

Cope, M., 1991. The application of near-infrared spectroscopy to non-invasive monitoring of cerebral oxygenation in the newborn infant. PhD thesis. University College London.

Correia, T., Banga, A., Everdell, N.L., Gibson, A.P. & Hebden, J.C., 2009. A quantitative assessment of the depth sensitivity of an optical topography system using a solid dynamic tissue-phantom. Phys. Med. Biol. 54, pp. 6277-6286.

Cronk, S., Fraser, C.S. & Hanley, H.B., 2006. Automatic Calibration of Colour Digital Cameras. Photogammetric Record. Vol. 21, Issue 116, pp. 355–372.

Crossman, A.R. & Neary, D., 1995. Neuroanatomy an illustrated colour text. 1st ed. Edinburgh: Churchill Livingstone.

CUDA, 2007. NVIDIA CUDA Compute Unified Device Architecture Programming Guide, V. 1.0.

Curry, S., Baumrind, S. & Anderson, J., 1986. Calibration of an Array Camera. Photogrammetry Engineering & Remote Sensing, 52(5), pp.627-636.

Cyberware: http://www.cyberware.com/.

Dahler, J., 1987. Problems in Digital Image Acquisition with CCD cameras. Intercommission Conference on Fast Processing of Photogrammetric Data, pp. 48-59.

Darrell, T., Maes, P., Blumberg, B. & Pentland, A.P., 1994. A novel environment for situated vision and behavior. In Workshop on Visual Behaviors, 19 June, Seattle, WA, USA, pp.68–72.

Debevec, P., Taylor, C.J. & Malik, J., 1996. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. SIGGRAPH'96, pp. 11–20.

Deng, T., 1987. Accuracy of position estimation by centroid. Intelligent Robots and Computer Vision, SPIE Vol. 848, pp. 141-150.

Dewitt, B.A., 1996. Initial approximations for the three-dimensional conformal coordinate transformation. Photogrammetric Engineering & Remote Sensing, 62(1), pp. 79-83.

DGAP, 2005:
http://www.ifp.uni-stuttgart.de/publications/software/openbundle/index.en.html.

Dimitrijevic, M., Lepetit, V., Fua, P., 2006. Human Body Pose Detection Using Bayesian Spatio-Temporal Templates. Computer Vision and Image Understanding, 104(2-3), pp.127–139.

DirectShow: http://msdn.microsoft.com/en-us/library/dd375454%28v=vs.85%29.aspx.

Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F.N. & Emslie, H., 2000. A neural basis for General Intelligence. Science 289 (5478), pp. 457-460.

dwBitRate definition, 2011: http://msdn.microsoft.com/en-us/library/dd407325%28v=vs.85%29.aspx.

Efford, N., 2000. Digital Image Processing - a practical introduction using Java, Addison Wesley.

El-Hakim, S.F., Beraldin, J.A. & Blais, F., 2003. Critical factors and configurations for practical image-based 3D modeling. Proceedings of 6th Conference Optical 3D Measurements Techniques. Zurich, Switzerland. Vol. II, pp. 159-167.

El-Hakim, S.F., 1986. Real-time image metrology with CCD cameras. Photogrammetric Engineering and Remote Sensing, 52(11), pp. 1757-1766.

El-Hakim, S.F., 2001. A flexible approach to 3D reconstruction from single images. ACM Proceedings of SIGGRAPH, Technical Sketches, Los Angeles, California, pp. 186.

El-Hakim, S.F., 2002. Semi-automatic 3D reconstruction of occluded and unmarked surfaces from widely separated views. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part 5, pp. 143-148.

Evans, A.C., Collins, D.L. & Milner, B., 1992. An MRI-based stereotactic atlas from 250 young normal subjects. Journal Soc. Neurosci. Abstr. 18, pp. 408.

Evans, A.C., Collines, D.L., Mills, S.R., Brown, E.D., Kelly, R.L. & Peters, T.M., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference, pp. 1813-1817.

Evans, A.C., Kamber, M., Collins, D.L. & Macdonald, D., 1994. An MRI-based probabilistic atlas of neuroanatomy. Magnetic Resonance Scanning and Epilepsy (NATO ASI Series A, Life Sciences), volume 264, Edited by Shorvon, S., Fish, D., Andermann, F., Bydder, G.M. & Stefan, H., Plenum Press, pp. 263-274.

Everdell, N.L., Gibson, A.P., Tullis, I.D.C., Vaithianathan, T., Hebden, J.C. & Delpy, D.T., 2005. A frequency multiplexed near-infrared topography system for imaging functional activation in the brain. Review of Scientific Instruments 76, 093705.

Faugeras, O. & Mourrain, B., 1995. On the geometry and algebra of the point and line correspondences between N images. Research Report No-2665, INRIA Sophia-Antipolis, BP 93 Sophia-Antipolis, Cedex, France.

Fischl, B., Sereno, M., Tootell, R. & Dale, A., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping, 8, pp. 272-284.

Fitzgibbon, A.W. & Zisserman, A., 1998. Automatic 3D Model Acquisition and Generation of New Images from Video Sequences. European Signal Processing Conference, pp. 1261-1269.

Fleishman, S., Cohen-Or, D. & Lischinski, D., 2000. Automatic camera placement for image-based modeling. Computer Graphics Forum, 19(2), pp.101 - 110.

Förstner, W., Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Cicular Features. Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland. pp. 281-305.

