Optical Tracking Using Commodity Hardware

Simon Hay* Joseph Newman† Robert Harle‡

Digital Technology Group, Computer Laboratory, University of Cambridge

ABSTRACT

We describe a method for using Nintendo Wii controllers as a stereo vision system to perform 3D tracking or motion capture in real time. Commodity consumer hardware allows a wireless, portable tracker to be created that obtains accurate results for a fraction of the cost of conventional setups. Consequently, tracking becomes viable in situations where cost or space were previously prohibitive. Initial results show an accuracy of ±2mm over a large tracking volume.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented and virtual realities; I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture

1 MOTIVATION

Optical tracking and motion capture techniques are widely used due to their high spatial accuracy and update rates. A number of commercial systems exist [7], but the specialised hardware required makes the cost of even the cheapest of these prohibitive for many applications. Since the release of Nintendo's Wii games console a vibrant Internet community has sprung up using its unique controllers separately to take advantage of the capabilities they provide in a convenient and inexpensive form. In particular, Johnny Chung Lee has created some excellent demonstrations and brought the potential of the controllers to the attention of millions[1]. We believe that these controllers contain all the necessary sensors to build an optical tracking system for a fraction of the conventional cost.

2 BACKGROUND

Wii controllers contain an optical sensor, which is used in conjunction with the 'sensor bar' (a strip containing two clusters of infrared LEDs positioned on top of the television) to determine the position and orientation of the controller with sufficient accuracy to control a mouse pointer on screen. The optical sensor consists of a 1024x768 CCD with an infrared filter. It uses a custom system-on-a-chip to detect up to four infrared 'hotspots' – bright point sources of light – and transmit their positions and sizes back to the host via Bluetooth at a rate of 120Hz.

We mount two controllers rigidly with overlapping fields of view and use stereo vision techniques to recover the 3D locations of points seen by both. Several people have demonstrated 'head tracking' by mounting the sensor bar on a user's head and viewing it with a single controller; while this is sufficient to give the impression of parallax in a display, it does not allow the full six degrees of freedom pose to be determined and, in the absence of any calibration steps, can not give absolute results.

*e-mail: sjeh3@cam.ac.uk
†e-mail: jfn20@cam.ac.uk
‡e-mail: rkh23@cam.ac.uk

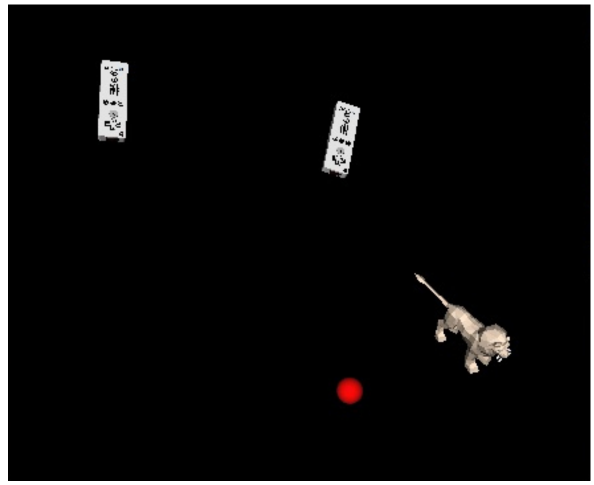

Figure 1: Demo application shows positions of controllers and targets

3 IMPLEMENTATION

We wrote a C application that runs on Linux and uses a slightly modified version of the libcwiid [2] library to communicate with the controllers via Bluetooth; this configures the controllers, collects and associates the data they return and passes it up the pipeline to subsequent stages for processing. It also provides a graphical user interface to allow the user to monitor the views seen by the cameras.

The markers used are 940nm wavelength infrared LEDs measuring just $1.5 \times 2.2 \times 2.8$mm (retroreflective passive markers are not detected by the controllers due to their size and the low intensity of the light reflected). The LEDs have a viewing angle of 160° and can be powered by a button cell and attached almost anywhere without difficulty or inconvenience. In particular, our outside-in design means that the amount of hardware carried by the person or object to be tracked is significantly less than that required for equivalent inside-out systems.

Given the quantised noisy (x, y) coordinates of the markers as seen by each controller, their 3D locations can be reconstructed by triangulation: minimising the square-sum of 2D image reprojection errors [3]. If three or more markers are attached to a rigid object in known positions the pose of that object can be recovered using point correspondences [1]; this stage in the pipeline is optional and may or may not be used depending on the intended application.

Figure 1 shows a screenshot of a demo application, recovering the full pose of an object with three markers (represented by the lion) and the position of a fourth single marker (represented by the sphere).

3.1 Calibration

To perform triangulation is is necessary to know the focal point of each camera; it is also useful to build a distortion model to allow correction of the images and improve accuracy. Furthermore, we

[1] `http://www.cs.cmu.edu/~johnny/projects/wii/`
[2] `http://abstrakraft.org/cwiid/wiki/libcwiid`

must know the positions and orientations of the cameras relative to each other. These parameters can all be determined by a calibration process. We use a square represented by infrared LEDs mounted at each corner as a planar pattern. Once each controller has observed the pattern at a number of different orientations the intrinsic and extrinsic parameters can be calculated with the Camera Calibration Toolbox for Matlab[3], using a method based on Zhang's [9].

This calibration step means that the controllers can be placed in arbitrary positions, and no manual measurement is required.

