

A Portable, Inexpensive Point-Tracking System for Validation of Wearable Biomechanics Sensors

G. P. Bailey and R. K. Harle

Computer Laboratory, University of Cambridge, William Gates Building, 15 JJ Thomson Avenue, Cambridge, U.K.

Keywords: Stereo Vision, Ground Truth, Wearable Sensors, In-Field Assessment, Running, Gait, Foot Kinematics, Continuous Sensing.

Abstract: In-field validation of the accuracy of wearable sensors is desirable since algorithms that perform well in a laboratory setting may not perform as well in real-world use. However, the use cases can be challenging. For example, a foot worn wearable designed to measure foot trajectory should expect to be used in a variety of scenarios ranging from straightforward (running track) to challenging (a woodland area with many undulations). Typically the more challenging the scenario the more difficult it is to get ground truth with conventional systems. We describe a low-cost, highly-portable, point tracking system that can be used where space and infrastructure is limited. The system is built around a pair of commodity video cameras in a stereo setup. We demonstrate how to configure the cameras, a novel technique to approximate shutter synchronisation to sub-frame interval, and we benchmark the system indoors against gold-standard motion capture systems. For a runner 3 m from the cameras were able to recover their foot trajectory with a mean spatial deviation of 1.7 ± 1.1 cm.

1 INTRODUCTION

Recent advances in wearable sensing have driven an interest in in-situ measurement of athletic performance. For the novice or amateur, wearable sensors promise much, including automatic diagnosis of faults in technique; motivation from quantitative performance metrics available at every session; and support for rehabilitation when injured. There is value even at the elite level—where detailed technique and performance analysis is already commonplace—by taking assessments out of the laboratory and into the training or competition venue, moving them into the background and making them more frequent.

A considerable challenge for the development of these sensors is validation outside the laboratory. Our motivation for this work is a foot-mounted inertial sensor that is able to track the gait of runners to high accuracy (Bailey and Harle, 2014a; Bailey and Harle, 2014b). The gold-standard measurement setups for gait typically involve expensive 3D motion capture systems such as Vicon. Unfortunately these systems are bulky, hard to configure, only enable capture in a small volume, and typically struggle outdoors, making them laboratory-bound. Gait is therefore assessed in the laboratory, with the athlete on a treadmill—

an approach we have previously used to evaluate our wearable system (Bailey and Harle, 2014a; Bailey and Harle, 2014b; Bailey and Harle, 2015).

However, the laboratory is often a poor simulation of the real training environment. Treadmills limit movement and force speeds, uneven terrain and gradients are not considered, changes due to fatigue are rarely captured due to short test durations, and the athlete is acutely aware of being assessed, potentially causing a change in behaviour. Because of these differences, a successful evaluation of a system in the laboratory does *not* necessarily imply the same—or even similar—success will be achieved outdoors. Unfortunately it is all too easy for a system to produce an erroneous but plausible result. This is not dissimilar to the situation with general fitness trackers that base activity levels on step counting. In the laboratory the step counters are very accurate because the tests are typically constrained and contrived. In the real world, step counting errors of 20% or more have been regularly reported. However, without a ground truth, users of such systems just trust the number they are given because it has a believable order of magnitude.

In this paper we describe the development and evaluation of a low-cost, highly portable point tracking system—i.e. a system capable of recovering the

3D trajectory of a point within some defined volume. Our system functions outdoors and serves as a valuable in-field ground truth for wearable sensors. The system was designed to facilitate the evaluation of our foot tracking inertial sensors, and we demonstrate its value in this context.

2 BACKGROUND

Video systems are widely used in sports analysis as they are easy to use and provide excellent qualitative data. Quantitative data can be extracted from single camera video systems by calibrating the video, however measurements must take place in a single plane in which a calibration object is present. This can be problematic for sports analysis as measurements may not always occur within the correct plane—many athlete motions are non-planar—leading to measurement error. However, a calibrated stereo vision system is able to estimate non-planar motions in 3D. The cameras must be carefully calibrated to allow triangulation of corresponding points in the two camera frames (Zhang, 2002; Hartley and Zisserman, 2004; Heikkilä and Silvén, 1997). This research area is very mature and robust stereo vision software is easily obtained: Matlab's *Computer Vision Toolkit* and OpenCV are two popular choices.

