Photogrammetric Record, 14(81): 459–468 (April 1993)

HEIGHT ESTIMATION USING VIDEO SECURITY IMAGERY

By J. H. Chandler
City University, London

(Paper read at the Thompson Symposium held at the University of York on 29th March, 1992)

Abstract

Security video imagery has been used by the Engineering Photogrammetry Unit at City University to determine the height of a person robbing a betting office. The problem of selecting the correct functional model where a network is geometrically weak and consequently unreliable is both illustrated and stressed. Controlled tests suggest that appropriate software and detailed analysis can produce both precise and accurate estimates. Benefit to the legal profession of this particular measurement technique is not apparent, due to non-photogrammetric limitations.

Introduction

The aim of this paper is to indicate how the Engineering Photogrammetry Unit (EPU) at City University, London made use of security video imagery to determine the height of a person robbing a betting office. The work was carried out on behalf of the Crown Prosecution Service (CPS) for the purpose of providing legal evidence in the British Crown Court.

A major problem with this type of analysis is the design of the photogrammetry. Typically, geometric strength of the network is weak due to a limited number of cameras, frames and non-stereoscopic coverage. Such a weak network yields an (externally) unreliable estimation, compounds the problem of selecting the optimum functional model and inaccurate results can be obtained. This particular problem is illustrated in the analysis carried out by the EPU and shows that not only is suitable software required, but extreme care in measurement, processing and particularly analysis is essential. Despite these concerns, evaluation of the technique in a controlled environment proves that accurate results can be obtained, although selecting the correct functional model remains a critical problem.

The Legal Application

In November 1990, the EPU was approached by the CPS to carry out a photogrammetric analysis of a series of video images recorded during an armed robbery of a betting office in London. The aim was to estimate the height of the robber which could then be compared with the measured height of the suspect. The supplied videotape did not provide ideal material for the determination of height of an individual by photogrammetric techniques (Fig. 1) for the following reasons.

(1) Although two cameras were installed in the betting office, the recorded images were alternately sequenced so that no stereoscopic overlap was created at any instant of time. This created the problem that true three dimensional coordinates could not be obtained for any moving object such as a robber.

(2) Although a number of static and well defined points which imaged on the video could serve as photogrammetric control, their distribution was poor and their total number limited.

(3) Two CCTV video cameras were ued to record the images and such cameras are not designed for photogrammetric use.

(4) Only two weeks were available for the whole project during which all survey, measurement, processing, analysis and report writing had to take place.

(5) The human body is not a fixed and well defined linear object and so a photogrammetric estimation of a person's height is dependent upon other factors. Typically these are associated with posture, stance and thickness of hair. In this particular case, the problem was compounded by the robber wearing a woolly hat.

FIG. 1. Video image of the robber.

Despite these concerns it was felt that the EPU could produce a result, qualified by a stochastical measure of quality, which would reflect the uncertainties indicated above. It would then be up to the jury to decide upon the relevance of a photogrammetric estimation compared with the known measured height of the defendant. The decision was taken to proceed.

Control Survey

A survey was carried out at the premises of the betting office using a tape and level. From these measurements, the three dimensional co-ordinates of prominent features imaged on the video were determined (Fig. 2). These points would serve the function of photogrammetric control and were determined to a precision of ± 3 mm in plan and ± 2 mm in height. The survey also encompassed the positions of the perspective centres of the two video cameras, their precision being determined to $\pm 0{\cdot}05$ m.

Frame Grabbing

Critical frames were identified and attempts made to frame grab and store these data on disc for consequent measurement. Access to a personal computer with a frame grabbing board was obtained but the task proved difficult. One major requirement of a video surveillance system is to record many hours of business before the videotape needs substitution. This particular system achieved this by recording a frame once every three seconds and alternating between each of the two