Fraser, C.S. & Al-Ajlouni, S., 2006. Zoom-dependent camera calibration in digital close-range photogrammetry. Photogramm. Eng. Remote Sensing 72, pp. 1017–1026.

Fraser, C.S., 1980. Multiple Focal Setting Self-Calibration of Close Range Cameras. Photogrammetric Engeering & Remote Sensing, 46(11), pp.1439-1445.

Fraser, C.S., 1997. Digital camera self-calibration. ISPRS, Journal of Photogrammetric & Remote Sensing, 52, pp. 149-159.

Fryer, J.G., 1988. Lens Distortion and Film Flattening: their Effect on Small Format Photogrammetry. International Archives of Photogrammetry and Remote Sensing 27(5), pp. 194-202.

Fryer, J.G., Clarke, T.A. & Chen, J., 1994. Lens distortion for simple 'C' mount lenses". International Archives of Photogrammetry and Remote Sensing, 30(5), pp. 97-101.

Fryer, J.G., Mitchell, H. & Chandler, J., 2007. Applications of 3D Measurement from Images. Whittles Publishing: Scotland, UK.

Furukawa, Y., Curless, B., Seitz, S.M. & Szeliski, R., 2010. Towards Internet-scale Multi-view Stereo. Computer Vision and Pattern Recognition, pp. 1434-1441.

Furukawa, Y. & Ponce, J., 2010. Accurate, Dense, and Robust Multi-View Stereopsis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 32(8).

Gavrila, D.M. & Davis, L.S., 1996. 3D Model-Based Tracking of Humans in Action: A Multi-View Approach. IEEE Conf on Computer Vision and Pattern Recognition, pp. 73-80.

Gerth, B., Berndt, R., Havemann, S. & Fellner, D.W., 2005. 3D modeling for non-expert users with the castle construction Kit v0.5t. 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage - VAST, Mudge/Ryan/Scopigno (Eds), pp. 1-9.

Gevins, A., Le, J., Brickett, P., Reutter, B. & Desmond, J.E., 1991. Seeing through the skull: advanced EEGs use MRIs to accurately measure cortical activity from the scalp. Brain Topography 4, pp. 125-131.

Gevins, A., Le, J., Martin, N.K., Brickett, P., Desmond, J. & Reutter, B., 1994. High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods. Electroencephalography and Clinical Neurophysiology 90, pp. 337-358.

Gibson, A.P., Hebden, J.C. & Arridge, S.R., 2005. Recent advances in diffuse optical imaging. Physics in Medicine and Biology 50, R1-R43.

Gibson, A.P., Riley, J., Schweiger, M., Hebden, J.C., Arridge, S.R. & Delpy, D.T., 2003. A method for generating patient-specific finite element meshes for head modeling. Phys. Med. Biol. 48, pp. 481– 495.

Goktepe, A. & Kocaman, E., 2011. Using Direct Linear Transformation Method in X-Ray Photogrammetry and an Illustrative Study. Experimental Techniques. doi: 10.1111/j.1747-1567.2011.00740.x.

Gonzalez, R.C. & Woods, R.E., 1992. Digital Image Processing. Addison Wesley Publishing Co.

Goodwin, J.A., van Meurs, W.L., Sa Couto, C.D., Beneken, J.E.W & Graves, S.A., 2004. A model for educational simulation of infant cardiovascular physiology. Anesth. Analg. 99, pp. 1655–1664.

Gottschalk, S. & Hughes, J.F., 1993. Autocalibration for virtual environments tracking hardware. In ACM SIGGRAPH 1993, pp. 65–72.

Granshaw, S.I., 1980. Bundle adjustment methods in engineering photogrammetry. Photogramm. Rec. 10, pp. 181–208.

Grimson, W.E.L., 1990. The combinatorics of object recognition in cluttered environments using constrained search. Artificial Intelligence Journal 44(1-2), pp. 121–166.

Gruen, A., 1976. Die simultane Kompensation systematischer Fehler mit dem Münchner Bündelprogramm MBOP. International Archives of Photogrammetry and Remote Sensing, Vol. 21, Part 3, Presented Paper, Commission III/1, ISP Congress, Helsinki.

Gruen, A., 1978. Accuracy, Reliability and Statistics in Close-Range Photogrammetry. In: Symposium of ISP, Commission V, Stockholm, Sweden.

Gruen, A., 1981. Precision and reliability aspects in close-range photogrammetry. Photogrammetric Journal of Finland, Vol. 8(2), pp. 117-132.

Gruen, A., 2000. Semi-automated approaches to site recording and modeling. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 33, Part 5/1, pp. 309-318.

Gruen, A. & Akca, D., 2005. Least squares 3D surface and curve matching. ISPRS Journal of Photogrammetry and Remote Sensing 59 (3), pp. 151-174.

Gruen, A. & Baltsavias, E., 1988. Geometrically constrained multiphoto matching. Photogrammetric Engineering and Remote Sensing, 54(5), pp.633-641.

Gruen, A. & Beyer, H.A., 2001. System calibration through self-calibration. In 'Calibration and Orientation of Cameras in Computer Vision' Gruen and Huang (Eds.), Springer Series in Information Sciences 34, pp. 163-194.

Gruen, A., Zhang, L. & Visnovcova, J., 2001. Automatic reconstruction and visualization of a complex Buddha Tower of Bayon, Angkor, Cambodia. Proceedings 21. Wissenschaftlich-Technische Jahrestagung der DGPF, pp. 289-301.

