REAL-TIME CLOSE RANGE WEB-CAM PHOTOGRAMMETRY SUITED TO THE COORDINATION OF OPTICAL TOPOGRAPHY SENSORS LOCATED ON THE HUMAN HEAD

Sen Wong[1], Stuart Robson[1], Adam Gibson[2] and Jeremy Hebden[2]

[1]Department of Civil, Environmental and Geomatic Engineering, UCL, London WC1E 6BT
[2]Department of Medical Physics and Bioengineering, UCL, London WC1E 6BT
Email: sen.wong@cege.ucl.ac.uk

Commission V, WG V/5

KEY WORDS: optical topography, close range photogrammetry, bundle adjustment, target detection, accuracy assessment

ABSTRACT:

Optical topography is an emerging neuroimaging technique for the non-invasive monitoring of human cerebral haemodynamics. The technique is capable of localising brain activity if the positions and orientations of its network of optical sensing devices (optodes) are known relative to the anatomy of the human head. This paper describes the application of a photogrammetric network of low cost web-cams in coordinating and subsequently tracking the locations of targets situated on an optode equipped sensing pad to better than 1mm. Reported investigations include camera calibration and accuracy assessment with a network of Logitech C500 web-cam images at a resolution of 1280 by 1024; a comparison between images from a single colour channel selection and RGB multi-channel frames in terms of real time target detection and coordination. The paper also presents a working solution of 3D target coordination from the real time web-cam tracking system.

1 INTRODUCTION

Optical topography (OT) is a non-invasive tool for monitoring brain activity. OT utilizes the tight coupling between neural activity and regional cerebral blood flow to monitor relative regional changes of haemoglobin concentration [Hoshi, 2003; Koizumi et al., 2003; Obrig and Villringer, 2003; Strangman et al., 2002]. 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 [Okamoto et al., 2004].

Neuroimaging is most valuable when it can map different brain activation data into a common anatomical platform or coordinate frame. Such mapping allows group analysis over multiple subjects and comparison across different studies. In order to provide a common map both structural and functional brain imaging data are "normalized" to a standard template brain by linear and nonlinear transformation processes [Brett et al., 2002]. Presenting functional mapping data within a standard coordinate space has become common practice for tomographic functional brain mapping methods, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). Without mapping OT can measure brain activation, but it cannot identify the source of activation on the cortical structure since there is no structural information of the underlying brain. In order to register OT data with the standard coordinate space, it is necessary to map it from the head surface to the brain anatomy that corresponds to the signal source. Appropriate methods are in their infancy [Okamoto et al., 2004; Okamoto and Dan, 2005].

Electroencephalography (EEG) is the recording of electrical activity along the scalp and requires coordination of electrodes against a standard coordinate space in a similar manner to OT. A popular space is defined by the 10-20 coordinate system. The 10-20 system is defined by bony landmarks on the head which can be located by an expert clinician. Test-retest measurements following the repeated establishment of physical locations placed at bony landmarks by an expert clinician have been shown to exhibit errors of up to 7 mm on a single patient. Patient variability of up to 7.7 mm has been reported even when electrodes were applied by an experienced senior registered EEG technologist [Towle et al., 1993]. Regarding the accuracy of optode localization and the determination of key bony anatomical landmarks such as nasion and inion, a system capable of coordinating to 1mm accuracy should be sufficient for clinical applications.

2 AIMS AND OBJECTIVES

In order to spatially assess OT data, we require the correspondence between the scalp location where an OT measurement is performed and its underlying cortical surface. To do this, the first step is to determine source locations. This paper presents a study to spatially register the surface measurement (optode channel positions) into the brain anatomy (cortical surface), using a network of eight low-cost cameras in combination with photogrammetric bundle adjustment techniques. The main goal is to investigate the accuracy of spatial registration. Since the photogrammetric process is subject to several sources of uncertainty, studies were carried out to test and confirm if low cost off-the-shelf webcam technology is appropriate for clinical OT studies.

3 METHODOLOGY AND RESULTS

3.1 Robust Target Detection

In our 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 of 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 computing either a centroid or ellipse fit (Robson et al 1998). However such detection methods are inappropriate when isolating target perimeters from complex background images such as those that might be found in our study. A computationally simple approach is required in order to rapidly extract and process target image coordinates from sequences of images.

Tests [Wong et al., 2009] demonstrated that although 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, it is extremely sensitive to background noise. For example the row of four white dots in the bottom left picture of Figure 1 are a result of image artefacts. Previous work applied morphological operation prior to Canny edge detection in order to filter background features. The morphological operation, dilation followed by erosion [Gonzalez and Woods, 1992], implements image smoothing, noise removal and connecting gaps and disconnected edges. In Figure 1 the locations on the box seen in the left image have been automatically filtered out of the right hand image.

