

A practical approach to using motion capture in performance dance

Paul Golz
University of Worcester
Institute of Sport and Exercise Science
Henwick Grove, Worcester, WR2 6AJ
p.golz@worc.ac.uk

Chris P. Bowers
University of Worcester
Department of Computing
Henwick Grove, Worcester, WR2 6AJ
c.bowers@worc.ac.uk

Marc Price
University of Worcester
Department of Computing
Henwick Grove, Worcester, WR2 6AJ
m.price@worc.ac.uk

1. INTRODUCTION

Motion capture is a key element of many aspects of digital performance including virtual and augmented realities (Golz & Shaw 2014; Chan, Leung, Tang & Komura 2011), digital art and visualisation (Gibson 2011). Extensive research and praxis exists on marked systems within live performance however the use of markers necessarily changes the aesthetic of the performance work (Hutchison & Vincs 2013). Whilst markerless motion capture systems are available, traditionally there has been significant overhead in terms of hardware, set-up, processing power and cost. However, with the introduction of low-cost consumer depth sensors such as Microsoft's Kinect in 2010 and Kinect2 in 2013, artists are now exploring the possibility of real-time motion capture within live performance applications (Golz & Smith-Nunes 2015).

A key element of motion capture within live performance is the ability to carry out real-time processing of the captured data. Indeed, a perceived advantage of the Kinect hardware is the availability of low-bandwidth, pre-processed skeletal model data. Here we will highlight the capabilities and limitations of various data streams produced by the Kinect2 within live dance performance. Whilst this work is applicable to all performance arts, dance is traditionally difficult to accurately motion capture due to the innate characteristics of the form. These include the dynamic and spatial range, multiple performers, contact and occlusion between performers, use of inversion and floor work and the presence of props.

2. PERFORMANCE CAPABILITIES

The Kinect2 captures high definition 1920x1080 pixel visible video and 512x424 infrared video, both at 30fps. As infrared images are more robust to changes in lighting conditions this makes them a useful addition in performance conditions.

The device utilises a time-of-flight camera to capture a depth image of the scene at 512x424 separate points and computes real-time cloud point data representing the 3D space in front of the camera. From this data, it attempts to track the position and state of 26 skeletal joints of up to six individual performers in real-time.

The combined data stream from the Kinect2 is around 2Gbs⁻¹ containing five separate components: real colour, IR, cloud point data, skeletal model data, and body mask (identification of which cloud point data elements belong to a body). Using the Kinect SDK it is straightforward to combine data from various streams (Figure 1), though any significant processing of the data streams is likely to decrease the frame rate¹. Even at full frame rate (30 fps) lag is evident within fast moving dance (Figure 2).

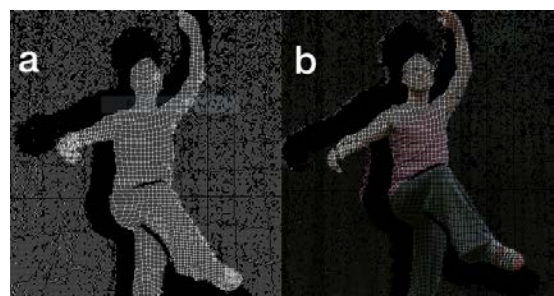


Figure 1: a) Cloud point data; b) Cloud point data colourised with data from real colour stream



Figure 2: Screen shows a 0.2s lag behind the jump

¹ Dependent on code quality and processing hardware.

The exact algorithm that is used for calculating the skeletal data is proprietary, however it is understood to make assumptions about likely skeletal arrangements and, as such, fails dramatically with many of the non-standard arrangements within dance, such as inversion, occlusion, floor-work and contact (Figure 3).

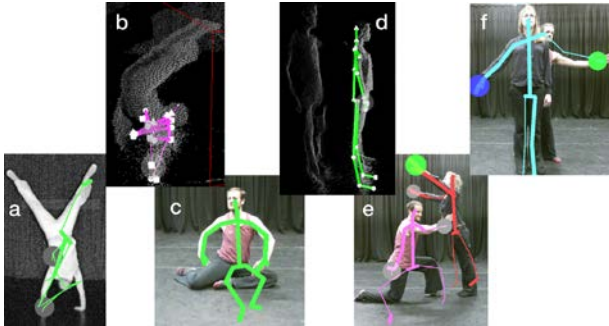


Figure 3: Errors in skeletal data. a & b) inversion c) floor-work, d & f) occlusion (d shown from side), e) contact

In these scenarios both the cloud point and body mask data are more reliable than the skeletal data, for example whilst the skeletal data is inaccurate in Figure 3a-e the cloud point data and body mask are correct.

