AN INEXPENSIVE AND SIMPLE 3-D MOTION CAPTURE PROCEDURE

Joanne Mikl

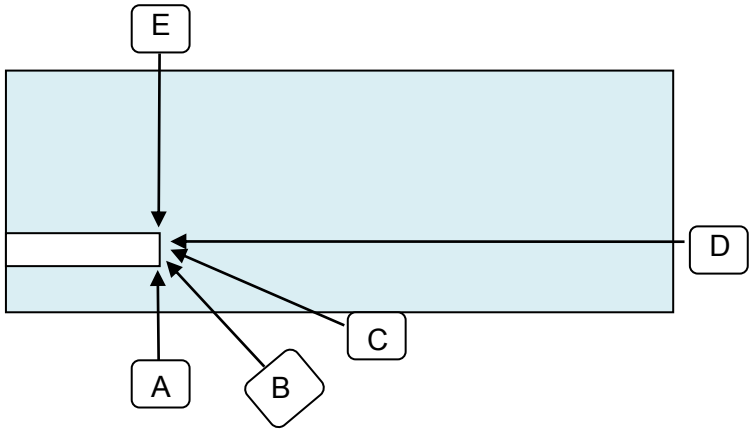
ACFR, University of Sydney, Sydney NSW, Australia

Commercial motion capture systems that are expensive, require permanent fixing, or require large and/or cumbersome calibration frames can preclude the use of video analysis in a training environment. Using inexpensive high-speed cameras, a checkerboard, and free open-source software a procedure was determined to achieve good motion-tracking accuracy for the study of divers at a pool. This procedure will be used in future studies, and may also be of use to other researchers with limited budgets.

INTRODUCTION: Commercial motion capture systems are typically expensive and generally require permanent fixing, or the use of a cumbersome calibration frame to cover the whole field of view of each camera. This precludes their use by researchers with limited budgets or working in environments where permanent set-ups are not achievable. There are, however, freely downloadable programmes and published methods which can be combined to overcome these problems. This paper presents a procedure which is simple, inexpensive and flexible to the training environment, along with the levels of accuracy anticipated using the procedure.

The Matlab Camera Calibration Toolbox (Bouguet, 2010) is a free software tool that may be used to determine the intrinsic and extrinsic parameters of a camera and hence calibrate the camera. It requires that a series of photographs of a checkerboard be taken by the camera, fixed in position and orientation and with the lens settings locked. The photographs may easily be obtained by filming a person who carries the checkerboard throughout the field of view of all cameras, and then places it in a fixed position that is visible to all cameras filming the performance. The checkerboard size should provide a compromise between the number of images required to cover the field of view and the ease of manoeuvring it around the area. This process of calibration is simple and relatively quick, and may be performed at any time as long as the cameras are not moved or the lens settings changed between calibration and the performances. Bouguet's toolbox may also be used to 'normalize' coordinates using the intrinsic camera parameters. Normalizing transforms digitised pixel coordinates to coordinates in the reference frame of an idealised camera with unit pixel focal length, no lens distortion, and origin on the optical axis of the camera. Along with the extrinsic camera parameters, the normalized coordinates may then be used to triangulate and hence determine the three-dimensional coordinates of a point with respect to one of the camera reference frames. For triangulation a minimum of two cameras are required. If their intrinsic and extrinsic parameters are known exactly, triangulation is a simple matter constructing viewing rays from each camera through the key point image location, and determining where they intersect (see Longuet-Higgins, 1981). If the camera parameters are not known exactly the viewing rays will not intersect and the location of the key point must be estimated. There are several approaches to this estimation. The oldest approach is a least squares estimate: the point chosen as the key point's 3-D location is the point that minimises the distance to each viewing ray in the world frame of reference (Szeliski 2011). A criticism of this approach (Hartley and Strum 1995) is that it does not consider the source of the error or the way that a camera maps from 3-D to 2-D. The dominant alternative is to minimise the L_2 reprojection error (see Hartley and Strum 1995). This involves seeking points in the image plane of each camera that produce intersecting viewing rays while being the minimum distance from the original digitised points. This approach assumes that the dominant source of error is in the image plane—that is, due to digitising or incorrect estimation of lens distortion—and that the extrinsic parameters are known exactly. Any of these three methods could be easily coded, for example in Matlab. It is unclear which method of triangulation would produce the best results when used with Bouguet's Toolbox in a typical sports scenario and so—in presenting a simple, inexpensive and flexible procedure for 3-D motion capture— these three methods of triangulation will be compared.

METHODS: Cameras and digitisation: Casio EX-FH100 are relatively low-cost cameras, and were available for use. The cameras were set to film at 120 fps and 640 x 480 pixel resolution and were used to film diving performances at the Sydney Olympic Park Aquatic Centre. Five possible camera positions surrounding the diving pool were identified. These positions are paired to reflect possible two-camera set-ups that could be used to film and then triangulate key points. Figure 1 illustrates the camera positions and pairings.