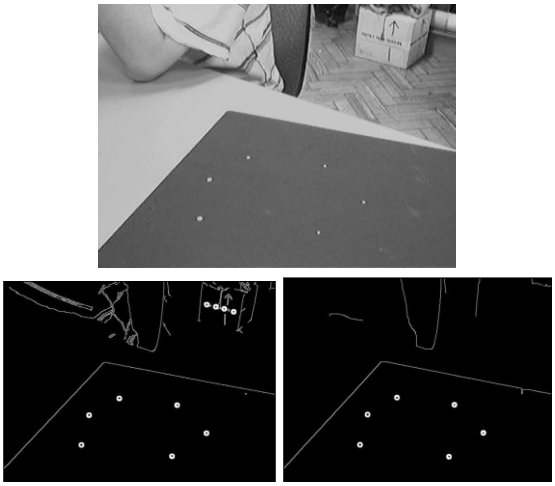

Figure 1. Comparison of target detection results.
Top: Original image;
Lower left: Canny detector only;
Lower right: Dilation, erosion and Canny detector.

Our earlier study has been extended to more complex environments, such as that shown in Figure 2. In this case although the erosion operation can eliminate some image noise, objects which are similarly sized to target images and form an obvious contrast against the background pass the criteria set by the Canny edge detector and are recognised as targets giving false positives.

Figure 2. Target detection in simulated clinic-environment.

In more extreme cases, such as Figure 3 (left) where a human head model with an OT optode pad have been imaged in a clinic-simulated laboratory environment, the contrast formed by the depth difference between the dark holes in the brighter optical table surface allow 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 passes the blob detection algorithm, even though it incorporates circularity criteria for target shape and size. Finally these holes are all recognised as targets and their individual centroids computed (middle picture of Figure 3). In the next stage the coordinates of both real and incorrectly recognised targets are 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].

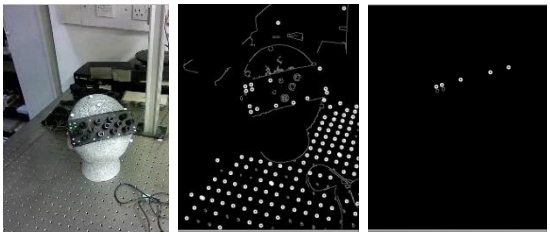

Figure 3. Comparison between multi-channel and single-channel target detections.

Incorrectly recognised targets are an obstacle to our goal of localising OT optodes. Incorrect measurements must be rejected before 3D correspondence. 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 search space is very important for real time processing.

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.

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 OT studies are usually carried out in locations such as at the patient bedside where lighting conditions are wide ranging.

The colour solution solves most of the image noise problem caused by these off-the-shelf cameras and we can concentrate on finding out what effects the target blurring of poor images would bring in terms of tracking accuracy. Utilising colour tuning targets, such as the seven green dots on the sensing pad in Figure 3, can be readily detected and measured with a single colour channel image. In Figure 3 (right) the single channel tracked 3 to 5 targets out of 7, whilst multiple channels tracked more than 50 targets. Figure presents the colour solution logic.

Although single channel target detection is able to achieve a much better result than multiple channel target detection, care must be taken to avoid unnecessary overhead which slow the real-time processing aspect of the algorithm.

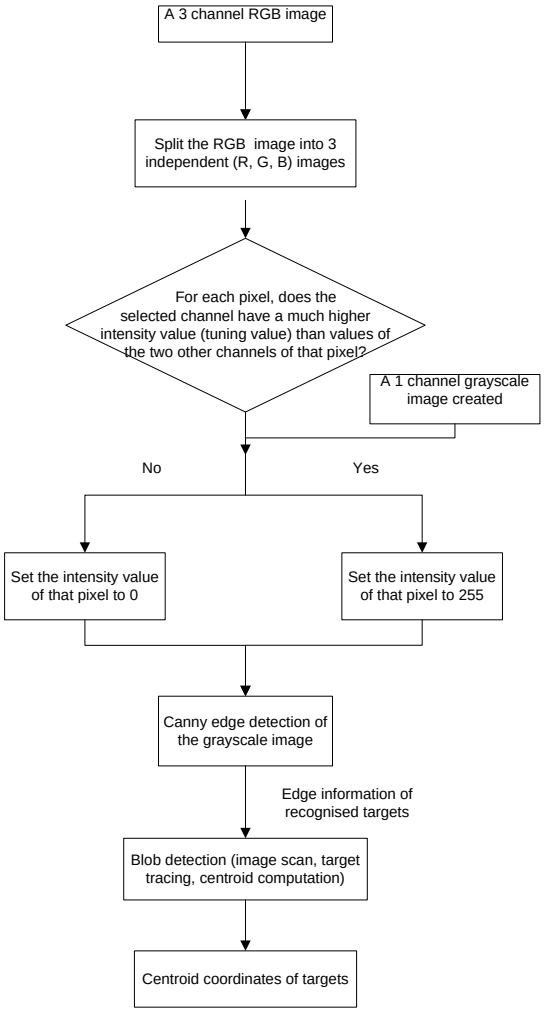

Figure 4. Colour Solution Logic.