Gulch, E., 1986. Calibration of CCD Video Cameras. Proceedings of ISPRS Symposium on Progress in Imaging Sensors, Stuttgart, pp. 391-403.

Haddock, R.E. & Hill, C.E., 2005. Rhythmicity in arterial smooth muscle. J Physiol (Lond)., 566: 645-656.

Haritaoglu, I., Harwood, D. & Davis, L.S., 1998. W4: Who, when, where, what: a real time system for detecting and tracking people. 3rd IEEE Int. Conf. Automatic Face and Gesture Recognition, Nara, Japan.

Hartley, R., 2000. Ambiguous configurations for 3-view projective reconstruction. European Conference of Computer Vision, Dublin, Ireland, Springer-Verlag, pp. 922-935.

Hebden, J.C. 2003. Advances in optical imaging of the newborn infant brain. Psychophysiology 40, pp. 501–10.

Hebden, J.C., Brunker, J., Correia, T., Price, B.D., Gibson, A.P. & Everdell, N., 2008. An electrically-activated dynamic tissue-equivalent phantom for assessment of diffuse optical imaging systems. Physics in Medicine and Biology 53, pp. 329-337.

Hillman, E., 2002. Experimental and theoretical investigations of near infrared tomographic imaging methods and clinical applications. PhD thesis. University College London.

Hilton, A., 1999. Towards model-based capture of a persons shape appearance and motion. In IEEE International Workshop on Modelling People, 20 Sept., Kerkyra, Greece, pp.37–44.

Hogg, D.C., 1984. Interpreting Images of a Known Moving Object. PhD thesis, University of Sussex, UK.

Homan, R.W., Herman, J. & Purdy, P., 1987. Cerebral location of international 10 – 20 system electrode placement. Electroencephalogr. Clin. Neurophysiol. 66, pp. 376– 382.

Huang, Y. & Trinder, J.C., 1993. A procedure for fully automated orientation of camera in digital close range photogrammetry. Optical 3-D Measurement Techniques II, Ed. A. Gruen & H. Kahmen, Pub. Wichmann, Karlsruhe, pp. 339-346.

Huppertz, H.J., Otte, M., Grimm, C., Kristeva-Feige, R., Mergner, T. & Lücking, C.H., 1998. Estimation of the accuracy of a surface matching technique for registration of EEG and MRI data. Electroencephalography and Clinical Neurophysiology 106, pp. 409-415.

Ives, J.R., Warach, S., Schmitt, F., Edelman, R.R. & Schomer, D.L., 1993. Monitoring the patient's EEG during echo planar MRI. Electroencephalography and Clinical Neurophysiology 87: 417-420.

Iwai, Y., Ogaki, K. & Yachida, M., 1999. Posture estimation using structure and motion models. In Seventh IEEE International Conference on Computer Vision, 20-27 Sept., Kerkyra, Greece, Vol. 1, pp. 214–19.

Jiang, A., Kennedy, D.N., Baker, J.R., Weisskoff, R.M., Tootel, R.B.H. & Woods, R.P., 1995. Motion detection and correction in functional MR imaging. Human Brain Mapping, 3, pp. 1-12.

Jordan, H.J., Wegner, M. & Tiziani, H., 1998. Highly accurate non-contact characterization of engineering surfaces using confocal microscopy. Meas. Sci. Technol. 9, pp. 1142–1151.

Jurcak, V., Okamoto, M., Singh, A.K. & Dan, I., 2005. Virtual 10-20 measurement on MR images for inter-modal linking of transcranial and tomographic neuroimaging methods. Neuroimage, vol. 26, pp. 1184-1192.

Jurcak, V., Tsuzuki, D. & Dan, I., 2007. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. NeuroImage 34, pp. 1600–1611.

Kahl, F., Hartley, H. & Åström, K., 2001. Critical configurations for N-views projective reconstruction. IEEE Computer Vision and Pattern Recognition, Vol. 2, pp. 158-163.

Kanade, T., Rander, P. & Narayanan, P.J., 1997. Virtualized reality: Constructing virtual worlds from real scenes. IEEE Multimedia, Immersive Telepresence, 4(1), pp. 34–47.

Kanade, T. & Saito, H., 1998. The 3D room: Digitizing time-varying 3d events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, Robotics Institute, Carnegie-Mellon University.

Kavanagh, R.H., Darcey, T.M., Lehmann, D. & Fender, D.H., 1978. Evaluation of methods for three-dimensional localization of electrical sources in the human brain. IEEE-Biomedical Engineering 25.

Keith Jack, 2007. Video Demystified, 5th Edition, Publisher: Newnes.

Kim, K., Chalidabhongse, T.H., Harwood, D. & Davis, L., 2004. Background modeling and subtraction by codebook construction. IEEE International Conference on Image Processing.

Koenig, K., 2001. Multiphoton microscopy in life sciences. J. Microsc. 200, pp. 83-104.

Koizumi, H., Yamamoto, T., Maki, A., Yamashita, Y., Sato, H., Kawaguchi, H. & Ichikawa, N., 2003. Optical topography: practical problems and new applications. Applied Optics, Vol. 42, pp. 3054-3062.

Kosar, V., Gomzi, Z. & Antunovi´c, S., 2005. Cure of polyester resin in a cylindrical mould heated by air. Thermochimica Acta., pp. 134-141.

Kozubek, M. & Matula, P., 2001. An efficient algorithm for measurement and correction of chromatic aberrations in fluorescence microscopy. J. Microsc. 200, pp. 206–221.