The cameras have a field of view of approximately 41° horizontally and 31° vertically, and can detect the markers at a range of over 5m in a normally-lit room, giving a tracking volume of the order of dozens of cubic metres depending on the camera positions.

The resultant camera matrix for one of our controllers in the format defined by Heikkilä [4] is shown in Figure 3.

4 Results

To assess the accuracy of the system, we mounted two infrared LEDs in pre-drilled holes on a length of stripboard 12 inches apart. The LEDs fit the holes tightly, and the stripboard was machined with very tight tolerances. We then moved the board throughout the tracking volume and recorded the reconstructed distance between the two markers for 7560 frames. The RMS distance error was 2.46mm, and the standard deviation was 2.23mm; these figures are comparable to those obtained for a much more expensive system using a similar testing method [5]. The histogram in Figure 2 shows the distribution of distance errors.

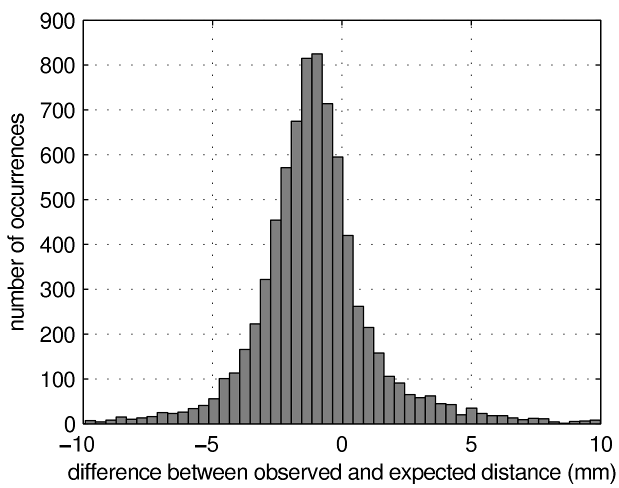

Figure 2: Accuracy of distance measurements

5 Limitations and Future Work

The biggest limitation with the system we have described is the built-in restriction of the controllers only to report four points. Although this is sufficient to recover the pose of a single object, it does not allow, for example, tracking all the fingers of both hands or full motion capture of the human body. This could be circumvented by using more markers but only turning on a small number at a time, in a manner similar to that used by the HiBall tracking system [8]; the tradeoff here is a reduced equivalent frame rate and additional hardware complexity. It may also be possible to work around this problem by using additional pairs of controllers, since the filters can easily be removed and replaced with ones that pass different wavelengths.

[3] http://www.vision.caltech.edu/bouguetj/calib_doc/

Figure 3: Camera matrix including focal length and principal point

$$\begin{pmatrix} 1.3063 & 0 & 0.5350 \\ 0 & 1.3023 & 0.4017 \\ 0 & 0 & 0.0010 \end{pmatrix}$$

No attempt is made to synchronise the shutters of the cameras beyond switching them both into streaming mode as close to simultaneously as possible. This could be improved by flashing an LED and identifying the frame from each camera where it is first seen, or by using an algorithm to estimate the phase difference between the two cameras by minimising the reconstruction error [6]. However, since the frame rate is comparatively high, any lack of synchronisation has only a very small impact on overall accuracy.

Of course, adding more controllers to the system would improve its accuracy and reliability; this is also an avenue for potential future work.

6 Conclusion

At the time of writing, Wii controllers sell for less than £30; the additional components to build the calibration pattern and markers cost no more than a few pounds. This makes the total cost of the system several orders of magnitude less than that of comparable commercial trackers and so opens up augmented reality techniques to a much wider audience: we have already received expressions of interest from artists, doctors and engineers as well as Computer Science researchers. It is ideally suited to applications such as virtual showcases where the pose of a single object must be recovered with high accuracy but at low cost [2]. We intend to make the source code of our system freely available so that anyone can build on it and take advantage of affordable tracking.

Acknowledgements

The authors wish to thank Andy Rice for his help and advice and Andy Hopper for his support of the work.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, 1987.

[2] O. Bimber, L. Encarnação, and D. Schmalstieg. The virtual showcase as a new platform for augmented reality digital storytelling. *Proceedings of the Workshop on Virtual Environments (EGVE 03), Zurich, Switzerland*, May 2003.

[3] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[4] J. Heikkilä and O. Silvén. A four-step camera calibration procedure with implicit image correction. *Proceedings of Computer Vision and Pattern Recognition (CVPR 97), San Juan, Puerto Rico*, pages 1106 – 1112, May 1997.

[5] T. Pintaric and H. Kaufmann. Affordable infrared-optical pose-tracking for virtual and augmented reality. *Proceedings of Trends and Issues in Tracking for Virtual Environments Workshop (IEEE VR 2007), Charlotte, NC, USA*, 2007.

[6] P. Pourcelot, F. Audigie, C. Degueurce, D. Geiger, and J. M. Denoix. A method to synchronise cameras using the direct linear transformation technique. *Journal of Biomechanics*, 33(12):1751–1754, Dec 2000.

[7] M. Ribo. State of the art report on optical tracking. *Technical Report VRVis 2001-25, TU Wien*, Nov 2001.

[8] G. Welch, G. Bishop, L. Vicci, S. Brumback, K. Keller, and D. Colucci. High-performance wide-area optical tracking - the hiball tracking system. *Presence: Teleoperators and Virtual Environments*, 10(1), 2001.

[9] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. *International Conference on Computer Vision (ICCV'99), Corfu, Greece*, pages 666–673, 1999.