Ideally in our application the correspondence computation only needs to be carried out once at the beginning until OT optodes move to obscure target images or new target(s) are detected. The aim is to deliver a consistent target detection result throughout the image sequence. We found that a tuning value (Figure 4) of 70, which is set manually based on the clinic simulated laboratory environment, has caused an edge disconnection in the target image perimeters (Figure 4 left). This is because even for the tuning target (green in our case), each imaged pixel has a different colour distribution. Setting a fixed tuning value can easily filter some imaged pixels inside the tuning target and therefore give rise to edge disconnection which will fail blob detection since a disconnected edge cannot form a single shape. In order to achieve the consistency of target detection across frames, a method is needed to compensate the drawback from image noise filtering.

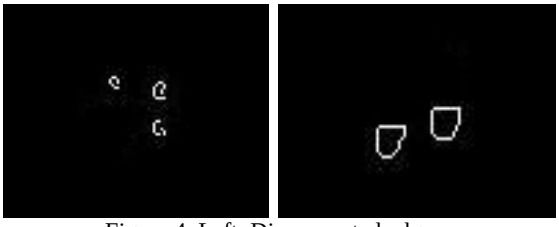

Figure 4. Left: Disconnected edges;
Right: Connected edges.

Tests in our simulated clinical environment have shown that by applying the dilation operation most disconnected edges can be reconnected and therefore fulfil the circularity requirement set by our blob detection algorithm. The advantage of applying the dilation between single channel filtering and Canny edge detection operation are its simple computation which does not significantly affect our real time processing. Tests with one webcam show that there is an average

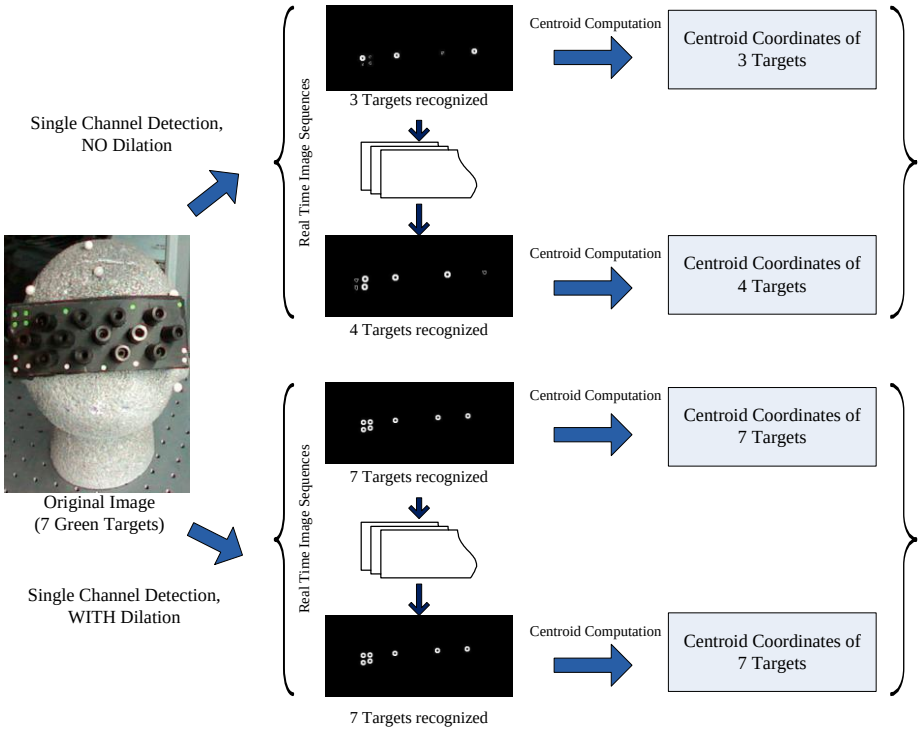

Figure 6. Comparison of two approaches of target detection in continuous image sequences.

time delay of 0.3ms per frame if dilation is applied. Note that the increased target detection consistency between image frames supports the use of simple inter frame tracking for correspondence rather than a full 3D correspondence solution at every frame. The lower part of Figure 6 shows the consistent detection and measurement of 7 target images.