Kusaka, T., Kawada, K., Okubo, K., Nagano, K., Namba, M., Okada, H., Imai, T., Isobe, K. & Itoh, S., 2004. Noninvasive optical imaging in the visual cortex in young infants. Human Brain Mapping 22, pp. 122-132.

Kwon, Y.H., 1989. The effects of different control point conditions on the DLT calibration accuracy. Unpublished class project report, Pennsylvania State University.

Läbe, T. & Förstner, W., 2004. Geometric Stability of Low-Cost Digital Consumer Cameras. In Proceedings of the ISPRS Congress, pp. 528-535.

Lagerlund, T.D., Sharbrough, F.W., Jack, C.R., Bradley, J.E., Strelow, D.C., Cicora, K.M. & Busacker, N.E., 1993. Determination of 10-20 system electrode locations using magnetic resonance imaging scanning with markers. Electro-encephalography and Clinical Neurophysiology 86, pp. 7-14.

Lamm, C., Windischberger, C., Ledolter, U., Moser, E. & Bauer, H., 2001. Co-registration of EEG and MRI data using matching of spline interpolated and MRI-segmented reconstruction's of the scalp surface, Brain Topography 14, pp. 93-100.

Latombe, J.C., 1991. Robot Motion Planning. Kluwer Academic Publishers.

Lee, S.C. & Nevatia, R., 2003. Interactive 3D building modeling using a hierarchical representation. International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis, IEEE Computer Society, pp. 58–65.

Li, Z.J., Hiroshi, M., Takashi, A. & Takashi, O., 2004. Gender difference in brain perfusion 99mTc-ECD SPECT in aged healthy volunteers after correction for partial volume effects. Nuclear Medicine Communication. 25(10), pp. 999-1005.

Li, F., Brady, M. & Wiles, C., 1996. Fast computation of the fundamental matrix for an active stereo vision system. In Proceeding of ICCV’96 International Conference on Computer Vision, pp. 157-166.

Liebowitz, D., Criminisi, A. & Zisserman, A., 1999. Creating Architectural Models from Images. EUROGRAPHICS, 18(3).

Logilink: http://www.logilink.eu/PCI%20express.htm.

Luhmann, T., Robson, S., Kyle, S. & Harley, I., 2006. Close range photogrammetry principles, methods, and applications. Whittles Publishing, Dunbeath, Scotland, UK.

Luo, Y.C. & Duraiswami, R., 2008. Canny edge detection on NVIDIA CUDA. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.

Maintz, J.B.A. & Viergever, M.A., 1998. A survey of medical image registration. Medical Image Analysis, 2(1), pp. 1-36.

Maki, A., Yamashita, Y., Ito, Y., Watanable, E., Mayanagi, Y. & Koizumi, H., 1995. Spatial and temporal analysis of human motor activity using noninvasive NIR topography. Med. Phys. 22, pp. 1997-2005.

Malmivuo, J. & Plonsey, R., 1995. Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press, New York.

Marieb, E.N. & Hoehn, K., 2006. Human Anatomy & Physiology. Seventh edition. New York: Pearson International Edition.

Mason, S. & Gruen, A., 1995. Automatic sensor placement for accurate dimensional inspection. Comput. Vision Image Understanding 61 3, pp. 454–467.

Mason, S., 1995. Conceptual model of the convergent multistation network configuration task. Photogrammetric Record 15 86, pp. 227–299.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R. & Mazoyer, B., 2001a. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R. Soc. London, Ser. B Biol. Sci. 356, pp. 1293–1322.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Feidler, J., Smith, K., Boomsma, D., Hulshoff Pol, H., Cannon, T., Kawashima, R. & Mazoyer, B., 2001b. A four-dimensional probabilistic atlas of the human brain. J. Am. Med. Inform. Assoc. 8, pp. 401–430.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J. & Woods, R., 2000. A probabilistic approach for mapping the human brain. In: Toga, A.W., Mazzoiotta, J.C. (Eds.), Brain Mapping: The Systems. Academic Press, San Diego, pp. 141–156.

McHugh, S., 2005. http://www.cambridgeincolour.com/tutorials/lens-quality-mtf-resolution.htm.

McKenna, S., Jabri, S., Duric, Z., Rosenfeld, A. & Wechsler, H., 2000. Tracking groups of people. Computer Vision: Image Understanding, Vol. 80, No. 1, pp. 42–56.

McNamee, L.P., Petriu, E.M. & Spoelder, H.J.W., 2001. Photogrammetry calibration of a mobile robot model. IEEE Instrumentation and Measurement Technology Conference: Vol. 1, pp. 245-250.

MeshLab: http://meshlab.sourceforge.net/.

Methley, B.D.F., 1986. Computational Models in Surveying and Photogrammetry. Thomson Press Limited, 346 pages.

Mills, J.P., Schneider, D., Barber, D.M. & Bryan, P.G., 2003. Geometric assessment of the kodak DCS Pro back. Photogram. Rec. 18, pp. 193-208.

Mitchell, H.L. & Newton, I., 2002. Medical photogrammetric measurement: overview and prospects. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 56, Number 5, pp. 286-294(9).

MNI brain and the Talairach atlas: http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach.

Motion Analysis Cooperation: http://www.motionanalysis.com/.

Narayanan, P.J., Rander, P.W. & Kanade, T., 1998. Constructing virtual worlds using dense stereo. In IEEE 6th International Conference on Computer Vision, 4-7 Jan, Bombay, India, pp. 3–10.