Camera pairs listed in order of increasing angle between the cameras

- A & B
- A & C
- A & D (near-orthogonal)
- C & E
- B & E

Figure 1: Possible camera positions

Lens zoom was used so that the camera's view covered a height that a diver may be expected to achieve and was centred horizontally on the diving board. To collect the required calibration images, a checkerboard (11 x 11 squares, each 100 mm side length) was filmed moving in view of each camera in turn and then placed statically in view of both cameras. Full coverage of the camera image spaces required the person to stand both on the diving board and the pool edge. Images from the footage were used with Bouguet's Toolbox to estimate the intrinsic and extrinsic parameters of the cameras by manually digitising the corners of the inner 10 x 10 square checkerboard, and following the prompts. Diving performance parameters are based on determining angles and lengths between anatomical points on a diver's body. The accuracy of triangulation should therefore be assessed by examining the accuracy of estimated angles and lengths. To do this, a cube framework of known size, with four corners marked with different colours, was filmed moving around the performance area. Footage was synchronised using a unique event visible to both cameras. The maximum error in synchronisation would be 1 frame (1/120sec). One thousand frames were then selected for digitisation, which was completed semi-automatically using the programme "Tracker" (Brown, 2012) to track the colours of the marked corners.

Triangulation: The three methods of triangulating using two cameras are designated here as "basic", "Lindstrom", and "midpoint". The basic method uses geometry as presented in Longuet-Higgins (1981), assuming that the viewing rays intersect. The "Lindstrom" method applies the 'niter2' code presented in Lindstrom (2010) to adjust the digitised points to minimise the L2 reprojection error, then triangulates via geometry as in the basic method. The "midpoint" method minimises the least-squares error with two cameras. Both the basic and Lindstrom methods discard some information since when solving for x, y, z coordinates only three equations are required, and four are available. The camera which will provide two equations will be referred to, from this point onwards, as the 'preferred' camera. The other camera will provide only one equation. The basic and Lindstrom methods were applied to each camera set-up with each camera 'preferred'. The same digitised points were used with each triangulation method. Once the 3-D coordinates were calculated, five lengths (two sides, two face diagonals, and one three-dimensional diagonal) and four angles (two of 90^{o} and two of 45^{o}) were found via coordinate geometry. These lengths and angles would cover various different positions in the performance area and combine the errors in positions of each corner in a variety of ways, some cancelling errors and others compounding them. The difference between calculated and actual values was used to determine accuracy. Methods and camera pairings were compared and the most accurate chosen.

RESULTS and DISCUSSION: Triangulation: Comparing the spread of errors in lengths or angles across camera positions, it was clear (Figure 2) that the basic method produced the least error. This was also true for individual camera pairings. The poor results achieved by the midpoint method support previous criticism of this method (Lindstrom, 2010). Since Lindstrom's method does not perform as well as the basic method this suggests that the extrinsic parameters are the dominant error. Bouguet's Toolbox is either better at estimating the intrinsic parameters than the extrinsic parameters, and/or the cameras used have little distortion.