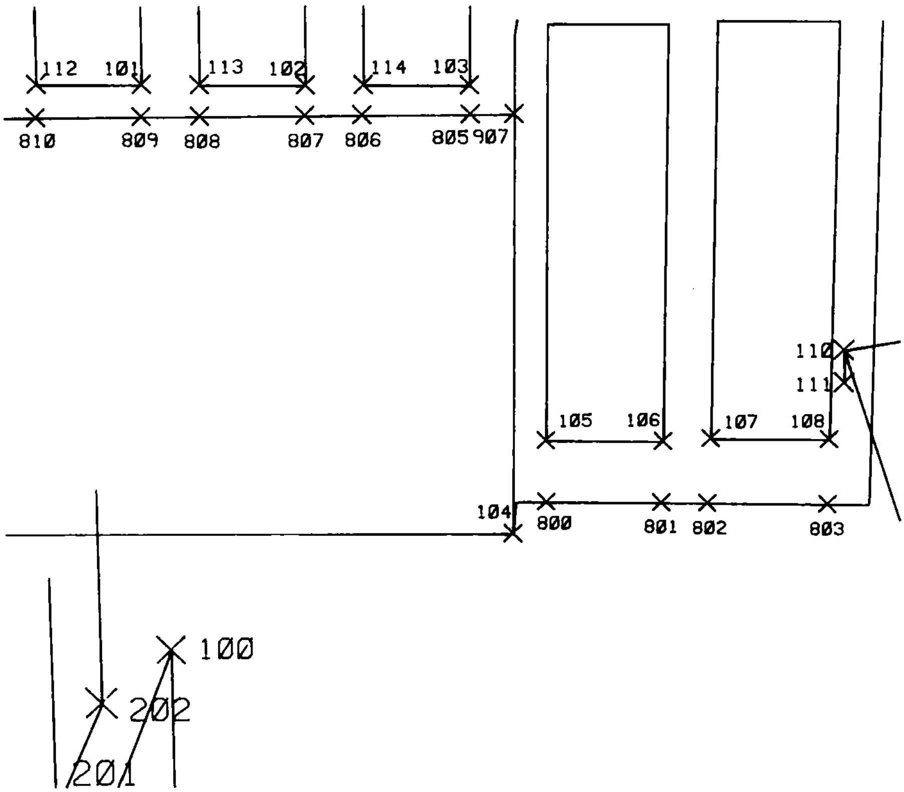

FIG. 2. Measured points in the betting office.

video cameras. The robber was only on the premises for a few minutes and so the number of frames where the robber was standing still, erect and with both feet and top of head imaged together was strictly limited. Moreover, it was found that if the video player was placed on "pause" for display and grabbing of any particular image, the frame often broke up and distorted badly. Frame grabbing was only possible when the video player was in "play" mode and it was then extremely difficult to grab any one specific image. After many hours spent attempting to grab specific frames, only four acceptable frames were obtained, three recorded by one of the cameras. The frame grabber had a resolution of 512 × 512 pixels, each with 256 grey levels.

Image Measurement

Using the personal computer and basic image processing software, the pixel co-ordinates of all visible control points were recorded, each measured three times. The pixel co-ordinates of a series of points representing the positions where the robber's feet were in contact with the floor were also measured (foot points). Similarly, co-ordinates representing points positioned around the robber's head were recorded (head points). These points were taken where the woolly hat appeared to be pressing hard against the robber's head and where a curve, taken to represent the head, was visible on the image. These image co-ordinates were measured with an estimated precision of $\pm 0{\cdot}5$ pixel.

Data Processing

The measured data were processed using analytical photogrammetry. Analogue methods would not have provided sufficient flexibility to obtain a solution.

A video camera is not necessarily designed to meet the condition of collinearity and the focal lengths of these particular cameras were unknown. A self calibrating bundle adjustment is often used to overcome this type of problem (Brown, 1976; Granshaw, 1980; Chandler and Cooper, 1988) and four separate estimations were used in which the focal length of the camera was determined. Although the program could determine parameters to model lens distortion, these additional parameters could not be estimated with statistical significance.

The lack of stereoscopic overlap due to the sequencing nature of the cameras prevented three dimensional co-ordinates from being derived by photogrammetric measurements alone. By making certain geometric assumptions, it was possible to obtain the three dimensional co-ordinates of the measured head and foot points. Two geometric assumption were made.

(1) The robber was standing with his feet flat on the floor. The Z ordinate of the foot could be assigned that value derived by levelling the floor, so allowing the X and Y ordinates of the feet to be determined.

(2) The plan position of the robber's head was midway between those of his feet. The orientation of the object co-ordinate system was selected so that the Y ordinate of the head points could be assigned the mean Y ordinate of the feet points. This allowed the X and most importantly the Z ordinates for the head to be computed.