Stereo vision is not without its issues, however. Calibration is labour intensive, the capture volume is relatively small (wide angle lenses allow for greater capture volumes, at the cost of accuracy), and the two cameras must be synchronised. The level of synchronisation required depends on the desired positioning accuracy and the expected speed of the object being tracked. High-end systems typically synchronise to milliseconds or better using a wired synchronisation signal.

Parallel work in wearable sensors has aimed to keep the portability of video systems while mitigating its limitations. For example, investigation of inertial sensors for foot tracking has been investigated previously for walking (Mariani et al., 2010; Sabatini et al., 2005) and running (Bailey and Harle, 2014a; Bailey and Harle, 2014b). We have previously proposed that such systems should constitute an always-on, in-situ, wearable system for gait assessment in runners. However, these systems have typically been evaluated in a lab-based environment, either on a treadmill, or in a contrived overground environment since the methods of assessment are not easy to use in more challenging environments. However, there is a question as to how such systems would perform in real world environments. A previous study attempted to assess iner-



Figure 1: GoPro Hero4 cameras and jig. The distance between the cameras is 0.25 m.

tial foot mounted trajectory recovery in a track environment (Bichler et al., 2012). The results were poor for some spatial metrics. The reason for this is unclear, although the authors suggest the particular video reference system used was not robust or accurate enough.

3 A LOW-COST PORTABLE STEREO VISION GROUND TRUTH

We sought to develop a lab-validated video reference system to make ground truth measurements in real-life scenarios. The requirements for our stereo vision motion capture system were:

- a fast shutter speed (to minimise motion blur);
- a high frame rate (≥ 100 frames per second (fps) to capture enough detail);
- a large depth of field (to prevent the need for constant refocusing);
- a wide angle lens (to capture more of the run); and
- a small form factor that is robust and easily operated.

We use two GoPro Hero4 consumer cameras capturing 1080p (1920×1080) video at 120 fps. We mounted them to a custom rig as shown in Figure 1. The mounting plates were fixed to the wooden platform, giving an inter-camera spacing of 0.25 m.

A sample pair of images is provided in Figure 2. The configuration allows capture of up to two typical strides with the athlete approximately 3 m away. The GoPro cameras use a wide angle lens to get this field of view, which introduces significant radial distortion. We account for this in the calibration step.

3.1 Camera Calibration

The system must be calibrated to establish both the intrinsic camera parameters (e.g. focal length, distortion, etc.) and the extrinsic parameters (camera positions and poses relative to each other). This is a key



(a) Left (b) Right
Figure 2: A pair of sample frames from the stereo camera system.



(a) Original (b) Undistorted
Figure 3: An original frame with an undistorted frame.

problem in computer vision and we apply a standard solution using a checkerboard pattern held at different positions and orientations in the field of view of the cameras. Once a sufficient number of images of the checkerboard are captured, calibration algorithms can be used to find the necessary parameters. Since we did not require instantaneous trajectory feedback, we ran the calibration software (Matlab's *Computer Vision Toolkit*, <http://uk.mathworks.com/products/computer-vision/>) post-hoc.

To facilitate fast in-field calibration we recorded continuously while the checkerboard was moved around. We then extracted a series of stills to be used for calibration. To avoid camera synchronisation errors influencing the calibration (see section 3.3), the checkerboard was held steady for a couple of seconds at each position where a still was to be taken. Doing so ensured that even a poor synchronisation (e.g. out by multiple frames) will not adversely impact the calibration.

Accurate estimation of the extrinsic parameters is key to an accurate tracking result. However, the cameras had to be removed from their housings to download data. Therefore we took a new calibration sequence with each new measurement session.

Table 1 lists a typical calibration output and Figure 3 shows a captured image and its undistorted output as an example. We note that the two cameras

feature similar intrinsic parameters, as expected for mass-produced hardware. The estimated position of camera 2 relative to camera 1 (247.6826 ± 0.3813 mm) closely matches our manually-measured 250 mm.