NDI: http://www.ndigital.com/.

Neugebauer, P.J., 1997. Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images. International Journal of Shape Modeling, Vol. 3 (1&2), pp. 71-90.

Niemz, M.H., 1999. Laser-Tissue Interactions. Springer-Verlag.

Nister, D., 2004. Automatic passive recovery of 3D from images and video. IEEE Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 438-445.

Obrig, H. & Villringer, A., 2003. Beyond the visible-imaging the human brain with light. J Cereb. Blood Flow Metab., Vol. 23, pp. 1-18.

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., Kohyama, K. & Dan, I., 2004a. Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping. Neuroimage, vol. 21, pp. 99-111.

Okamoto, M., Dan, H., Shimizu, K., Takeo, K., Amita, T., Oda, I., Konishi, I., Sakamoto, K., Isobe, S., Suzuki, T., Kohyama, K. & Dan, I., 2004b. Multimodal assessment of cortical activation during apple peeling by NIRS and fMRI. Neuroimage, vol. 21, pp. 1275-1288.

Okamoto, M. & Dan, I., 2005. Automated cortical projection of head-surface locations for transcranial functional brain mapping. Neuroimage, vol. 26, pp. 18-28.

Olague, G. & Mohr, R., 1998. Optimal camera placement to obtain accurate 3d point positions. In Fourteenth International Conference on Pattern Recognition, Vol. 1, pp. 8–10.

OpenCV – Open Computer Vision library: http://www.intel.com/research/mrl/research/opencv/.

Page, A., Moreno, R., Candelas, P. & Belmar, F., 2008. The accuracy of webcams in 2D motion analysis: sources of error and their control. European Journal of Physics: Vol. 29, No. 4.

Parikesit, G.O.F., Darmawan, M. & Faisal, A., 2010. Quantitative low-cost webcam-based microscopy. Optical Engineering, Vol. 49, Issue 11, pp. 113205.

Pavlidis, T., 1982. Algorithms for graphics and image processing. Published by Springer-Verlag, Berlin, 416 pages.

Peipe, J. & Stephani, M., 2003. Performance evaluation of a 5 megapixel digital metric camera for use in architectural photogrammetry. In Proceedings of the XX International Congress for Photogrammetry and Remote Sensing, Ancona, Italy, pp. 259-262.

Perrin, F., Pernier, J., Bertrand, O., Giard, M.H. & Echallier, J.F., 1987. Mapping of scalp potentials by surface spline interpolation. Electroencephalography and Clinical Neurophysiology 66, pp. 75-81.

Pervölz, K., Nüchter, A., Surmann, H. & Hertzberg, J., 2004. Automatic reconstruction of colored 3D models. Proc Robotik, Vol. 1841, pp. 215-222.

Pollak, C. & Hutter, H., 1998. A webcam as recording device for light microscopes. J. Comput. Assist. Microsc. 10, pp. 179–183.

Pollefeys, M., Koch, R., Vergauwen, M. & Van Gool, L., 1999. An Automatic Method for Acquiring 3D Models from Photographs: applications to an Archaeological Site. International Archive of Photogrammetry and Remote Sensing, Vol. XXXII, Part 5W11, pp.76-80.

Rander, P., Narayanan, P.J. & Kanade, T., 1997. Virtualized reality: Constructing time-varying virtual worlds from real events. In Proceedings of IEEE Visualization 1997, pp. 277–283.

Remondino, F. & Börlin, N., 2004. Photogrammetric Calibration of Image Sequences Acquired with Rotating Camera. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W16.

Remondino, F. & Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. Int. Soc. Photogramm. Remote Sens., XXXVI, pp. 266-272.

Rieke-Zapp, D.H. & Peipe, J., 2006. Performance evaluation of a 33 megapixel alpa 12 medium format camera for digital close range photogrammetry. In Proceedings of the ISPRS Commission V Symposium of Image Engineering and Vision Metrology, Dresden, Germany.

Robson, S., Clarke, T.A. & Chen, J., 1993. The suitability of the Pulnix TM6CN CCD camera for photogrammetric measurment. SPIE Vol. 2067, Videometrics II, pp. 66-77.

Robson, S. & Shortis, M.R., 1998. Practical influences of geometric and radiometric image quality provided by different digital camera systems. Photogrammetric Record, 16(92), pp. 225-248.

Rosales, R. & Sclaroff, S., 2000. Inferring Body Pose without Tracking Body Parts, IEEE Conf on Computer Vision and Pattern Recognition, pp.II, pp. 721-727.

Roth, D.G., 2004. Automatic correspondences for photogrammetric model building. Proceedings of the XXth ISPRS Congress, Istanbul, Turkey.

Rusinkiewicz, S. & Levoy, M., 2001. Efficient variants of the ICP algorithm. IEEE International Conference on 3D Digital Imaging and Modeling, pp. 145-152.

Sabel, J.C., van Veenendaal, H.L.J. & Furnee, 1993. PRIMAS, a real time 3D motion analysis system. Optical 3-D Measurement Techniques II. Published by Wichmann, Karlsruhe, pp. 530-537.

Saito, H. & Kanade, T., 1999. Shape reconstruction in projective grid space from large number of images. In 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp.49–54.

Sanz-Ablanedo, E., Rodríguez-Pérez, J.R., Armesto, J. & Taboada, M.F.A., 2010. Geometric Stability and Lens Decentering in Compact Digital Cameras. Sensors 10, pp. 1553-1572.

