Evaluating the Feasibility of Low Cost Motion Capture Systems

Alexandre G. Szykman João P. Gois

Federal University of ABC, Dept. of Mathematics, Computation and Cognition, Brazil

Abstract

Recent technological advances have enabled several alternatives for developing motion capture systems at low cost. These systems, in turn, can be used in different applications, such as independent games or movies. This paper aims at evaluating and discussing about this statement. Based on the evaluation, it proposes what would be the best cost-benefit system that can be build with the current available technology.

Keywords: motion capture, low cost, infrared sensors, depth sensors

Authors' contact:
alexandre.szykman@aluno.ufabc.edu.br
joao.gois@ufabc.edu.br

1. Introduction

"I'm not interested in being an animator... That's what Pixar does. What I do is speaking with actors: Here's the scene, let's see what you can come up with it..." - Said James Cameron while talking about the production of "Avatar" - budget of US$237 million, box office of US$2.8 billion [Sito 2013]. The technology behind the words of Cameron is named motion capture, mocap for short, or human motion capture (HMC), more precisely in this case [Calderita et. Al 2013]. Sito [2013] explains that mocap is the technique of covering actors with sensors and recording their movements to be used as base for animation of figures modeled by Computer Graphics. This is a contemporary, yet classically explored, topic. Up to our knowledge, Jules Marey was the first to capture the movements of a living being [Sitio 2013]. The study was made with a black leopard with major joints painted in white and a photographic gun in 1882. Sito [2013] also argues that the modern motion capture was developed by scientists who did not want to depend on artists, people with "aberrant and incompatible" genes to make animations. Besides having to contradict this common sense that "science and art are not compatible", the cost can be a major barrier for people who want to make these fields work together to create animations. Hollywood's level mocap-based animations are obviously overpriced, and these innovators "scientist-artists" are usually independent professionals whose profit is not big enough for such. Recent research, as performed by Dominic [2013], however, shows that new devices and concepts, resulting from technological advances, provide them alternatives to the use of mocap.

2. Related Work

A starting point for assessing the viability of low-cost motion capture is to understand how the already known high technology equipment is budgeted in *Hollywood* and study which alternatives can be found. A *high-end mocap* kit (basically composed of cameras and software*)*, as the *Vicon MX-f20,* costs around US$250,000 [Vicon 2014]. It seems expensive for an independent producer. However, once that there was a huge increase in the demand for *mocap* for both animation and biomechanics industry, not only the hardware manufacturing process has been optimized, but also sales and support contracts have been reduced. Consequently, prices were reduced [Dominic 2013]. The release of the kit *OptiTrack Flex* System V100R2 at US$15,000 is a good example [Natural Point 2014]. Dominic [2013] compares static accuracy and quality of linear and angular kinematic momentum of the two systems. After finding minimal differences, the author suggested that the low-cost systems tend to be ruling the market. Based on this information, it can be assumed that the alternative *mocap* systems, such as the *OptiTrack* (at US$15,000), have the minimum quality required by independent producers. The objective would be, therefore, to explore other types of systems to reduce this cost as much as possible without decreasing too much its quality.

2.1 Classifications of *Mocap*

Considering the technology and concepts used, there are different ways of classifying *human motion* capture. Bodies in movement can be tracked with the help of markers in a setup named *online,* or without them, in an *offline* se*tup*. *Online* setups are generally more expensive and only make sense when used in environments very well controlled, such as *Hollywood's* studios. For low cost projects, only the *offline* setup must be considered. Such configuration, however, requires more intense computational processes [Castro 2006]. Generally, algorithms for offline tracking bodies, consist of, frame-by-frame, first estimating and extracting the background, then, the silhouettes of human bodies, their poses in 2D and, finally, their 3D poses. With the 3D poses, the coordinates of the joints are recognized and linked into a virtual skeleton. A kinematic skeleton is created by building up the skeletons extracted from every frame. This skeleton can then be rigged to the objects modeled in 3D [Gudukbay et al. 2013]. Another way of classifying h*uman motion capture* is by tracking one or more than one person simultaneously. Tracking multiple people have been shown as much more complex than tracking one. One of the most complete studies in this area was done by *Liu et al.* [2013]: An algorithm that tracks movements of people interacting closely was developed. However, some issues, such as failures to resolve ambiguities when people are in physical contact and long time required to process the data still persist. The last classification of *human motion capture* is the tracking being accomplished by mapping the scene directly or with the help of algorithms that recognize common silhouettes in humans by using a library of predefined data. These models are respectively named *model-free* or *model-based* approaches. *Model-free* simply track points previously defined, so they are simplest and most common. *Model-based* compares the model being tracked with a predefined human model during the capture process; consequently, they are generally faster and more robust. *Calderita et al. [2013]* proposed a mixed approach to take advantage of both concepts. In one of his experiments, they used an already-developed *model-free* approach to perform the *mocap*, and a *model-based* approach to fix some problems of the original algorithm. An affordable and effcient *mocap* system must use an *off-line* configuration, make use of the balance between *model-free* and *model-based* approaches, and be able to track more than one person simultaneously.