These two geometric constraints could be enforced readily in the self calibrating bundle adjustment by "fixing" the appropriate ordinate. By processing these measured data for each frame, it was possible to derive three dimensional co-ordinates for points representing the robber's feet and head.

Determination of Height of Robber

The co-ordinates of both foot and head points were read into a MicroStation graphics design file for graphical display on the screen of an Intergraph CAD system. It was impossible to identify the top of the head directly because the robber was wearing a woolly hat and so the highest point on the robber's head was estimated by fitting a best fit arc to the measured head points. It was then possible to derive an estimate of the robber's height by measuring the vertical distance between the foot points and the highest point on the arc (Fig. 3). This procedure was repeated for all four frames and there was a variation of 0·084 m (3·3″) between these four estimates. This variation was thought to be attributable to differences in posture between frames, the poor distribution of control data and low resolution of the images.

Statistical analyses of these four estimations gave a mean of 1·888 m and a standard deviation of 0·039 m. Therefore it was concluded initially, with 95 per cent confidence, that the top of the robber's head was between 1·826 m and 1·950 m (between 5′ 11·9″ and 6′ 4·8″) above the carpeted floor surface. No account could be taken of the thickness of the heel of the shoe.

With the benefit of hindsight it is apparent that the *a posteriori* variance factors associated with the original bundle estimates were larger than would perhaps be expected. More exhaustive analysis was not carried out at this stage because the Metropolitan Police confirmed that the height of their suspect was within 0·01 m of the EPU mean estimate.

The Missing Systematic Error

The EPU was not the only expert witness making an attempt to derive an estimate of the robber's height. Dr. A. D. Linney, Department of Medical Physics and Bioengineering, University College London, who has expertise in facial mapping, had produced an estimate which differed from the initial EPU solution.

The Linney estimate was based upon analysis of one single frame (Frame 2728, Fig. 1). A known distance on the object (Fig. 2, Point 104 → 907) and the same distance measured on a photographic print were used to derive a single scale factor for the whole image (Linney and Coombes, 1990). This single scale factor was applied

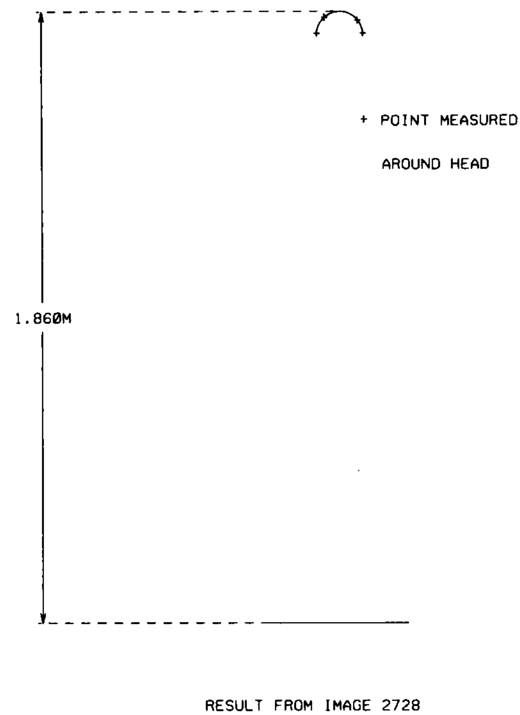

FIG. 3. Estimate of the height of the robber.

to the distance measured on the print between the feet and head of the robber and used to estimate the robber's height at 5′ 9·9″ (Linney and Coombes, 1990).

Photogrammetrists are aware that there are several systematic error sources ignored by such an approach. Most obvious is relief and tilt displacement, but also the inner geometry of the video camera is disregarded. Distortions are further introduced by taking a photograph of a video screen and producing the necessary paper print enlargement. Originally it was thought that these omissions in the simple Linney functional model were sufficient to explain the difference between the two estimates.

The effect of each type of normal photogrammetric source of systematic error was considered individually, but none yielded a sufficient explanation for the large difference in estimations. Background reading on the nature of video imaging systems was carried out and it was realised that one other source of systematic error had been ignored in the original EPU solution. In a video camera, variations in beam scanning velocity and geometry can lead to deformations in the image (McGlone, 1989). Tests carried out by El-Hakim *et al.* (1989) show that electronic distortions of video tube cameras are largely systematic and stable and that a large proportion of the electronic distortion is caused by scale differences between the horizontal and vertical axes of the video image. More significantly, large amounts of video distortion can be removed by a simple scale correction (El-Hakim *et al.*, 1989).