3.2 Determining Trajectory

Once the cameras are calibrated and synchronised (see Section 3.3) images can be undistorted and paired. The next step is to identify the 2D pixel coordinates of the point(s) being tracked in each image frame. If each point exhibits high contrast to its surroundings (e.g. a brightly coloured sticker) this process can be automated using standard computer vision techniques. In this work we were tracking a single point post-hoc so we preferred manual point identification to avoid any error that may be introduced by an algorithm. The final step is to triangulate the 3D position of the point relative to (arbitrarily) camera 1. We used the Matlab toolkit's `triangulate` function for this computation.

3.3 Camera Synchronisation

For successful output from a stereo vision system, it is important to synchronise the two camera shutters, or to know the period between the two shutters firing—the *shutter offset*. Full shutter synchronisation is hard

Table 1: Example calibration results.

(a) Intrinsic Parameters

Parameter	Camera 1	Camera 2
Focal length (pixels)	897.24±3.56, 898.41±3.52	894.80±3.51, 895.24±3.49
Principal point (pixels)	991.14±3.79, 552.04±3.08	968.52±3.82, 539.68±2.96
Skew	2.81±0.61	2.74±0.57
Radial distortion	-0.265±0.0020, 0.099±0.002, -0.020±0.001	-0.268±0.002, 0.103±0.002, -0.022±0.001
Tangential distortion	-0.0000±0.0003, 0.0005±0.0002	-0.0000±0.0003, 0.0002±0.0002

(b) Extrinsic Parameters (Relative to Camera 1)

Camera 2 Parameter	Value
Rotation	0.0234±0.0004, - 0.0026±0.0009, - 0.0067±0.0001
Translation (mm)	247.68±0.38, -4.31±0.26, - 6.45±1.08

to achieve without a physical wire between cameras, which is not available on the Hero4 (or almost any consumer-grade camera). *Wireless* synchronisation is offered on the Hero4 via its WiFi radio, but this is not intended for accurate synchronisation—we observed offsets of multiple frames in our tests. This is acceptable for the intended use—creating stereoscopic video—since our visual systems cannot perceive a lag of a few frames at 120 fps. For motion tracking, however, tighter synchronisation is necessary, as we show here.

At 120 fps, the inter-frame period is a little over 8 ms. At its fastest point, a typical jogger’s foot will move approximately 7 cm in this time. For a camera–runner distance of approximately 3 m and a camera separation of 0.25 m as we envisage here, the geometry of the angulation is not forgiving of such an error—see Figure 4. If the foot is at point A when the first (topmost) camera captures an image and at point B (7 cm away from A) when the second camera captures, the triangulation will predict a position at C. This is 37 cm away from the true position in this example. The observed difference will depend on where in the frame the foot is, but errors of this magnitude are not unexpected for the geometry we describe. If we assume a maximum foot speed of 10 ms^{-1} then we need a synchronisation of $\frac{d}{10}$ s to ensure the foot does not move d m between frames. For $d = 5$ mm this requires synchronisation of 0.5 ms or better—i.e. sub-frame synchronisation (or equivalently much

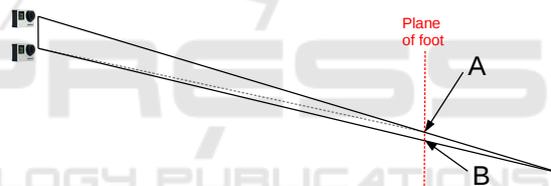


Figure 4: A scale drawing of the geometry for cameras 0.25 m apart and 3 m from the runner.

faster framerate) is necessary to give reliable depth perception.

In order to estimate the shutter offset, we first assume it to be constant—i.e. that the shutter clocks do not drift significantly during an experiment. This is a reasonable proposition for such short durations, validated by the results we present in section 3.5.2. Two challenges remain: to determine the shutter offset value; and to estimate the marker position at the appropriate point *between* frames.

Addressing the latter first, we assume that the movement between two consecutive frames can be approximated by a straight line and we linearly interpolate the position from the measured positions of the marker in the two frames in camera 2 that are either side of the shutter event in camera 1.