2.2 Infrared Sensors

So far, only motion capture kits with cameras was mentioned. However, in the context of production of games and animations at low cost, depth sensors generally provide data with better quality than monocular video sensors (cameras) [Wang et al 2013]. This fact is also illustrated by *Conlin* [2010] when showing the lack of quality for real time *mocap* in Games of the camera *"Playstation Eye", Sony,* in comparison to the *Microsoft's "Kinect for* Xbox" depth sensor. Depth sensors radiate infrared lights with specular patterns. These patterns are altered when reaching objects and return different values to the sensor. These values are used to create depth maps of scenes [Jorge et al. 2013]. *Dutta* [2011] evaluated the accuracy of the depth sensor mapping from the *Microsoft's "Kinect for Windows*" by comparing it with a *high-end Vicon* kit. It was concluded that, despite its quality being at least one order of magnitude lower than the *Vicon* system, *Kinect* offsets due its portability, simplicity, and, mainly, low cost. An alternative depth sensor to *Kinect* is the *Asus' "Xtion".* The comparison made by *Jorge et al* [2013]. demonstrate that both devices have software functions equivalent in accuracy. Considering the *hardware,* unlike *Xtion, Kinect* includes a small motor to help its positioning during the scene setup. *Kinect* showed very good industrial reliability. The reliability of *Xtion* was not tested. In addition, *Kinect* is cheaper: US$220 versus US$266 for the *Xtion* [AMAZON, 2013]. Another point to be considered is that, due to a *Microsoft* policy, Kinect works with a better support in *Windows Operational System* [Ope13]. Overall, depth sensors are more effective than cameras setups for *mocap* at low cost. Therefore, theoretically, with depth sensors, the minimum quality for independent producer could be kept, but at a considerable lower cost. For development on *Windows*, Kinect has proven to be the best choice among the depth sensors due its possibility of an easy positioning in the scene and low cost.

Similar to the most traditional approaches to motion capture, depth mapping with *Kinect* exhibits instability when tracking bodies in fast movements, rotating, or whose parts are occluded - occlusions are common in tracking mutual persons or when the tracked person is interacting with objects. The consequences are blurred results, incomplete tracking or false positives - when the system tracks data that do not belong to the target body. There are two main causes for these problems. Firstly, *Kinect* has a single infrared sensor. Therefore, if there is any obstruction in its light beams, the target will not be correctly tracked. Secondly, *Kinect* does not recognize human bodies as single assemblies during the scanning. The body parts are tracked independently, increasing the possibility of distortion of its segments. *Shum et al.* measured the severity of these problems and proposed an algorithm to reduce them. The code works, but it requires intensive and expensive computational processes [Shum et al. 2013]. *Wang et al.* [2013] increased the robustness of the system by introducing an algorithm that helps *Kinect* to recognize not only humans, but also its movements when interacting with daily life objects. These movements were categorized and named *actionlets*. *Data mining* was used to correlate the actions database with the tracked body and building the *actionlet set model* - the final captured skeleton. This model proved to be resilient to noise and temporal misalignment. Besides the technique introduced, *Wang et al.* [2013] use multiple *Kinect* devices simultaneously to improve the tracking accuracy. The efficiency of the mutual configuration of sensors is reaffirmed by some companies specialized in *mocap*, such as *IPISoft LLC* [IPI SOFT LLC 2012]. It is a fact, therefore, that there are some problems in tracking with *Kinect,* but they can be minimized through the use of correction algorithms and mutual sensors.