In the original EPU approach, image scale is defined by the relationship between control co-ordinates in the object space and their associated image co-ordinates. By making use of a standard self calibrating bundle adjustment, a single mean scale factor was derived for both x and y image ordinates, introducing an important scaling error. The Linney solution used one height difference to define scale in the y image direction. Scale in the x direction was never determined but was not required for the purpose of deriving the height of the robber. The error sources inherent with the simple Linney approach discussed above were indeed present, but their effects were insignificant in comparison to the scaling error introduced by EPU. Linney's solution was therefore originally more accurate than the EPU approach, despite the simplifying assumption underlying it. By modifying the functional model used in the bundle estimation to account for differential scale between x and y co-ordinates, a more accurate and precise solution could be obtained.

Revised Data Processing

Additional programming was required to allow the General Adjustment Program (GAP) to derive a differential scale factor between x and y image co-ordinates. This was incorporated into the software by estimating corrections to two focal lengths, one associated with the x image ordinates, the second with those in the y image direction. The differential scale factor was calculated simply from the ratio between these two estimated focal lengths.

When these data were processed using the new version of GAP, the scale factor was found to be 0·84 which represents a large difference of scale between the x and y image co-ordinates. With the effective removal of this particular source of systematic error, it was found possible to include data from all four images in one combined solution, with two sets of inner orientation parameters associated with the two original cameras. This increased the geometric strength of the estimation to the extent that a parameter used to model the lens distortion of one camera could be derived with statistical significance.

Revised Determination of Height of Robber

The revised three dimensional co-ordinates of both the "foot" and "head" points were again read into an Intergraph MicroStation graphics design file and the height of the robber estimated using the techniques discussed earlier. There was now a variation of 0·049 m (1·9″) between the four estimates, considerably lower than the range obtained before. Statistical analyses of the four estimations gave a mean of 1·781 m and a standard deviation of the mean of 0·011 m. Therefore it was concluded, with 95 per cent confidence, that the top of the robber's head was between 1·745 m and 1·817 m (between 5′ 8·7″ and 5′ 11·5″) above the carpeted floor surface.

Concerns

The difference between the initial and revised solutions was over 0·10 m (1·888 m to 1·781 m), which was of extreme importance when put into the context of a person's height and liberty! What concerned the EPU most was missing the critical systematic error during analysis of output from the original self calibrating bundle adjustments. The functional model selected originally was inadequate and examination of the output had not indicated this error.

Undoubtedly a major factor in this omission was the poor design of the photogrammetric network. If the EPU had been able to specify the number, position and type of the cameras, their recording rate and the number and type of control points, the geometric strength of the network would have been significantly stronger. Not only would this have produced a precise solution, but the resulting large and perhaps systematic pattern of residuals would have indicated the presence of undetected systematic errors. Once the correct functional model had been identified, accurate results would then be obtained. A network with high internal reliability (Hottier, 1976) can yield both a precise and, perhaps more importantly, an accurate estimation.

In these types of analyses the photogrammetry cannot be designed and is generally, if not always, poor. Although it is normally possible to obtain a solution, this will be only accurate if the selected functional model accurately simulates those original ray bundles. In these circumstances only experience, patient examination and thorough testing will help the photogrammetrist from making potentially disastrous mistakes. Unfortunately the pressure of commercial contracts leaves little time for such analysis!

A Controlled Simulation

The application of photogrammetric techniques to solve this general type of problem was of interest to the EPU and the concerns indicated justified further examination into the various factors affecting a solution.

The Rigid Robber

It was decided to replicate the geometric relationships between robber, security video camera and photocontrol in a controlled environment. The robber was replaced by a survey staff located in front of an array of control targets normally used for camera calibration. The relationship between the staff and the control field was determined using a combination of taped distances and height differences. Black circular targets were placed over the control points and on the staff at heights of 0·000 m, 1·600 m and 1·800 m. These targets were of sufficient size to be imaged on the video camera at a range of image scales. A single Panasonic NV-MC 10B CCD home video camera was used to obtain a series of images of the staff and control field from various positions (Fig. 4), these images being stored on TDK EHG EC30 video cassette.