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

Schoberl, J., 1997. NETGEN—An advancing front 2D/3D-mesh generator based on abstract rules. Comput. Vis. Sci. 1, pp. 41–52.

Shier, D., Butler, J. & Lewis, R., 2001. Hole's human anatomy and physiology, 9th. McGraw-Hill.

Shortis, M.R., Ogleby, C.L., Robson, S., Karalis, E.M. & Beyer, H.A., 2001. Calibration modelling and stability testing for the Kodak DC200 series digital still camera. In Proceedings of SPIE Videometrics and Optical Methods for 3D Shape Measurement, pp. 148-153.

Shortis, M.R., Robson, S. & Beyer, H.A., 1998. Principal point behaviour and calibration parameter models for Kodak DCS cameras. Photogram. Rec., 16, pp. 165-186.

Shortis, M.R., Seager, J.W., Robson, S., Harvey, E.S., 2003. Automatic recognition of coded targets based on a Hough transform and segment matching. SPIE Proceedings Series Vol. 5013, pp. 202-208.

Shortis, M.R., Snow, W.L., Childers, B.A. & Goad, W.K., 1993. The influence of storage media on the accuracy and repeatability of photogrammetric measurements using CCD cameras. Proceedings, SPIE Vol. 2067 Videometrics II, pp. 80-92.

Shortis, M.R., Clarke, T.A. & Short, T., 1994. A comparison of some techniques for the subpixel location of discrete target images, Videometrics III. SPIE Vol. 2350, pp. 239-250.

Sidney, F.R., 2002. Applied photographic optics, 3rd ed., Focal Press.

Simpson, G.V., Pflieger, M.E., Foxe, J.J., Ahlfors, S.P., Vaughan, H.G.J., Hrabe, J., Ilmoniemi, R.J. & Lantos, G., 1995. Dynamic neuroimaging of brain function. Journal of Clinical Neurophysiology 12, pp. 432-449.

Singh, A.K., Okamoto, M., Dan H., Jurcak, V. & Dan I., 2005. Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. Neuroimage, vol. 27, pp. 842-851.

Singh, K.D., Holliday, I.E., Furlong, P.L. & Harding, G.F., 1997. Evaluation of MRIMEG/EEG co-registration strategies using Monte Carlo simulation. Electro-encephalography and clinical Neurophysiology 102, pp. 81-85.

Snavely, N., Seitz, S.M. & Szeliski, R., 2006. Photo tourism: Exploring photo collections in 3D. ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), pp. 835-846.

Snow, W.L., Childers, B.A. & Shortis, M.R., 1993. The calibration of video cameras for quantitative measurements. Presented paper, 39th International Instrumentation Symposium, Albuquerque, New Mexico, 28 pages.

Sony datasheet of CCD image sensor ICX098BQ, 2003: http://www.sony.net/Products/SC-HP/datasheet/90203/data/a6809927.pdf.

Souza, M.A.de, 2009. Acquiring accurate head surfaces of newborn infants for optical tomography using digital photogrammetry. PhD Thesis.

Sprawls, P., 2000. The Physical Principles of Medical Imaging, 2nd Edition. Medical Physics Publishing, Madison, WI.

Stauffer, C. & Grimson, W.E.L., 1999. Adaptive background mixture modelsfor real-time tracking. Proceeding of Computer Vision and Pattern Recognition, pp. 246-252.

Taga, G., Asakawa, K., Maki, A., Konishi, Y. & Koizumi, H., 2003. Brain imaging in awake infants by near-infrared optical topography. Proc. Nat. Acad. Sci. 100(19), pp. 10722-10727.

Talairach, J. & Szikla, G., 1967. Atlas d'anatomie stereotaxique du telencephale: etudes anatomo-radiologiques. Masson and Cie, Paris.

Talairach, J. & Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Tarabanis, K.A., Allen, P.K. & Tsai, R.Y., 1995. A survey of sensor planning in computer vision. IEEE Transactions on Robotics and Automation, 11(1), pp. 86 – 104.

Tecklenburg, W., Luhmann, T. & Hastedt, H., 2001. Camera modelling with image-variant parameters and finite elements. Optical 3-D Measurement Techniques V, Gruen, A., Kahmen, H. (Eds.), pp. 328-335.

Toga, A.W. & Thompson, P.M., 2001. Maps of the brain. Anat. Rec. 265, pp. 37–53.

Towle, V.L., Balanos, J., Suarez, D., Tan, K., Grzeszczuk, R., Levin, D.N., Cakmur, R., Frank, S.A. & Spire, J.P., 1993. The spatial location of EEG electrodes: locating the best fitting sphere relative to cortical anatomy. Electroencephalography and Clinical Neurophysiology 86, pp. 1-6.

Triggs, B. & Laugier, C., 1995. Automatic camera placement for robot vision tasks. In Proceedings of 1995 IEEE International Conference on Robotics and Automation, pp. 1732–1737.

Triggs, B., McLauchlan, P., Hartley, R. & Fitzgibbon, A., 1999. Bundle Adjustment - A Modern Synthesis. ICCV '99: Proceedings of the International Workshop on Vision Algorithms. Springer-Verlag, pp. 298–372.

Trinder, J.C., 1989. Precision of digital target location. Photogrammetric Engineering & Remote Sensing, Vol. 55, No 6, pp. 883-886.