To estimate the shutter offset, we have already established that the 120 Hz sampling of the video is insufficient. However, the 48 kHz audio streams can be synchronised to tens of microseconds by

cross-correlating the signal made by a synchronisation sound. Unfortunately, we found the synchronisation between the video and audio streams is itself only approximate. This is common on mass-market consumer cameras; the brain cannot perceive a video/audio offset below approximately 20 ms. There is thus little motivation to increase device complexity and cost to achieve synchronisation better than 2–3 frames.

Our novel synchronisation technique is based on the observation that a given *erroneous* shutter offset will give larger depth estimates for faster moving objects for motion parallel to the camera sensor (‘horizontal’ motion). i.e. *the depth estimate is correlated with the horizontal speed of the object when the shutter offset is incorrect*. When the shutter offset is correct, we expect no correlation between horizontal speed and depth estimate.

To exploit this we record an object accelerating horizontally at approximately constant depth from the camera. We use a simple pendulum, although any *planar* motion will do so long as it has sufficient velocity and acceleration range, ideally incorporating a period of zero velocity. We then consider a sequence of offset values from $-0.5f$ to $+0.5f$ increasing in units of $0.01f$, where f is the frame interval. For each we compute the correlation coefficient between the speed of the motion and the depth estimate. Note that the true object speed is unobservable without the depth information. However the ‘image speed’ (in pixels per second) is an acceptable surrogate since it is proportional and thus exhibits the same correlation properties.

Figure 5(a) shows a typical progression of the correlation coefficient as the shutter offset is changed. The true offset is associated with zero correlation. In the example shown, this corresponds to $0.18f$. Figure 5(b) shows a top-down view of the pendulum trajectory for different offset values. Assuming the pendulum moved only in the vertical plane parallel to the camera sensor, we expect to see a straight line when the offset is correct—we see that an offset of $0.18f$ did indeed produce the expected result. Note that all offsets agree on the depth for the extremes of the motion: this reinforces the observation that the depth of a stationary object (which the pendulum is at either extreme) is independent of the shutter offset accuracy. The full details of the shutter offset determination is given in Algorithm 1.

As an aside we note that the pendulum is, in principle, redundant when the runner passes the camera rig parallel to the camera sensors. In this case limbs will typically exhibit the necessary acceleration range. In practice, we found many amateur runners

Algorithm 1: extract_shutter_offset.

```

input : video1, video from camera 1
input : video2, video from camera 2
output: Shutter offset in (fractional) frames

 $m_1 \leftarrow \text{extract\_marker\_path}(\textit{video1})$ 
 $m_2 \leftarrow \text{extract\_marker\_path}(\textit{video2})$ 
 $f \leftarrow$ 
   $\text{find\_sync\_to\_nearest\_frame}(\textit{video1}, \textit{video2})$ 
 $v \leftarrow \text{differentiate}(\text{low\_pass\_filter}(m_1))$ 
 $\textit{offset} \leftarrow -0.5$ 

while ( $\textit{offset} < 0.5$ ) do
   $rs \leftarrow$  new array
   $m'_2 \leftarrow \text{interpolate\_image\_coords}(m_2, \textit{offset})$ 
   $t \leftarrow \text{extract\_3d\_trajectory}(m_1, m'_2)$ 
   $d \leftarrow \text{low\_pass\_filter}(\text{depth}(t))$ 
   $r \leftarrow \text{pearson\_coefficient}(v, d)$ 
   $rs.append(r)$ 
   $\textit{offset} \leftarrow \textit{offset} + 0.01$ 

return:  $f - 0.5 + 0.01 \times \text{argmin}(rs)$ 

```

Table 2: Median trajectory errors for different step and shutter offsets.

Step no.	Median error (cm)			
	Previous	Nearest	Interpolated	Next
0	7.6	3.6	1.8	15.1
1	7.9	4.1	1.6	16.6
2	9.3	3.0	2.2	16.7
3	10.2	4.8	2.9	16.7
4	9.1	3.9	2.4	16.1

did not keep their limb motions planar and we had more reliable results using an explicit synchronisation process with the pendulum.

