Motion Capture: An Evaluation of Kinect V2 Body Tracking for Upper Limb Motion Analysis

Silvio Giancola[1(✉)], Andrea Corti[1], Franco Molteni[2], and Remo Sala[1]

[1] Vision Bricks Laboratory, Dipartimento di Meccanica, Politecnico di Milano, Via La Masa, 1, 20156 Milan, Italy
{silvio.giancola,remo.sala}@polimi.it, andreacorti@outlook.it
[2] Movement Analysis Lab of Valduce Hospital, "Villa Beretta" Rehabilitation Centre, Via Nazario Sauro, 17, 23845 Costa Masnaga, LC, Italy
franco56molteni@gmail.com

Abstract. In this study, we evaluate the performances of the body tracking algorithm of the Kinect V2 low-cost time-of-flight camera for medical rehabilitation purposes. Kinect V2 is an affordable motion capture system, capable to monitor patients ability to perform the exercise programs at home after a training period inside the hospital, which is more convenient and comfortable for them. In order to verify the reliability of the body tracking algorithm of the Kinect V2, it has been compared with an actual stereophotogrammetric optoelectronic 3D motion capture system, routinely used in a Motion Analysis Laboratory in a Rehabilitation Centre, focusing on the upper limb rehabilitation process. The results obtained from the analysis reveal that the device is suitable for the rehabilitation application and, more generally, for all the applications in which the required accuracy related to the joint position does not exceed a couple of centimetres.

Keywords: Motion capture · Kinect V2 · Stereoscopic system

1 Introduction

Due to nervous, muscle's or skeletal system lesions, people may lose a part of their motor control with an impairment of abilities and performances. In order to try to restore those functions, intensive, complex and long term rehabilitation procedures are necessary.

Depending on the distance from the onset of the disability and the level of impairment, the rehabilitation team plans short-medium-long term rehab program that the patients start to perform as in or out-patient. The rehabilitation team have to maintain under control the training session and to perform regular follow up. They need to monitor functional changes of the ability of the patients in order to select the best exercise programs that fit the level of difficulty depending on the patient abilities. For neurorehabilitation, the duration of the treatment can last several months.

P. Perego et al. (Eds.): MobiHealth 2016, LNICST 192, pp. 302–309, 2017.
DOI: 10.1007/978-3-319-58877-3_39

Regarding upper limb rehabilitation, task and goal-directed exercises are planned to involve arms, forearms, shoulders and backbone, executing activity of daily life. Patients are asked to control upper limb to manipulate small objects and/or to explore peripersonal and personal space using an echological approach to increase the engagement of the patient. Serious game [1], with the support of Virtual Reality (VR), helps patients to project themselves in a virtual world where they can naturally interact and the surrounding environment will not interfere with the neutrality of their movements.

Motion analysis focuses on the biomechanical study of the human body. The skeleton can be interpreted as a complex multi-body system composed of bones, linked together in joints and actuated by muscles. Shippen [2–4] presents a complex model composed of 31 rigid body and 35 joints actuated by 539 muscles for human motion analyses in sports and dance.

Direct and inverse kinematics and dynamics study analyses the body capability to perform a defined movement. Muscle activity are generally tracked with electromyography (EMG) that transforms contractions in an interpretable signal [5]. Regarding motion tracking, different techniques already exist [6].

In the late 1970's, Bajd et al. [7] developed electrogoniometers based on potentiometers able to record joint angles and to realise online gait analysis of the lower limb. This technique was improved in the 1980's with the use of triaxial goniometers, with a complex setup. In the same period, Furnee et al. [8] and successively Jarret et al. [9] were starting to use computer vision system to register human motion. They were using markers on people and animal body in order to tracking their motion in 2D and 3D spaces. Vision-based systems are contact-less methods, that does not introduce any load for the patient, more suitable for rehabilitation since patients keep their movement's freedom.