3. Current Offered Solutions

Given the complexity of algorithms for processing motion capture, hybrid independent artists-scientists may not have the time needed to develop them alone. Turnkey solutions offered by the market can then be considered. IPISoft provides systems with multiple depth sensors. It offers three types of software, differing in accuracy and features. The "Express Edition", which costs US$295, uses a depth sensor and tracks simple movements of one person. For

US$595, the "Basic Version" employs two devices and can track complex movements of one user. Finally, for US$1195, the "Standard Edition" uses two depth sensors to track any movement of several people mutually. In addition, the software supports the use of extra equipment, such as the "Playstation Move", Sony, to improve the tracking of the hands. Although the IPI Soft offers good solutions, its software tracks just bodies; there is no tracking for facial expressions [IPI SOFT LLC 2012]. The software "NuiCapture 1.4", NuiCaptureInc, is an alternative that offers face tracking and the same features found in the Standard Edition of the IPISoft for US$399. NuiCapture supports only Kinect as depth sensor [NuiCapture 2013]. Kinect, in turn, has a Software Development Kit that allows the creation of applications for both body and face tracking, it is named "Microsoft Kinect for Windows SDK 1.8" [Microsoft 2013]. There are, therefore, commercial solutions for various categories of mocap. Meanwhile, there is also the possibility of developing these solutions for free via, for example, the Microsoft Kinect for Windows SDK.

There are alternatives to the Microsoft Kinect for Windows SDK. OpenNI is a non-profit, industry-led, organization that offers open source framework to develop middleware for mocap [Jorge et al. 2013]. Built up with this middleware, the Development Kit "NITE 2.2" tracks mutual people with mutual devices. As OpenNI is sponsored by PrimeSenseLTD and ASUSTec Compute Inc, companies that have their own depth sensors ("Carmime" and "Xtion", respectively), its kit originally supports only their devices. There is, however, a free adaptation algorithm for Kinect to be used with OpenNI [OpenNI 2013]. A final alternative to Microsoft Kinect for Windows SDK is provided by Microsoft itself: a new Kinect model, named Kinect V2, launched in the USA in July of 2014 at US$199 [Microsoft Store 2014]. This model has its own kit, the "Microsoft Kinect for Windows SDK 2.0", and provides enhanced scope and quality in comparison to the original model. Because of the required bandwidth, though, applications with multiple Kinect V2 referring to the same computer are not possible yet. Its prototype development kit is currently available for select developers [Microsoft 2014]. The best alternative, in summary, by considering cost-benefit would be developing with the Kinect V2 and its SDK. For configurations with mutual sensors, hybrid applications with sensors Kinect V1 and V2 are considerable. In Fig. 1, it is possible to see the sensor of Kinect V2 emitting infrared lights. In Fig. 2, it is illustrated the skeleton tracked with the sensor.

Figure 1: Kinect V2 emitting infrared lights

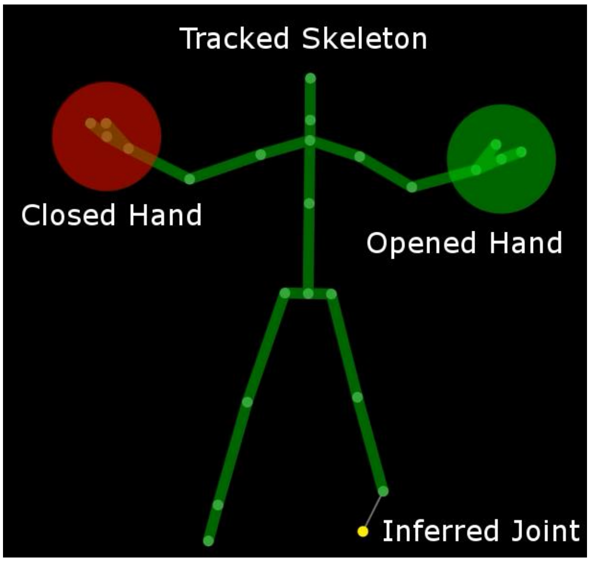

Figure 2: Example of skeleton tracking

4. Conclusion

Low cost motion capture settings are available today. These settings are possible due to recent advances in technology and intense demand in the market. They do not achieve the same accuracy of Hollywood's high-end equipment, but, at a cost hundreds of times lower, reach a satisfactory level for applications such as independent films and games. Considering the trade-off between cost and quality, the best configuration evaluated is using a setup formed by one Kinect V2, possible auxiliary sensors Kinect V1 and their respective SDKs. The proposed system can be developed, tested and evaluated as future work.