Figure 6 illustrates the importance of this synchronisation scheme. It shows the raw trajectories generated from the stereo vision system for a sample step using nearest-frame synchronisation and interpolated synchronisation. These steps were recorded indoors with Vicon ground truth (dashed red lines). We see that the error was predominantly in the depth coordinate. This is due to the camera being side-on to the treadmill.

The interpolated result is also notably closer to the ground truth. We quantitatively assess the error by taking the median value of the distances between corresponding points in the stereo vision and Vicon trajectories. Table 2 shows the results for a series of different steps, confirming that the interpolated offset is at least as good as taking the nearest frame, and usually significantly better.

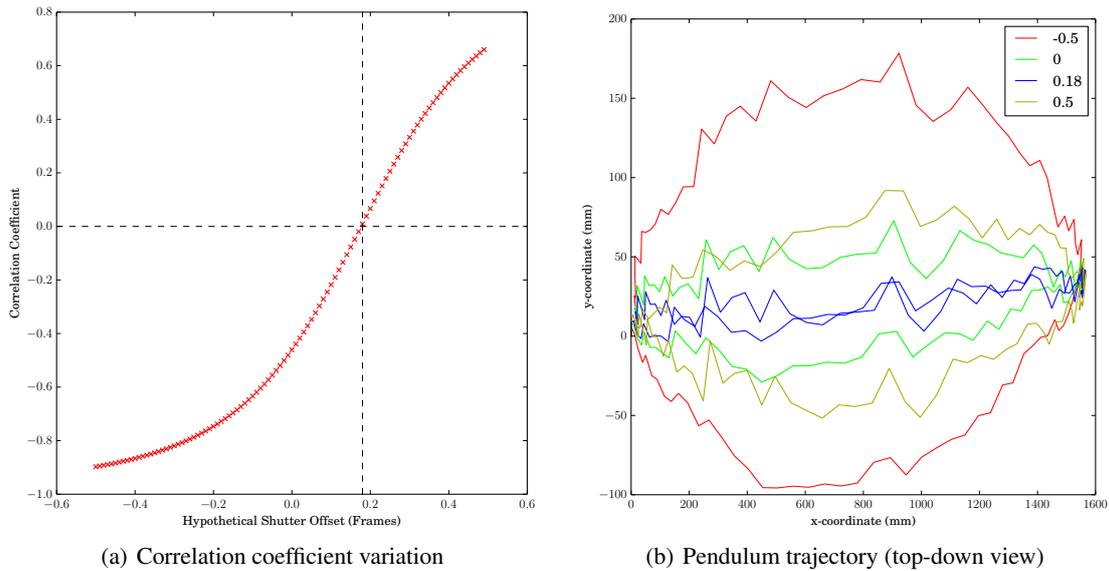


Figure 5: Determining the shutter offset.