Our study compare two vision-based systems for human body tracking. The ground truth one is a Multi-View Stereoscopy (MVS) system that tracks markers in space. The position of those markers approximates selected joints that represent the articulation of a partial skeleton. Since it is the most widely used system for human body tracking, due to its maturity and its 0.1 mm accuracy, it will be used for reference data. The second system is a 3D Time-of-Flight (TOF) camera that evaluate the human body position in space from a measured point cloud. Similar comparison with MVS systems has been done with structured light depth camera in [10], providing interesting results. Since TOF technology provides better results in depth measurement [11], we are expecting improvements for body tracking precision.

In a first part we will present both techniques. Then, we will analyse the Kinect V2 TOF performances for absolute position and relative orientation estimation respect to a BTS Smart-DX 7000 MVS system. Finally, uncertainty in position tracking will be estimated for the Kinect V2 system.

2 Vision System for Gait Analysis

Vision systems are non-contact optical measurement techniques that do not introduce any loading effects like IMUs or goniometers do and that could lead

in changing the mechanical properties of the multi-body system and impede the patient natural movement.

2.1 Multi-view Stereoscopy: BTS Smart-DX 7000

The multi-view stereoscopy is an active vision technique that use 2 or more cameras for tracking markers in space. They are typically composed of infrared (IR) emitters that enlighten the camera field of view and IR filters that permeate other light wavelength than the one emitted. Reflective objects, typically spheres, are emphasised and tracked from multiple points of view. Using epipolar geometry between the cameras, single points are reconstructed in 3D.

In order to evaluate the performance of the Kinect V2 body tracking algorithm, we use the BTS multi-view stereo system as ground truth measurement provided by the Movement Analysis Lab of Valduce Hospital "Villa Beretta" Rehabilitation Centre in Costa Masnaga, Lecco, Italy. It is a multi-view stereo system composed of 8 cameras, with a resolution of 2048 * 2048 pixel each and a maximum frame rate of 250 fps. The lightening system strobe a 850 nm wavelength light on spherical markers fixed on the patient body. The setup provides marker position measurement in a 6 * 6 * 3 m working space with an uncertainty of 0.1 mm. The cameras are beforehand calibrated in order to correct eventual lens distortion and register the cameras respect to a common reference system.

In order to get the joint position in space, markers have to be placed astutely on he patient body in order to measure the actual articulation. Many marker placement exists: Body Segment CM, Plug-in-gait (Vycon), Helen Hayes (Davis), Cleveland Clinic and Golfer Full-Body are the more common. They typically use multiple markers for a single joint in order to return a better identification of the articulation position. In most rehabilitation cases, simpler marker placement are preferred, using a single marker per joint in order to reduce preparation time. A limited number of markers does not reconstruct the exact position of the centre of a joint, but guaranty an acceptable similarity with the real movement, with a satisfying repeatability. Also, the expected performances are usually not reached for the articulation measurements since markers are fixed on soft tissues that are not rigidly fixed with the skeleton.

2.2 Time-of-Flight Camera: Microsoft Kinect V2

Time-of-Flight cameras are depth sensors that return dense point clouds. A TOF camera is composed of a pulsed cIR lightening system and an IR matrix sensor that measures the phase between the codified light sent and the received one for every single pixel. The phase between emitted and received signals is actually proportional to the distance covered by the light back-and-forth. Every single pixel measure the distance of the first obstacle it sense; put together it returns a 2.5D representation of a scene.

The second version of the Microsoft Kinect (Kinect V2) is an RGB-D camera based on the TOF technology. The Kinect V2 is composed of a 512 * 424 pixel TOF IR camera and a 1080 * 1920 pixel RGB camera. They are registered and

(a) 25-joint skeleton (b) Kinect V2

Fig. 1. Human body skeleton tracking with Kinect V2 according to Shotton et al. [12]

return a 217.088 organised coloured point cloud at a 30 Hz frequency. The depth goes from 0.5 to 4.5 m and previous study claimed a best-case precision of 1.5 mm for the point cloud reconstruction [11].

