Accuracy of the Microsoft Kinect System in the Identification of the Body Posture

Paolo Abbondanza, Silvio Giancola, Remo Sala, and Marco Tarabini(✉)

Department of Mechanical Engineering, Politecnico di Milano, Via La Masa 1, Milan, Italy
{paolo.abbondanza,silvio.giancola,remo.sala,
marco.tarabini}@polimi.it

Abstract. Markerless motion capture systems have been developed in an effort to evaluate human movements in a natural setting. However, the accuracy and reliability of these systems remain nowadays understudied. This paper describes a study performed to evaluate the accuracy and repeatability of the identification of posture using the Microsoft Kinect V2 markerless motion capture system. The measurement repeatability has been studied by observing a mannequin from different positions, with different light conditions, with obstacles partially hiding the lower limbs and with different clothes. The metrics for the evaluation of repeatability were the length of forearms, arms, thighs, legs and spine and the angle of the elbows and knees. Results showed the preferential positions of measuring in terms of distance and angular position between the sensor and the target. The presence of occluded or hidden limbs and close subject represent the most critical problems of body detection returning misleading results.

Keywords: Kinect V2 · Body posture · Uncertainty estimation

1 Introduction

Motion capture techniques are used in several applications, starting from digital animation for entertainment to biomechanics analysis for clinical, sport applications and rehabilitation. There are different human body tracking systems available on the market and their performances mainly depends on the adopted measurement principle. According to the literature review on the human motion tracking for rehabilitation [1], there are three main categories of tracking systems, i.e. the visual-based, non-visual based and the robot-aided methods.

Then visual-based methods use one or more cameras to identify the different body segments, using markers placed in known position or recognizing the body posture using proper algorithms. The non-visual based methods use inertial sensors (accelerometers [2] and gyroscopes), magnetometers or acoustic sensors [3, 4] to detect the relative position of the sensors with respect to fixed elements located in known positions. The robot-aided techniques basically identify the position of the limbs starting from the geometrical configuration of the exoskeleton that is assisting the patient movements.

The Kinect V2 is a Time of Flight camera manufactured by Microsoft; it is available in the market since 2014. The device was originally built for the Xbox console and

P. Perego et al. (Eds.): MobiHealth 2016, LNICST 192, pp. 289–296, 2017.
DOI: 10.1007/978-3-319-58877-3_37

allowed the video games to be controlled by voice recognition and body movements. The Kinect is based on a continuous wave time of flight camera, that uses an array of 3 lasers and a monochrome camera with a resolution of 512×424 pixels; the field of view is $70° \times 60°$ and the measurement ranges between 0.5 and 4.5 m. The device also integrates an RGB camera and an array of 4 microphones for the voice command interpretation. The SDK allows to visualize in real-time and to acquire color images, infrared images, depth images and audio streams. The most useful feature for our purposes is the identification, performed with a machine vision algorithm, of the position and orientation of the 25 joints of the human body (Fig. 1).

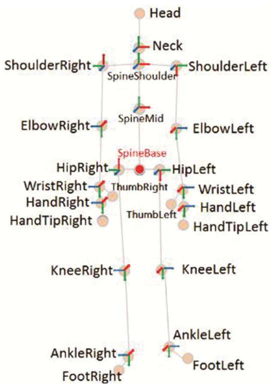

Fig. 1. Name and position of the 25 points detected by the Kinect V2.

The metrological performances of the first version of Kinect were studied in several works. The most comprehensive study of the posture was performed by Plantard, Auvinet et al. [5]. Asteryadis studied the motion of subjects using different Kinect sensors [6] while Lun and Zhao reviewed the possible application in the human motion recognition with the Kinect [7]. The Kinect was also used for the fall detection of older adults in houses [8].

The number of studies focused on the Kinect V2 is more limited [9]. The performances of the Kinect V2 were compared to those of the first version by Zennaro, Munaro et al. [10]. Authors studied the accuracy in the identification of common objects (a ball, a book, a bear puppy). Results showed that Kinect V2 is approximately two times more accurate in the near range; the gap increased at large distances. The new sensor was also more robust versus the artificial illumination and the sunlight. The metrological performances of the Kinect V2 in the reconstruction of geometrical features were analysed in [11]; results evidenced that the temperature of the Kinect V2 has an influence in the distance measurement, that uncertainty increases with the depth and the radial coordinate and that the measurement error usually depend on the material and other surface characteristics. The performance of the Kinect V2 in 3D reconstruction and people tracking also improved significantly with respect to the Kinect V1. Recently, performances of the open source software for multi-camera people tracking OpenPTrack has been studied by Munaro, Basso and Menegatti [12]. OpenPTrack is an open source project for people tracking and uses a network of heterogeneous 3D sensors all to track people in colour, infrared and disparity images. Results evidenced the supremacy of Microsoft Kinect V2 over the other sensors in people tracking and the benefits deriving from the calibration procedure.

In this paper we describe the accuracy of the Kinect V2 system in the identification of the human body posture. The Kinect V2 can be used in several fields where the Kinect V1 showed some limitations as, for instance, in the monitoring of workers' posture or for the identification of elders' falls in dwellings. Our work is divided in two parts: in the first we analyse the accuracy in the identification of the posture of a mannequin. The mannequin position was fixed and its position was measured by the Kinect V2 placed in 39 positions inside the room. For each Kinect position the posture was measured with and without clothes, in two sensors heights and in two different light conditions. The paper is structured as follows: Sect. 2 describes the experiments and the data processing techniques; experimental results are presented in Sect. 3. The discussion and the conclusions were grouped in Sect. 4.