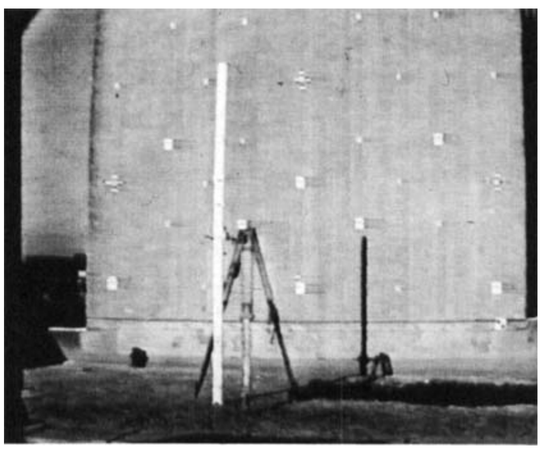

FIG. 4. Laboratory simulation of geometric relationships.

Image Measurement

Access to a personal computer based video frame grabber was provided and six images were grabbed at a resolution of 748 × 548 pixels in the red, green and blue bands. Only the red images were used for further analysis. Image positions of the control and staff targets were measured using similar procedures to those discussed earlier. The geometric configuration between the six measured frames and the targets was convergent (Fig. 5) and represented a network with stronger internal reliability than was possible in the betting office.

Data Processing

These measured data were processed using the self calibrating bundle adjustment with and without the differential scale factor carried as an additional parameter. Staff object points were treated in an identical manner as foot and head points previously, with the Z ordinate held fixed for the staff reading of zero and the

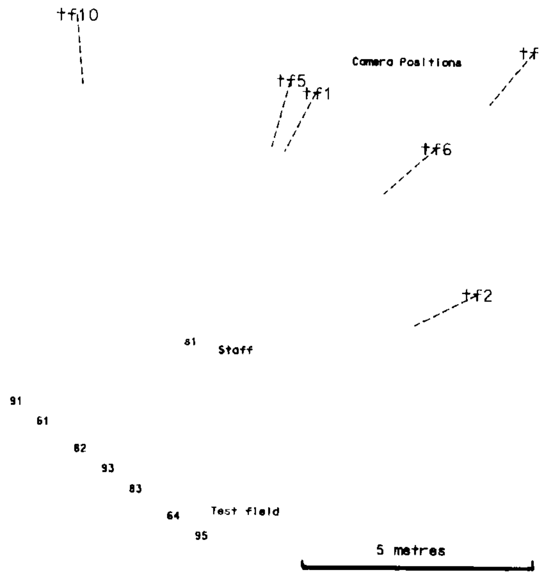

FIG. 5. Relative positions of staff, test field and video camera.

resulting XY co-ordinates held fixed for upper staff points. During the running of the self calibrating bundle adjustments a problem of slow convergence was encountered, although only when differential scale factor was included. This problem was found to be associated with correlation between the exterior and inner orientation parameters and was particularly acute where the control field was planar.

Four separate estimations were carried out in order to look at the role of the differential scale factor and the number of frames measured, either one or all six. An additional estimation was carried out using all measured frames, the optimum functional model and those taped distances and differences in height between the staff and the control points. This could be regarded as the best estimate for the staff position and staff height using all measured data.

Results

Table I indicates results obtained from these five different options.

TABLE I. Results obtained from simulation.

Number of frames	*Differential scale factor*	*Estimated staff height (m) (known height 1·600m)*	*Plan position of staff (X)*	*Plan position of staff (Y)*	*Variance factor* a posteriori	*Option*
1	none	1·598 ± 15 mm	102·445	112·065	2·56	A
6	none	1·568 ± 2 mm	102·499	112·183	2·00	B
1	0·978	1·600 ± 15 mm	102·495	112·169	0·74	C
6	0·978	1·598 ± 3 mm	102·504	112·183	0·90	D
6 + survey observations	0·978	1·591 ± 3 mm	102·506	112·195	0·87	E

The difference in scale between x and y image co-ordinates is clearly not as large as in the case cited earlier. This suggests that the value varies significantly between various frame grabbing systems and so a means of estimating differential scale factor is clearly important. The estimated value is close to unity and the difference in estimated staff height with and without this additional parameter is not large, particularly with single frame estimations (options A and C). Indeed, the estimated staff height is extraordinarily close to the known value of 1·600 m, although reference should be made to the low precision (±15 mm) of these particular estimates. The most accurate and reliable estimate is represented by Option D in which all six frames are used and differential scale is included. The estimated plan position of the staff also agrees closely with the known position represented by Option E and the *a posteriori* variance factor is also below unity, suggesting that a suitable functional model has been selected.