The Kinect V2 has an integrated SDK function for markerless human-motion capture based on Shotton et al. [12] algorithm based on SVMs and Randomized Decision Forests. This markerless human-motion tracking method can fully track up to 6 human body simultaneously, defined with 25 joints as shown in Fig. 1, respect to the reference system defined by the TOF sensor.

3 Experimental Setup and Preliminary Results

The Kinect V2 system seems to be a cheap alternative to the expensive MVS technology. Since the 2 vision-based devices have different reference systems, it is necessary to register one system with another, in order to compare both upper limb trajectories in a common reference system.

We define the BTS reference system as the main one since it provides ground truth data. The (X, Y) plane is horizontal and the Z axis is vertical, centred in the room and directed versus top. The Kinect V2 is fixed on a tripod that frames the pedestal where the patient will be tracked.

For the registration, a set of markers composed of a thin black 80 mm-diameter disk and a 15 mm semi-sphere are disposed casually on the scene, in order to measure the position of those custom markers with both systems. The Kinect V2 IR camera identify the black disk and its barycentre is measured through a blob analysis with sub-pixel precision, which is then reprojected into the point cloud in order to obtain the 3D points in the Kinect V2 reference system. The BTS system directly returns the 3D position of semi-sphere in its own reference system. The 2 set of points are then aligned through the solving of the Procrustes problem [13] with an SVD-based algorithm [14] that minimise

the root mean squared distances between the sets of markers. The transformation that align the 2 sets of point corresponds to the registration between the 2 reference systems.

For the comparison, the person to track sit down on a chair placed at around 2.5 m of the TOF camera. Regarding the body motion measurement, spherical reflective markers are placed on the joint to track, following the classical routine for the patients. While the BTS system acquires data at 250 Hz, the Kinect V2 is limited at 30 Hz. We have interpolated and re-sampled Kinect V2 data at 250 Hz, transformed the trajectory in the BTS reference system and synchronises times with the time-stamps and a cross-correlation analyses.

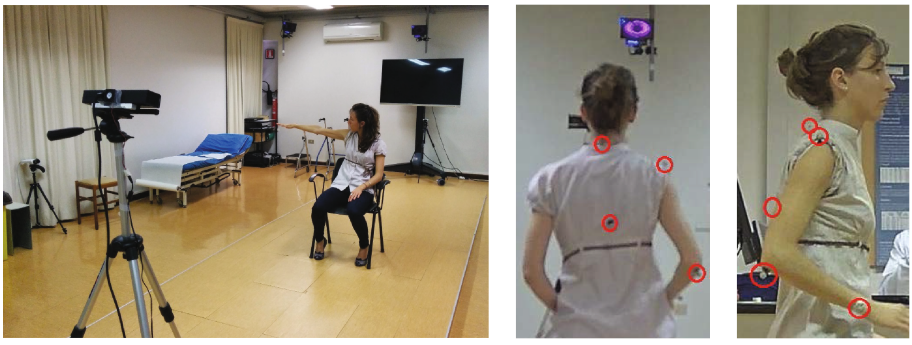
In a first motion recording, the patient is asked to rotate its right arm around the lateral axis of its shoulder 2. The X, Y and Z coordinates of the wrist position are recorded and compared between our systems. Different posture have been tested, frontally and laterally behind the camera, as well as intermediate posture of the body. The Kinect V2 body tracking system seems more accurate when the body is placed in front of the sensor oriented at 45° along the medial axis.

4 Neuro-Rehabilitation Motion Analysis

The following study will focus on comparing performances in motion tracking between the 2 techniques during neuro-rehabilitation exercises.

Exercises for an upper limb rehabilitation program [15] were performed in the "Villa Beretta" rehabilitation centre, the which require the use of 5 reflective markers for the analysis of a single upper limb as shown in Fig. 2. The exercises are 3, during which he performs 10 times the same simple daily life movement in a seated position.