Tsuzuki, D., Jurcak, V., Singh, A.K., Okamoto, M., Watanabe, E. & Dan, I., 2007. Virtual spatial registration of stand-alone functional NIRS data to MNI space. Neuroimage, Vol. 34, pp. 1506-1518.

Turk, G. & Levoy, M., 1994. Zippered polygon meshes from range images. Proceedings of the Annual Conference Series on Computer Graphics, pp. 311-318.

Vaithianathan, T., Tullis, I.D.C., Everdell, N., Leung, T., Gibson, A.P. & Delpy, D.T., 2004. The Design of a portable infrared mapping system for functional imaging on babies. Review of Scientific Instruments, 75(10), pp. 3276-3283.

Van den Heuvel, F.A., 1999. A line-photogrammetric mathematical model for the reconstruction of polyhedral objects. In El-Hakim, Gruen (eds.), SPIE Proceedings, Videometrics VI, Vol. 3641, pp. 60-71.

Vicon Motion Systems: http://www.vicon.com/.

VideoInput library: http://muonics.net/school/spring05/videoInput/.

VMS, 1999: Vision Metrology System. Developed by Robson, S. & Shortis, M.. http://www.geomsoft.com/VMS/index.shtml.

Wackrow, R., Chandler, J.H. & Bryan, P., 2007. Geometric consistency and stability of consumer-grade digital cameras for accurate spatial measurement. Photogram. Rec. 22, pp. 121-134.

Wang, L.W., Zhang, L., Yu, Z.J., Chen, F., Si, X.Y. & He, D.R., 2008. Precision circular target location in vision coordinate measurement system. Advanced Materials and Devices for Sensing and Imaging III. Proceedings of the SPIE, Volume 6829, pp. 68290J.

Wang, B., Toro, C., Zeffiro, T.A. & Hallet, M., 1994. Head surface digitization and registration: a method for mapping positions on the head onto magnetic resonance images. Brain Topography 6**,** pp. 185-192.

Weik, S., 1997. Registration of 3-D partial surface models using luminance and depth information. IEEE International Conference on Recent Advances in 3D Digital Imaging and Modeling, pp. 93-100.

Welch, G., 1996. SCAAT: Incremental Tracking with Incomplete Information. Ph.D. thesis, TR96-051, University of North Carolina – Chapel Hill.

Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K. & Colucci, D., 1999. The hiball tracker: high-performance wide-area tracking for virtual and augmented environments. In M. Slater, editor, VRST'99. Proceedings of the ACM Symposium on Virtual Reality Software and Technology, 20-22 Dec., London, UK, pp.1–188.

Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K. & Colucci, D., 2001. High performance wide-area optical tracking - the hiball tracking system. Presence, 10(1), pp. 1– 21.

West, G.A.W. & Clarke, T.A., 1990. A survey and examination of subpixel measurement techniques, Close range photogrammetry meets machine vision. SPIE Vol. 1395, pp. 456-463.

Wong, K.W. & Ho, W., 1986. Close range mapping with a solid state camera. Photogrammetric engineering and remote sensing, Vol. 52, No 1, pp. 67-74.

Wong, S., Robson, S., Gibson, A. & Hebden, J., 2009. Low cost real-time web-cam photogrammetry to determine the locations of optical topography sensors located on the human head. Proceedings of RSPSoc 2009 Annual Conference, 8-11th September, Leicester, UK. pp. 463-470.

Wong, S., Robson, S., Gibson, A. & Hebden, J., 2010. Real-time close range webcam photogrammetry suited to the coordination of optical topography sensors located on the human head. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, pp. 608-612.

Woodlief, T. & Whiley, J., 1973. SPSE handbook of photographic science and engineering. pp.204.

Wren, C.R., Azarbayejani, A., Darrell, T. & Pentland, A.P., 1997. Pfinder: real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp. 780-785.

Wren, C.R., Clarkson, B.P. & Pentland, A.P., 2000. Understanding purposeful human motion. In Fourth International Conference on Automatic Face and Gesture Recognition, 28-30 March, Grenoble, France, pp.378-383.

Wu, J.J., Sharma, R. & Huang, T.S., 1998. Analysis of uncertainty bounds due to quantization for three-dimensional position estimation using multiple cameras. Optical Engineering, 37(1), pp. 280-292.

Yamashita, Y., Maki, A. & Koizumi, H., 1996. Near-infra-red topographic measurement system: Imaging of absorbers localized in a scattering medium. Rev. Sci. Instrum. 67 (3), pp. 730-732.

Yi, S., Haralick, R.M. & Shapiro, L.G., 1995. Optimal sensor and light source positioning for machine vision. Computer Vision and Image Understanding, 61(1), pp. 122-137.

Yu, W., 2003. An embedded camera lens distortion correction method for mobile computing applications. IEEE Trans. Consumer Electron. 49, pp. 894-901.

Zeiss, C. 2010. Depth of Field and Bokeh: http://www.zeiss.com/C12567A8003B8B6F/EmbedTitelIntern/CLN_35_Bokeh_EN/$File/CLN35_Bokeh_en.pdf

Zhang, Z., 1994. Iterative point matching for registration of free-form curves and surfaces. In: International Journal of Computer Vision, Vol. 13, No.2, pp. 119-152.

Zhao, Q., Ji, L.J., Shi, F. & Jiang, T.Z., 2006. Distinguish activations on sensorimotor cortex using high-resolution diffuse optical tomography. 5th International Conference on Photonics and Imaging in Biology and Medicine (PIBM'06), September 1-3, Wuhan, China.