Conclusion

The original height determination of the robber and subsequent tests illustrate that photogrammetrists can determine geometric data from poor quality video images. The value of strong network geometry should not be underestimated because such strength leads to an estimate which is both internally and externally reliable. Reliability can help identify a suitable functional model and will ultimately lead to both a precise and, perhaps more importantly, an accurate result. When network geometry is weak, extreme care is necessary at all stages of processing and analysis.

The relevance of a photogrammetric estimate of a person's height can perhaps be questioned, particularly for legal purposes. A comparison between an estimated height from an incidental photograph and a physical measurement of a person standing erect, is not necessarily valid. Posture, stance and footwear all effect the height of an individual. The effects of such unusual sources of systematic error surely go beyond the expertise of the photogrammetrist and, once opposing Counsel have sensed this, the value of this particular expert evidence is put into doubt.

References

Brown, D. C., 1976. The bundle adjustment: progress and prospects. *International Archives of Photogrammetry*, 21(3). Paper 3 03: 33 pages.

Chandler, J. H. and Cooper, M. A. R., 1988. Monitoring the development of landslides using archival photography and analytical photogrammetry. *Land and Minerals Surveying*, 6(11): 576–584.

El-Hakim, S. F., Burner, A. W. and Real, R. R., 1989. Video technology and real-time photogrammetry. *Non-topographic photogrammetry*. (Second Edition, Ed. H. M. Karara.) American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia. 445 pages: 279–304.

Granshaw, S. I., 1980. Bundle adjustment methods in engineering photogrammetry. *Photogrammetric Record*, 10(56): 181–207.

Hottier, P., 1976. Accuracy of close-range analytical restitutions: practical experiments and prediction. *Photogrammetric Engineering and Remote Sensing*, 42(3): 345–375.

Linney, A. and Coombes, A. M., 1990. *Report on comparison of photographs and video images. re:* Regina v. McVicar. 4 pages.

McGlone, J. C., 1989. Analytic data-reduction schemes in non-topographic photogrammetry. *Non-topographic photogrammetry*. (Second Edition, Ed. H. M. Karara.) American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia. 445 pages: 37–57.

Résumé

On a pu utiliser l'imagerie d'une vidéo d'un système de sûreté, au laboratoire de photogrammétrie d'ingénierie de la City University, *pour déterminer la taille d'une personne dévalisant un bureau du pari mutuel.*

On met plus particulièrement l'accent sur le problème du choix d'un modèle fonctionnel correct, là où le réseau est géométriquement très lâche et par conséquent peu fiable. Des essais contrôles ont montré qu'en procédant à une analyse détaillée et en utilisant un logiciel approprié on pouvait obtenir des évaluations à la fois précises et exactes. Toutefois le bénéfice que peut tirer la profession juridique de cette technique de mesures particulière n'est pas évident, par suite des limitations qui ne sont pas d'ordre photogrammétrique.

Zusammenfassung

Aufzeichnungen eines Sicherheits-Video-Systems wurden durch die Abteilung für Ingenieurphotogrammetrie der City University *genutzt, um die Größe einer Person zu bestimmen, die ein Wettbüro ausgeraubt hat. Die Auswahl eines korrekten funktionalen Modells bei schwacher und folglich unzuverlässiger geometrischer Grundlage wird veranschaulicht und als Problem herausgearbeitet. Kontrollierte Versuche zeigen, daß eine geeignete Software und detaillierte Analyse sowohl genaue wie auch zuverlässige Schätzwerte liefern können. Ein Vorteil dieser besonderen Meßmethode für Juristen ist wegen der nicht-photogrammetrischen Begrenzungen nicht erkennbar.*