The first exercise is called *"abduction"*, the patient needs to rigidly stand its arm along the lateral axis inside the coronal plane, starting from a relaxed caudal direction. In the second exercise, called *"reaching"*, the patient needs to extend its arm in front of him along the sagittal plan anterior direction, starting from



(a) Movement Analysis Laboratory (b) Back View (c) Lateral View

Fig. 2. Experimental setup and marker placement

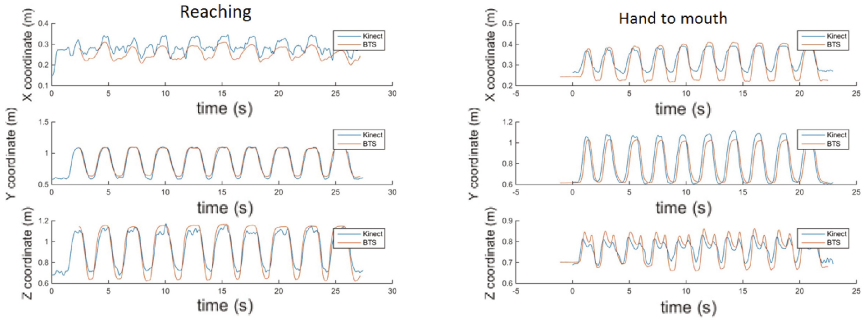


Fig. 3. X, Y and Z coordinates of the wrist position during Reaching and Hand-to-mouth exercises

the same relaxed position than in the previous exercise. In the third exercise, called *"hand-to-mouth"*, the patient is asked to reach his mouth with his hand, starting with his hand on his thigh.

During the exercise, the position in space and time of the wrist, elbow, shoulder, cervical C7 and thoracic T5 vertebrae. The last 2 correspond to the SpineShoulder and SpineMid joints in the Kinect V2 skeleton representation (see Fig. 1a).

Some of the plots of the wrist trajectory during the exercise are shown in Fig. 3. Note that a 5 Hz 2^{nd} order low-pass Butterworth filter has been applied in order to remove high frequency noise in the trajectories.

Even if the trajectories shape look similar, they are not perfectly superimposed. First of all, the position of the marker introduce an offset in the measurement, as seen previously. As long as the markers are placed in the same position on the body, the process can be considered as repeatable. Also, a limb position close to the body is challenging to distinguish in the point cloud, which result in more erroneous full body reconstruction. Finally, a frontal position for the camera produce an occlusion of the torso with the limb, especially in the reaching and the hand-to-mouth exercises.

We can actually denote that the trajectory have a maximum deviation of 20 mm, especially shown in the range of motion extrema, when switching direction occurs. We can assume that the body tracking algorithm takes into consideration both speed and previous positions [12] to estimate the current position of the joint, since it returns worse results when switching direction. Also we can note a systematic error in the trajectories, which has been assigned to the inaccurate marker placement on the ulnar styloid, and not on the centre of the wrist. Nevertheless, the system actually provides a repeatable measurement for the position.

5 Uncertainty Estimation

In this section we estimate the uncertainty for the wrist motion tracking with a Kinect V2 device.

The arm of the patient is kept in a static position along the lateral axis. A metallic structure ensures the immobility of the limb for more than a minute. The position of the wrist is measured at 30 Hz, more than 1800 samples are recorded.

Then, the same 5 Hz 2^{nd} order low-pass Butterworth filter is used to remove high frequency noises, as well as a 0.01 Hz high-pass filter that removes an eventual drift due to the patient strain in staying still during 60 s. Test with different limb attitude has been done, return the same precision order of magnitude.

Precision is defined as the standard deviation of the static measurement, and the root mean square of the standard deviation along the 3 main axis, which has been estimated around 1 mm for the wrist position measurement. Accuracy is not provided since it is impossible to estimate the wrist as a single point, nevertheless an offset of around 20 mm has been estimated in the previous exercises.