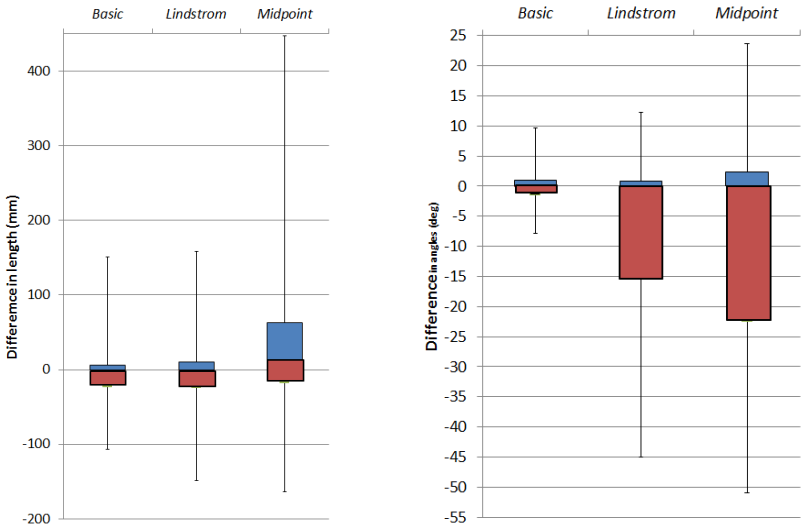


Figure 2: Box plots showing error across all camera pairs by method

Camera position when using the basic method: Orthogonal views are expected to produce the minimum error, due to the reduction in the size of the "region of uncertainty" (Hartley & Zisserman, 2003). However, the greater the angle between the cameras means that the checkerboard, when placed in view of both cameras, is at a larger angle to each of the cameras. This will make it more difficult for Bouguet's Toolbox to extract the checkerboard corners and estimate extrinsic parameters. The BE camera pair performed the worst, especially on length reconstructions, and pair EC performed worse that all pairings using camera A. This may possibly be attributed to the errors in extrinsic parameter estimation: with camera E, when manually digitising the corners the errors in corner extraction of the fixed checkerboard position were apparent to the eye, and this was confirmed by the larger estimated extrinsic pixel errors reported by the calibration toolbox. Camera pairs AB, AC, and AD had angle errors that were centred on zero with an interquartile range of ~2° and a range of ~10–17°. The same camera pairs tended to under-estimate length with the median error ranging from -9.1mm to -14.5mm across the pairs. Interquartile ranges were ~22–27 mm and ranges 114–146 mm. Of the three 'A' camera pairings, AD produced the least error for length in terms of mean, median, interquartile range, and for range and standard deviation when averaging the results from 'preferring' either camera. Pair AB tended to perform worst in terms of both the angles and lengths, although it was not the worst for all descriptive statistics. Which camera was 'preferred' did affect the basic triangulation results, which was mostly seen as a change in the range of error. Considering the effect of camera positions it is desirable to avoid an angle between cameras greater than 90°. It is also desirable to, as far as practical, have this angle close to orthogonal; AC and AD performed well. Considering the physical constraints and other users at the Sydney Olympic Park Aquatic Centre, set-up AC was achievable across a number of sessions and produced good results, and thus will continue to be used.

Accuracy: Regarding pair AC, with C 'preferred', (Table 1) very good accuracy is obtained for the middle 50% of estimates, when considering both values and percentages. Such error is insignificant in practical terms. For angles, the error range is reasonable in practical terms considering the angular measurements, and the fact that divers rotate at high angular velocities showing clear postural phases. The length error range is relatively larger, although

whether this is a concern depends on the “effect size” expected in any particular study. No conclusions can be drawn about whether the magnitude of error was dependant on the position in the field of view since the cube was moved generally and not in set intervals.

Table 1: Descriptive statistics of error for camera pair AC, with C ‘preferred’

	Length error, mm	Length error, %	Angle error, deg.	Angle error, %
Minimum	-75.6	-11.8	-5.2	-11.1
Lower Quartile	-24.4	-2.9	-1.0	-2.0
Median	**-13.4**	**-1.7**	**0.0**	**0.0**
Upper Quartile	-1.5	-0.2	1.0	1.8
Maximum	38.2	6.1	6.8	9.9
Interquartile Range	**22.9**	**2.7**	**2.0**	**3.8**
Range	113.8	17.8	12.0	21.0
Mean	**-13.2**	**-1.6**	**0.0**	**-0.1**
Standard Deviation	16.5	2.1	1.6	2.9

Recommended Procedure:

1. Set and fix camera positions. It is preferable that the camera views are at a relative angle of 90° or just under
2. Lock camera lens settings
3. Film the checkerboard being moved throughout field of view of both cameras, and stationary at a position visible to both cameras
4. Record athlete performances including a unique start event for synchronisation
5. Synchronise cameras using the start event
6. Use Bouguet’s Toolbox within Matlab to determine camera parameters
7. Normalize pixel coordinates using the Toolbox
8. Triangulate via the “basic” method to obtain 3-D coordinates.

CONCLUSIONS: This paper presented a procedure for obtaining 3-D coordinates of a point using inexpensive cameras, no cumbersome calibration frames or fixed set-ups, freely downloadable software, and an easily coded triangulation method. The “basic” triangulation method was selected since it produced the most accurate estimates of lengths and angles between two points. It was determined that cameras should be set at an angle of 90° or slightly less for best accuracy. This was the result of Bouguet’s Toolbox having more difficulty estimating extrinsic parameters for angles greater than 90° and the region of uncertainty increasing as the angle differs from 90°. Good accuracy was achieved in general for the configuration AC in Figure 1, and it was thus recommended.

REFERENCES:

Bouguet, J.Y. (2010). Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/ version dated 9 July 2010 used, accessed 5 February 2014.

Brown, D. (2012). Tracker, version 4.71, http://www.cabrillo.edu/~dbrown/tracker/ accessed 5 July 2012.

Lindstrom, P. (2010). Triangulation made easy. *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition*, 13–18 June 2010, pp. 1554–1561.

Longuet-Higgins, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.

Hartley, R.I. and Strum, P. (1995) Triangulation. http://users.cecs.anu.edu.au/~hartley/Papers/triangulation/triangulation.pdf

Hartley, R. & Zisserman, A. (2003) *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Szeliski, R. (2011) *Computer Vision: Algorithms and Applications*, Springer, 2011. Electronic draft http://szeliski.org/Book/ accessed 5 February 2014.