Acknowledgements

The authors gratefully thank the Federal University of ABC for the financial support for this paper.

References

AMAZON, 2013. Online Shopping for Electronics, Apparel, Computers, Books, DVDs & more. Available from: http://www.amazon.com [Accessed 8 November 2013]

CALDERITA, L.V., BANDERA J.P., BUSTOS P., SKIADOPOULOS, A., 2013. Model-based reinforcement of date for Kinect human motion capture applications. *Sensors, 13 (7):. 8835-8855*. Available from: http://www.mdi.com/journal/sensors [Accessed 23 October 2013]

CASTRO, J., 2006. Design and evaluation of a new three-dimensional motion capture system. *Gait & Posture, p. 126 -. 129.* Available from: https://uark.illiad.oclc.org/illiad/illiad.dll?Action=10&Form=75&Value=778807 [Accessed 10 October 2013]

CONLIN, S., 2010. Makes its Playstation Move to catch up. *Toronto Star (Canada), pp. 1-2*.

DOMINIC, T., 2013. Next-Generation Low-Cost Motion Capture Systems Can Provide Spatial Accuracy Comparable to High-End Systems. *Journal of Applied Biomechanics, 1, 112-117.*

DUTTA, T., 2011. Evaluation of the Kinect sensor for 3-D kinematic measurement. *Applied Ergonomics, 1, 645 -. 649.*

GUDUKBAY, U., DEMIR, I., DEDEOGLU, Y., 2013. Motion Capture and human pose reconstruction from a single-view video. *Digital Signal Processing, 23, 1441 -. 1448*.

IPI SOFT LLC, 2012. iPi Motion Capture Releases Version 2.0. *Business Wire, pp. 1 – 3.*

JORGE, H., G., RIVERO, B., FERNANDEZ E.V., SANCHEZ, J.M., ARIAS, P., 2013. Metrological evaluation of Kinect sensor and Asus Xtion sensors. *Measurement, 46 (6), 1800 -. 11806.*

LIU, Y., DAI, Q., GALL, J., STOLL, C., SEIDEL, H., THEOBALT. MARKERLESS, C., 2013. Motion Capture of Multiple Characters Using Multiview Image Segmentation. *IEE Transactions on Pattern Analysis and Machine Intelligence, 35, 2720 -2735.*

MICROSOFT, 2013. Face Tracking, 2013. Available from http://msdn microsoft.com/en-us/library/jj130970.aspx [Accessed 8 November 2013]

MICROSOFT, 2014. Kinect for Windows v2 Developer Preview Program. Available from http://www.microsoft.com/en-us/kinectforwindowsdev/newdevkit.aspx [Accessed 31 May 2014]

MICROSOFT STORE, 2014. Pre-order the Kinect for Windows v2 sensor. Available from: http://www.microsoft.com/en-us/kinectforwindows/Purchase/developer-sku.aspx [Accessed 5 July 2014]

NATURALPOINT, 2014. TrackIR. Available from: http://naturalpoint.com [Accessed 3 June 3 2014]

NUICAPTURE, 2013. nuiCapture - Record, export, and playback Kinect data for analysis effortlessly. Available from: http://nuicapture.com [Accessed 8 November 2013]

OpenNI, 2013. OPENNI NITE. Available from: http://www.openni.org/files/nite [Accessed 8 November 2013]

SHUM H.P., HO, E.S., JIANG, Y., TAKAGI, S., 2013. Real-Time Posture Reconstruction for Microsoft Kinect. *Cybernectics IEEE Transactions on, 43 (5), 1357 -. 1369.* Available from: http://rapidill.org/redir.ashx?id=NTYyMDMz [Accessed 23 October 2013]

SITO, T., 2013. Motion Capture: The Uncanny Hybrid. Moving Innovation: A History of Computer Animation *(pp. 199-215). Cambridge, MA, USA:. MIT Press.*

VICON, 2014. Quantic Dream. Available from: [Accessed 3 June 2014]

WANG, J., LIU, Z., WU, Y., YUAN, J., 2013. Actionlet Ensemble Learning for 3D Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, 1 – 14.*