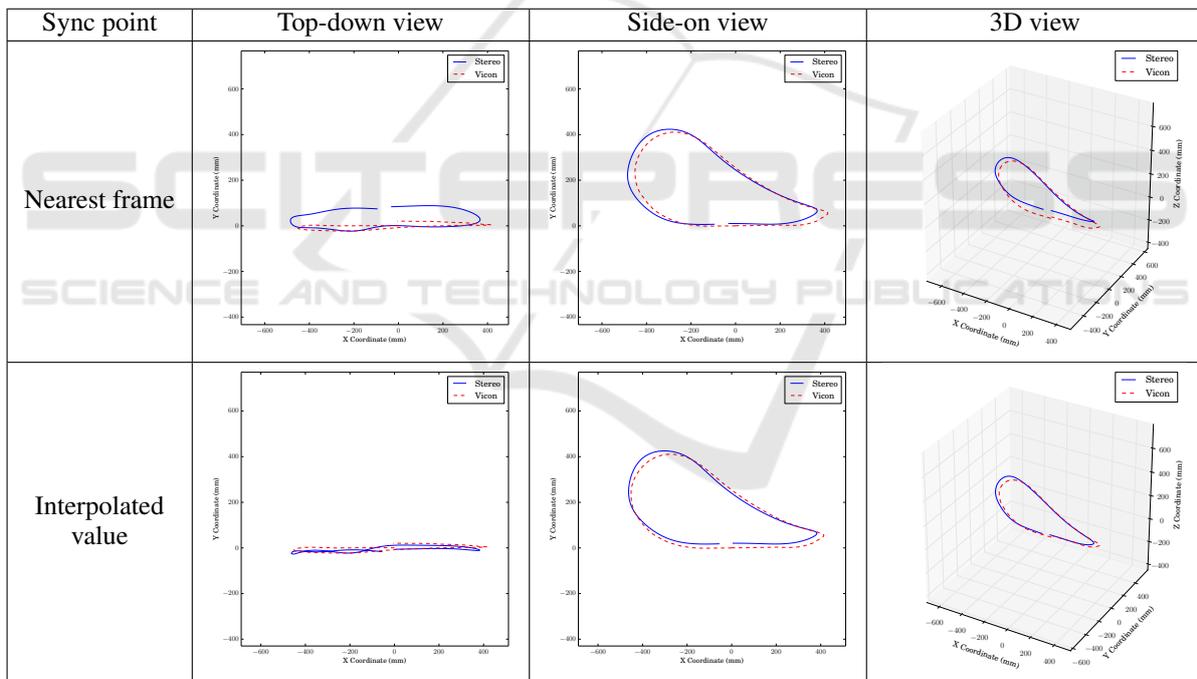


Figure 6: Effect of synchronisation error for an example step (ground truth provided by Vicon).

3.4 Trajectory Smoothing

The trajectory results in Figure 6 demonstrate that even the interpolated synchronisation trajectory has noise of a few centimetres in the depth axis. We expect smooth foot trajectories (as confirmed by the Vicon tracks) so we smooth the stereo vision trajectory. We used a fourth order Butterworth low-pass filter with a 10Hz cut-off frequency. This was found em-

pirically to give the best results (Figure 7).

3.5 Accuracy Evaluation

3.5.1 Method

To assess the accuracy of the stereo camera system, we evaluated it in a lab-based environment. We used a Vicon motion capture system to record the trajectory

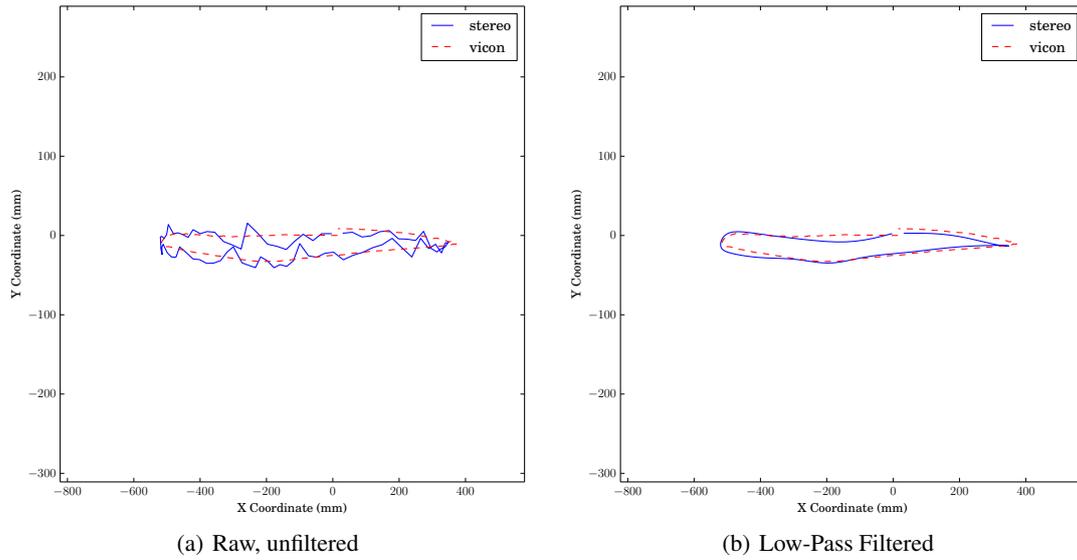


Figure 7: Raw and smoothed stereo vision trajectories vs Vicon ground truth (top-down view). (Depth (y – axis) exaggerated for visibility).

of a marker attached to a treadmill runner’s shoe. We selected a treadmill speed of (3.4 ms^{-1} , a typical running pace). The stereo vision system was used to track the same marker across five steps in each trial. As before a treadmill–camera distance of approximately 3 m was used.