A disadvantage of the dilation process is that an increase in the detected target perimeter since more pixels will be used in the centroiding process. Since the dilation applies the same operation on every edge pixel, all the target edge pixels are moved outwards, depending on which mask of dilation is applied. This equal increase of edge should not bias the intensity square weighted target centroid since these additional edge target pixels will be of low intensity (Equation 1). :

$$\begin{cases} x_{cent} = \dfrac{\sum_{i=1}^{n} \left[i^2 \bullet x_i \right]}{\sum_{i=1}^{n} i^2} \\ y_{cent} = \dfrac{\sum_{i=1}^{n} \left[i^2 \bullet y_i \right]}{\sum_{i=1}^{n} i^2} \end{cases} \quad (1)$$

where i = the image intensity value;
x_i, y_i = the image coordinates of each pixel;
n = the total number of pixels in one blob

Based on our clinic simulated laboratory environment, a system consisting of eight Logitech C500 webcams, each delivering 1280x1024 pixels connected with USB 2.0 ports to a 2GHz quad core computer is able to acquire and process at least one set of eight frames per second.

3.2 Accuracy Assessment of Eight Webcam System

A validation of the previous generation Logitech QuickCam Pro 4000 series webcams is reported in [Wong et al., 2009].. This paper presents a validation of next generation Logitech C500 webcams which are fitted with glass fixed focus lenses. The validation was carried out by comparing target intersection results from the webcam system with an accepted result of significantly higher precision. Compared to the previous study, the distance between cameras and targets has been increased (60-80cm) whilst the physical diameter of the targets on the calibration object is smaller. Both of these variations require higher image quality from the off-the-shelf webcams.

As a baseline the 3D target coordinates of some 20 white circular targets were estimated from an image network taken with a 6MP Nikon D100 fitted with a 28mm lens. This camera system has been in use for photogrammetric tasks for many years. Inter-target distance measurements made with Mitutoyo digital callipers (σ = 20μm) were used to ensure scale. Weights of 250μm were assigned to the distance measurements to denote the fact that targets were measured by overlaying the calliper with the target edges rather than through measurement to mechanical surfaces. Photogrammetric network processing was carried out with a self calibrating bundle adjustment in which all eight web cameras were simultaneously calibrated using eight sets of camera parameters, one for each physical camera (results presented in Table 1). Datum definition was by inner constraints using identical target coordinate starting values for all adjustments. [Granshaw, 1980].

Figure 7 Network configuration (left) and positions of intersected targets on the calibration object (right).

Program Control Variables	**Value**
Network datum definition type	generalised internal constraints
Maximum iterations for a solution	10
Default target image precision	1
Minimum images for a network target	4
Rejection criterion for image errors	5
Input Summary	**Value**
Number of camera calibration sets	8
Number of target image observations	1342
Total number of exposures	64
Number of exposures in the network	63
Total number of targets	61
Number of targets in the network	60
Results for the calibration solution	**Value**
Unit weight estimate (sigma zero)	1
RMS image residual (microns)	0.98
Number of observables in the network	2727
Number of unknowns in the network	605
Number of redundancies in the network	2122
Mean number of images per target	22.3
Target precision summary	**Value**
Mean precision of target coordinates (microns)	38.6
Relative precision for the network	10000

Table 1. 8 webcam self calibrating bundle adjustment.

A low cost system also requires minimal effort on the part of the user in setting up and operating the system. To routinely calibrate the system, a rigid targeted calibration object was rotated within the field of view of the eight cameras and a sequence of eight images was taken with each camera to provide a convergent network [Atkinson, 1996].

To investigate the accuracy of this system, summary statistics of the target coordinate differences between the Logitech C500 webcams at their maximum resolution of 1280 by 1024 pixels (4μm square pixels) were compared against reference target coordinate estimates from the Nikon D100 network. Data from this comparison are presented in Table 1 and Figure 8 where it can be seen that mean absolute coordinate discrepancies are within

0.2mm and webcam RMS image measurement residuals are of the order of ¼ of a pixel.

Figure 8 specifically demonstrates the discrepancy between coordinates calculated by two processes. An intersection solution taking all camera interior and exterior orientation parameters as known which would be the case for real time clinical measurement. A self calibrating bundle adjustment using identical input data to the intersection case. Results show that intersection only leads to greater coordinate discrepancies as might be expected since the solution does not allow for any physical instability. However comparison against the Nikon network reference shows that all coordinate axes are significantly better than the 1mm tolerance.