The body mask data likely looks for contiguous cloud data points and is accurate when the skeletal data is also available (Figure 3e), however when skeletal model fails it identifies all contacted objects as part of the same body (Figure 3f), which can also be an issue in the case of props/set (Figure 4).

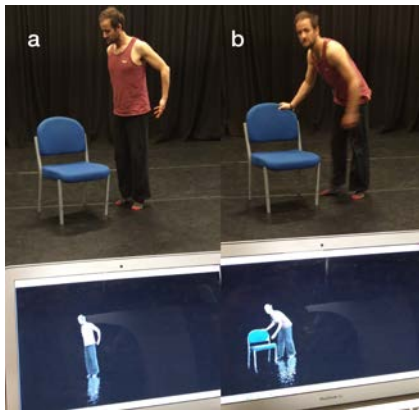


Figure 4: Only elements with body mask are rendered – a) chair is not rendered, b) chair is identified with body mask 1 and displayed

A Kinect2 camera has a 70° horizontal field of view, and is reported to be able to resolve depth images in a range of 0.5 – 4.5 meters. We found significant noise in the cloud point data at short ranges (<1m) (Figure 5). The far distance limit appears to be a software limit: the body mask and skeletal data vanish at this point. Cloud point data remains fully accessible (though obviously the camera resolution can be significant).

We observed a comfortable, workable depth of around 8m with an upstage width of around 11m. Although a

reasonable space for dance, downstage right and left cannot be captured due to the effective field of view.

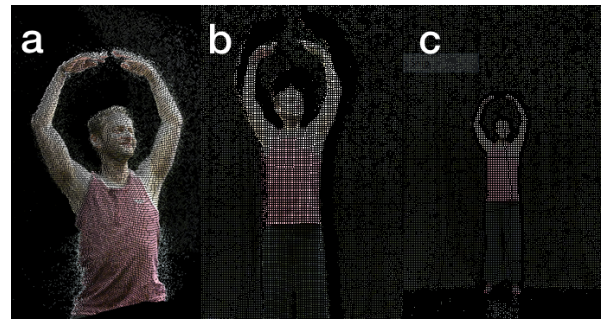


Figure 5: Cloud point data at a) 1m, b) 4.5m, c) 8m

3. CONCLUSIONS

The Kinect2 represents a huge step forward in the capacity to capture and process live performance, however it does have limitations. Most significantly the skeletal model data should not be relied upon, whereas the cloud point and body mask data are more reliable. Occlusion will always be a problem for any single device approach, and combining cloud point data from multiple Kinect2's may partially resolve this (Kitsikidis, Dimitropoulos, Douka & Grammalidis 2014). Even with these limitations, it is likely that performance artists will find the Kinect2 a good solution for many of their motion capture requirements.

4. REFERENCES

- Chan, J.C.P., Leung, H., Tang, J.K.T. & Komura, T., 2011. A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Transactions on Learning Technologies*, 4(2), pp.187–195.
- Gibson, R., 2011. Capturing Stillness: Visualisations of dance through motion / performance capture. In *Electronic Visualisation and the Arts*. pp. 210–212.
- Golz, P. & Shaw, A., 2014. Augmenting live performance dance through mobile technology. In *BCS-HCI '14 Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014*. pp. 311–316.
- Golz, P. & Smith-Nunes, G., 2015. [arra]stre: A data driven ballet. In *Electronic Visualisation and the Arts*. pp. 90–91.
- Hutchison, S. & Vincs, K., 2013. Dancing in Suits: a Performer's Perspective. In *Proceedings of the 19th International Symposium of Electronic Art*.
- Kitsikidis, A., Dimitropoulos, K., Douka, S. & Grammalidis, N., 2014. Dance analysis using multiple kinect sensors. In *2014 International Conference on Computer Vision Theory and Applications*. pp. 789–795.