Three runners were recruited and each performed three runs on the treadmill: one where they were in the centre of the vision system’s frame; one to the left; and one to the right (see Figure 8). This resulted in three trials of five steps for three camera angles, and a total of over 3800 3D points for evaluation. To compare the Vicon and vision trajectories, we manually aligned the two co-ordinate systems and looked at the co-ordinate errors.

3.5.2 Accuracy Results and Discussion

Table 3 gives the mean and standard deviation of the errors observed between the Vicon and vision trajectories. We found a 3D euclidean error of approximately 2 ± 1 cm. The errors at the edge of the camera frame (where lens distortion is strongest) are marginally greater but surprisingly similar.

These errors might be expected to improve outdoors, where the higher ambient light level permits a faster shutter speed (the GoPro cameras adapt the shutter speed automatically). Faster shutter speeds correspond to reduced motion blur and thus more accurate pixel co-ordinates of markers. Nonetheless, accuracies around 2 cm are of value to verify the operation of in-situ sensors, especially given the higher portability and significantly reduced cost compared to motion capture systems.

Table 3: Laboratory trajectory errors.

Aspect	Mean error \pm Standard deviation (cm)			
	x-axis	y-axis	z-axis	3D error magnitude
Left	-0.0 ± 0.8	-0.0 ± 1.1	-0.1 ± 1.6	1.8 ± 1.1
Centre	-0.4 ± 0.7	-0.3 ± 1.1	0.3 ± 1.4	1.7 ± 1.1
Right	-0.1 ± 0.8	-0.6 ± 1.2	0.3 ± 2.0	2.1 ± 1.4

4 COMPARISON WITH INERTIAL SENSORS

We have previously used foot mounted inertial sensors to extract spatial (Bailey and Harle, 2014a; Bailey and Harle, 2014b) and temporal (Bailey and Harle, 2015) measurements of running gait. This involves integrating gyroscope and accelerometer data to form a strapdown Inertial Navigation System (INS). These systems require drift to be accounted and corrected for, using de-drifting techniques (Mariani et al., 2010; Bailey and Harle, 2014a), a Kalman Filter (Foxlin, 2005; Bailey and Harle, 2014a) or similar.

We ran a pilot study to compare our inertial results to the stereo vision system. We collected 10 running trials. During each trial the participant ran through the camera field of view allowing capture of a single stride with the vision system. The videos were post-processed to extract 3D foot trajectory by identifying a coloured marker attached to the shoe and co-located with the inertial sensor.

The inertial sensor was a modified Shimmer Shimmer3 Inertial Measurement Unit (IMU, www.shimmersensing.com). We added a 200g accelerometer to the package following the dis-

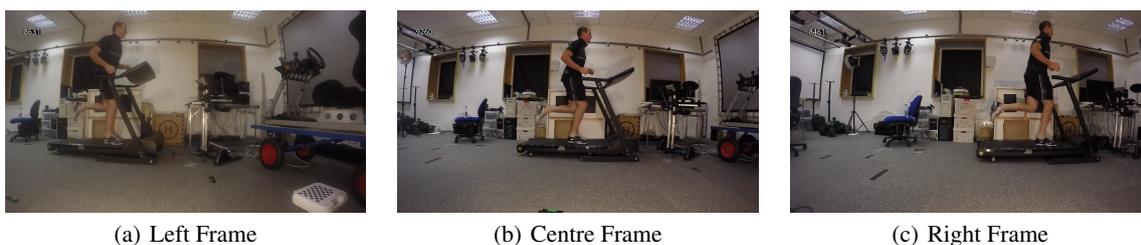


Figure 8: Field of view coverage of the three camera angles tested.

covery that the integrated 16g sensor saturates during a typical running stride (Bailey and Harle, 2014b). We generate individual step trajectories from the inertial data using a de-drifted strapdown algorithm (Mariani et al., 2010; Bailey and Harle, 2014a).