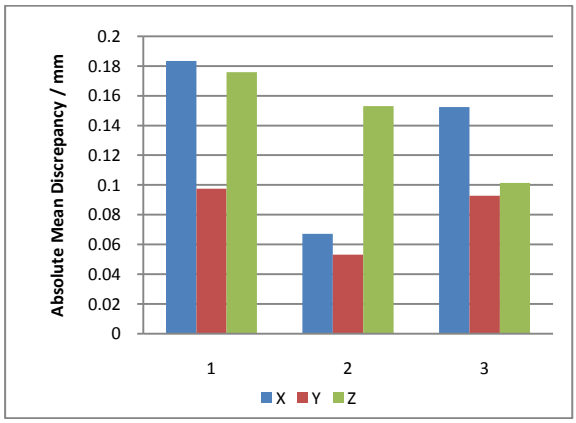

Figure 8. Comparison of absolute mean discrepancies between target coordinates.

1. 8-webcam system (using intersection) vs. Nikon reference;
2. 8-webcam system using self calibration vs. Nikon reference;
3. 8-webcam system using intersection vs. 8-webcam system using self calibration

4 DISCUSSION AND CONCLUSION

This assessment of an eight webcam system demonstrates that a close range network of calibrated webcams, with a stand-off distance of ~0.6m, is capable of 3D target coordination within 0.2mm of those computed from a network of images taken with a calibrated Nikon D100 camera. The RMS residual in the distance measurements made with the calipers was 180µm further confirming the capability of the system. These data are considered to be appropriate for OT localisation and tracking applications given the uncertainty of defining the bony landmark locations necessary to define the 10:20 coordinate system. The assessment has therefore demonstrated that fixed lens off-the-shelf webcams can be used for target positioning for optical topography studies, provided the camera units are in close proximity to the head. Furthermore, the work shows that low cost cameras can provide a suitable platform for real time photogrammetry.

REFERENCES

Atkinson K.B., 1996. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, pp. 18-19.

Brett M., Johnsrude I.S., Owen A.M., 2002. The problem of functional localization in the human brain. Nat. Rev., *Neurosci.* 3, 243–249.

Canny, J., 1986. A Computational Approach To Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679–714.

Darrell T, P. Maes, B. Blumberg, and A.P. Pentland, 1994. A novel environment for situated vision and behavior. *Workshop on Visual Behaviors*, 19 June 1994, Seattle, WA, USA, pages p.68–72.

Gonzalez R.C. and Woods R.E., 1992. *Digital Image Processing*. Addison Wesley Publishing Co.,.

Granshaw S.I., 1980. Bundle adjustment methods in engineering photogrammetry. *Photogrammetry Rec.* 10, pp. 181-208.

Hilton A., 1999. Towards model-based capture of a persons shape appearance and motion. *IEEE International Workshop on Modelling People*, 20 Sept. 1999, Kerkyra, Greece, pages p.37–44.

Hoshi Y., 2003. Functional near-infrared optical imaging: utility and limitations in human brain mapping. *Psychophysiology*, vol. 40, pp. 511-520.

Iwai Y, Ogaki K., and Yachida M., 1999. Posture estimation using structure and motion models. *Seventh IEEE International Conference on Computer Vision*, 20-27 Sept. 1999, Kerkyra, Greece, pages p.214–19 vol.1.

Koizumi H., Yamamoto T., Maki A., Yamashita Y., Sato H., Kawaguchi H., Ichikawa N., 2003. Optical topography: practical problems and new applications. *Appl. Opt.*, vol. 42, pp. 3054-3062.

Luhmann, T., Robson, S., Kyle, S., Harley, I., 2006. *Close Range Photogrammetry: Principles, Methods and Applications*. Whittles, Scotland, 510 pages.

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., 2004. 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., 2004. 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.

Robson, S. and 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.

Strangman G., Boas D.A., Sutton J.P., 2002. Non-invasive neuroimaging using near-infrared light. *Biol. Psychiatry* 52, 679–693.

Towle V.L., Balanos J., Suarez D., Tan K., Grzeszczuk R., Levin D.N., Cakmur R., Frank S.A., and Spire J.P., 1993. The spatial location of EEG electrodes: locating the best fitting sphere relative to cortical anatomy. *Electroencephalography and Clinical Neurophysiology* 86: 1-6.

Wong S., Robson S., Gibson A.P., Hebden J., 2009. Low cost real-time web-cam photogrammetry to determine the locations of optical topography sensors located on the human head. *Remote Sensing and Photogrammetry Society Annual Conference*, 8-11 Sept. 2009, Leicester, UK.