Similar test has been carried on the elbow angle, which return a 0.25° precision. We believe the accuracy of the Kinect V2 algorithm for angle measurement is better than the multi-view stereo system with single marker placement at the joints, since a completely extended arm return 180° with the Kinect V2 but only 160° with the BTS system.

6 Conclusion

In this paper the Kinect V2 motion tracking algorithm has been evaluated to analyse movement of the upper limb. It has been applied in a rehabilitation exercise program and compared in terms of precision of detection of the movement with a state-of-the-art MVS marker tracking system already used in medical field.

We found that the Kinect V2 body tracking system has a good 1 mm precision. On the other side, the accuracy is larger but hard to improve due to the difficulty to define the position of the wrist as a single point. In any case, the Kinect V2 accuracy is better than a single marker placement per joint with the multi-view stereo marker tracking system analysis.

Kinect V2 is a markerless technique that reduces preparation time for medical staff. We have shown that this low-cost system is user-friendly, by not being invasive to the patient. We believe patients will be able to use it at home during custom training sessions, associated with serious game frameworks.

Microsoft initially reveals the possibility to contemporary use multiple Kinect V2 systems. Combining information from different system, it is possible to solve the occlusion problem as well as improving the human body tracking performances by meaning information.

Future exploitation of this work will be extended with complete body tracking and real-time inverse dynamics evaluation.

References

1. Abt, C.C.: Serious Games. University Press of America, Lanham (1987)
2. Shippen, J., May, B.: Teaching Biomechanical Analysis Using the Bob Matlab/Simulink Model
3. Shippen, J.M., May, B.: Calculation of muscle loading and joint contact forces during the rock step in irish dance. J. Dance Med. Sci. **14**(1), 11–18 (2010)
4. Wagner, D.W., Stepanyan, V., Shippen, J.M., DeMers, M.S., Gibbons, R.S., Andrews, B.J., Creasey, G.H., Beaupre, G.S.: Consistency among musculoskeletal models: caveat utilitor. Ann. Biomed. Eng. **41**(8), 1787–1799 (2013)
5. Sutherland, D.H.: The evolution of clinical gait analysis part l: kinesiological EMG. Gait Posture **14**(1), 61–70 (2001)
6. Sutherland, D.H.: The evolution of clinical gait analysis: part II kinematics. Gait Posture **16**(2), 159–179 (2002)
7. Bajd, T., Stanič, U., Kljajić, M., Trnkoczy, A.: On-line electrogoniometric gait analysis. Comput. Biomed. Res. **9**(5), 439–444 (1976)
8. Furnée, E., Halbertsma, J., Klunder, G., Miller, S., Nieukerke, K., Van der Burg, J., van der Meché, F.: Proceedings: Automatic analysis of stepping movements in cats by means of a television system and a digital computer. The Journal of Physiology **240**(2), 3P (1974)
9. Jarrett, M., Andrews, B., Paul, J.: A television/computer system for the analysis of human locomotion. In: IERE Golden Jubilee Conference, An exhibition on Application of Electronics in Medicine, Southampton University, Southampton, England (1976)
10. Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., Rochester, L.: Accuracy of the microsoft kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture **39**(4), 1062–1068 (2014)
11. Corti, A., Giancola, S., Mainetti, G., Sala, R.: A metrological characterization of the kinect v2 time-of-flight camera. Robot. Auton. Syst. **75**, 584–594 (2016)
12. Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., Moore, R.: Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)
13. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. Psychometrika **31**(1), 1–10 (1966)
14. Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. In: Robotics-DL tentative, International Society for Optics and Photonics, pp. 586–606 (1992)
15. Caimmi, M., Carda, S., Giovanzana, C., Maini, E.S., Sabatini, A.M., Smania, N., Molteni, F.: Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients. Neurorehabilitation Neural Repair **22**(1), 31–39 (2008)