After processing we had 10 overground running steps at a mean running speed of 5.4ms^{-1} that were captured by both the stereo vision and inertial subsystems.

4.1 Results

The inertial system estimates both the foot position and attitude of the foot during each stride, while the stereo vision system is limited to foot position. Since we have verified both the trajectory and attitude estimates in the laboratory, we focus on evaluating the trajectory under the assumption that a good trajectory estimate implies correct attitude.

We used spatial metrics to assess agreement between the two systems. Foot clearance, mean step velocity and stride length were chosen as commonly used statistics for gait assessment (Mariani et al., 2010; Bailey and Harle, 2014a). The results in Table 4 show strong correlation coefficients from measurements taken from each system implying that the stereo vision system is working well and that the algorithms used in treadmill running work as expected for basic overground running on flat ground. This clearly does not constitute a full evaluation of the inertial derived metrics, but serves to show the stereo system can plausibly be used for such a purpose. The full evaluation falls outside of the scope of this paper and is left as future work.

Table 4: Mean and Standard Deviation of Error.

Measurement	Error	Correlation Coefficient
Mean Step Velocity	$-0.28 \pm 0.07\text{ms}^{-1}$	0.99
Stride Length	$-0.1 \pm 5.2\text{cm}$	0.96
Foot Clearance	$-4.2 \pm 2.1\text{cm}$	0.87

5 CONCLUSIONS AND FURTHER WORK

We have shown that off-the-shelf commodity video hardware can be used to track a point using a stereo camera setup with centimetre level accuracy. This is difficult due to the issue of synchronisation between cameras, but our novel approach to camera synchronisation makes it possible to use such a set up for athletic activities despite the faster motion inherent in these scenarios. We have used the system to assess foot kinematics in overground running and compared them to a previously evaluated wearable system during a small pilot study. Further work should use this system to fully evaluate the ability of foot mounted inertial sensors to assess spatial metrics in challenging environments including on slopes or undulating ground.

ACKNOWLEDGEMENTS

This work was funded by Qualcomm Inc. In addition to financial support the authors would like to recognise the technical contributions of the Advanced Location Algorithms group at Qualcomm Cambridge.

REFERENCES

Bailey, G. P. and Harle, R. (2014a). Assessment of foot kinematics during steady state running using a foot-mounted IMU. In *Procedia Engineering*, volume 72, pages 32–37.

Bailey, G. P. and Harle, R. K. (2014b). Investigation of Sensor Parameters for Kinematic Assessment of Steady State Running Using Foot Mounted IMUs. In *Proceedings of the 2nd International Congress on Sports Sciences Research and Technology Support*, pages 154–161. SCITEPRESS - Science and Technology Publications.

Bailey, G. P. and Harle, R. K. (2015). Measuring Temporal Parameters of Gait with Foot Mounted IMUs in

- Steady State Running. In *Proceedings of the 3rd International Congress on Sport Sciences Research and Technology Support*, number icSPORTS, pages 24–33. SCITEPRESS - Science and Technology Publications.
- Bichler, S., Ogris, G., Kremser, V., Schwab, F., Knott, S., and Baca, A. (2012). Towards high-precision IMU/GPS-based stride-parameter determination in an outdoor runners' scenario. In *Procedia Engineering*, volume 34, pages 592–597.
- Foxlin, E. (2005). Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Computer Graphics and Applications*, 25(6):38–46.
- Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.
- Heikkilä, J. and Silvén, O. (1997). A Four-step Camera Calibration Procedure with Implicit Image Correction. *Cvpr*, pages 1106–1112.
- Mariani, B., Hoskovec, C., Rochat, S., Büla, C., Penders, J., and Aminian, K. (2010). 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *Journal of Biomechanics*, 43:2999–3006.
- Sabatini, A. M., Martelloni, C., Scapellato, S., and Cavallo, F. (2005). Assessment of walking features from foot inertial sensing. *IEEE Transactions on Biomedical Engineering*, 52:486–494.
- Zhang, Z. (2002). A Flexible New Technique for Camera Calibration (Technical Report). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