Appendix 1 - DirectShow

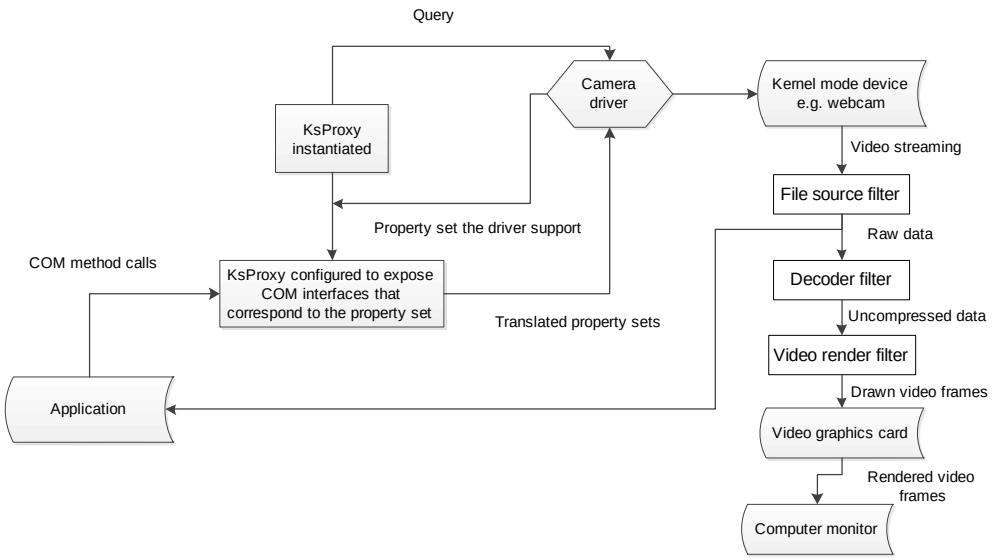
DirectShow [DirectShow] is a professional tool kit optimised for video collection, processing and display. It is used in this research to capture live video frames and implement webcam control functions, where the utilised features are described below.

DirectShow is composed of COM (Component Object Model) objects. A chain of selected objects is usually required to build a filter graph in order to accomplish a task such as live video capture. Each individual processing is performed by a COM object called filter, and data is transferred between filters by a COM object called pin. Common types of filters for video streaming include the source filter (loading video data), the transform filter (encoding or decoding the stream data) and the render filter (drawing video data on monitor).

Filters within a filter graph are coordinated by a COM object called filter graph manager, which not only decides when a specific data should be passed from one filter to the next filter but also communicates with applications to build a specific filter graph for each application. To transfer data from one filter to another, a common format of media type has to be agreed first. The media type defines the overall data types (e.g. video) in a structure called major type as well as more specific types (e.g. RGB24) in a structure called subtype. The detailed physical layout of the data (e.g. image size and frame rate) is defined in a structure called format block, where VIDEOINFOHEADER is an example of format block used for video data. Therefore it is possible for two media files with the same major type (e.g. MEDIATYPE_Video) and subtype (e.g. MEDIASUBTYPE_RGB24) but completely different image sizes and frame rates specified in the VIDEOINFOHEADER structure. Appendix 1 - Flowchart 1 summarises the procedures in order to set up a video streaming in DirectShow:

Appendix 1 - Flowchart 1. *Setup of video streaming in DirectShow.*

The communication between applications and hardware device (webcam) is summarised in Appendix 1 - Flowchart 2.



Appendix 1 - Flowchart 2. *Communication between application and webcam.*

KsProxy: a wrapper filter provided by DirectShow representing the capabilities of Windows Driver Model (WDM) streaming devices.

Appendix 2 - OpenCV

OpenCV [OpenCV] is a cross-platform middle-to-high level API (Application Programming Interface) library consisting of C/C++ functions which allow users to implement, test or modify various image processing algorithms. OpenCV uses Intel IPL (Image Processing Library) for its lower level image processing calls and also uses COM to create filter-like functions which combine with DirectShow to perform several synthetic functions.

Appendix 3 - Videoinput library

Although OpenCV contains a lot of image processing functions, its library is rather limited in that it supports only a few models of webcams/cameras. On the other hand, DirectShow supports a large number of cameras and meanwhile a large number of cameras support DirectShow. Videoinput [VideoInput library] is a library utilising DirectShow to import images into OpenCV for image processing. The following SDKs are used in this research to implement videoinput library:

1. Microsoft Platform SDK for Windows Server 2003 R2;
2. The DirectX SDK;
3. The DirectShow SDK.

Videoinput works by first identifying the number of connected and recognised cameras, followed by camera setups to start streaming. During the setup process, it identifies the model and connection type for each camera (Logitech C500 webcam and USB in this research), and also allows users to set hardware resolution and the colour model for each camera. Appendix 3 - Figure 1 is a screenshot of an eight-camera setup in this research:

Appendix 3 - Figure 1. *Screenshot of initial setup of 8 Logitech C500 webcams.*

Videoinput uses a method built on the COM concept to control video-acquiring devices. Therefore, the complete camera control and acquisition chain can be established on the DirectShow architecture (Appendix 1). The library is also very flexible in terms of modifications and method extensions suited to particular applications, in which image processing functions (Section 3.4.4.1 - 3.4.4.6) and 3D correspondence solutions (Section 3.5.2 - 3.5.3) can be appropriately embedded and implemented.