2 Method

The Kinect V2 SDK was used to identify the position and the orientation of the 25 joints shown in Fig. 1 with a sampling rate of 30 Hz. Data are stored in an ASCII file and afterwards processed with Matlab.

The dummy used in our tests is shown in Fig. 2. The two elbows have angles of approximately 90 and 160° while the legs are straight and the feet are in contact with the ground. The dummy is located at approximately 0.5 m from a room wall.

Fig. 2. Positions from where the subject was observed.

Two series of tests were performed. In the first one the mannequin was fixed and the Kinect was moved to the positions shown in Fig. 2. The dummy is always observed from the left side (see Fig. 3): this consideration is driven by the necessity of assessing the performances of the Kinect when the subject is not observing the sensor and by the fact that the mannequin trunk is symmetrical with respect to the sagittal plane. The sensor was located on circumferences with radii of 2, 3 and 4 m each 15°. The mannequin posture was observed in different conditions:

- Naked/Dressed mannequin
- Lights on/off
- Horizontal sensor (sensor height 1.45 m)/Tilted sensor (Sensor height 1.80 m, 15° tilted downward)
- Mannequin partially hidden by a table.

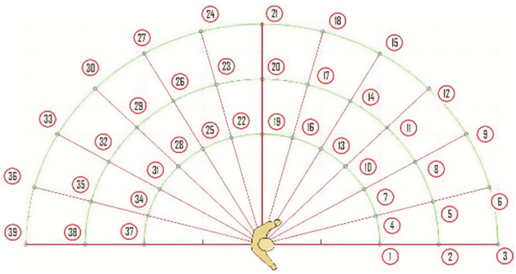

Fig. 3. Positions from where the subject was observed.

In static tests the dummy position was observed for around 30 s. Globally, more than 200 observations were performed. The metrics used to evaluate the measurement repeatability were:

- The length of specific body segments (arms, forearms, legs, trunk)
- The left and right elbow extension

The median value was chosen instead of the mean for the error estimation because of the presence of different outliers (points in which the mannequin position was not correctly recognized). The tracking algorithm used by the Kinect is time-dependent [13] and takes into account previous depth measurement. In static conditions, the algorithm have difficulties in the initialization and can return erroneous joint position (i.e. outliers).

In each configuration, the actual position of the limb (or the actual segment length) was estimated by the median of all the measures (900 frames for each position, 39 positions of the naked dummy in best light conditions, both horizontal and tilted Kinect sensor, no obstacles).

In the second series of tests the mannequin oscillated around its equilibrium position thanks to the force generated by an electrodynamic shaker. The motion measured by the Kinect was compared to the motion measured by a laser Doppler vibrometer (Polytec OFV 500 with displacement decoder). The frequency of the motion imposed by the shaker ranged between 1 and 10 Hz with steps of 0.5 Hz. The displacement varied with the frequencies and with the observed body part (head, torso, wrist). The dynamic tests last 10 s per frequency, in best light conditions.

3 Results

Experimental results of the static tests showed that, when the upper limb is not hidden by the body, the measurement error (the median of the measurements performed in a specific condition minus the reference value) is lower than 2 cm. Figure 4 show the measurement errors (in m) for the left (top figure) and right (bottom figure) forearms.

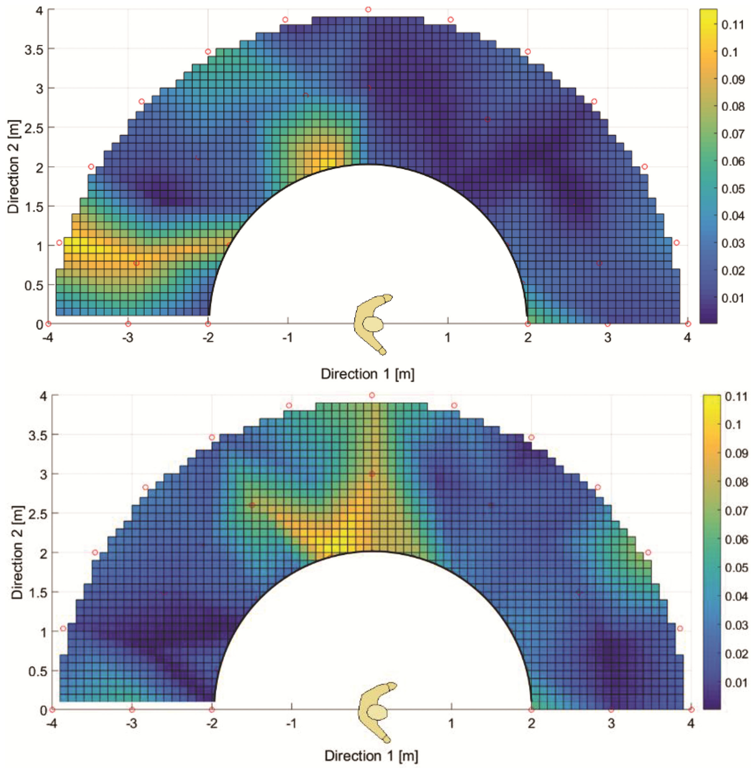

Fig. 4. Errors (median in m) for the estimation of the length of the left and right forearm

Results showed that the lightning conditions do not influence the skeleton measurement and that results obtained with the dressed dummy are more reliable than those obtained with the naked dummy. Larger errors occur from angles between 90° and 135° for both the right and left forearm. In these positions, the arms are not completely visible, and the machine vision algorithm fails in the identification of the upper body posture. The same analyses were performed on the arms, on the upper and lower part of the leg and on the trunk. Errors were always lower than 2 cm when the observed parts were clearly visible. The analysis on the elbow angle are shown in Fig. 5.

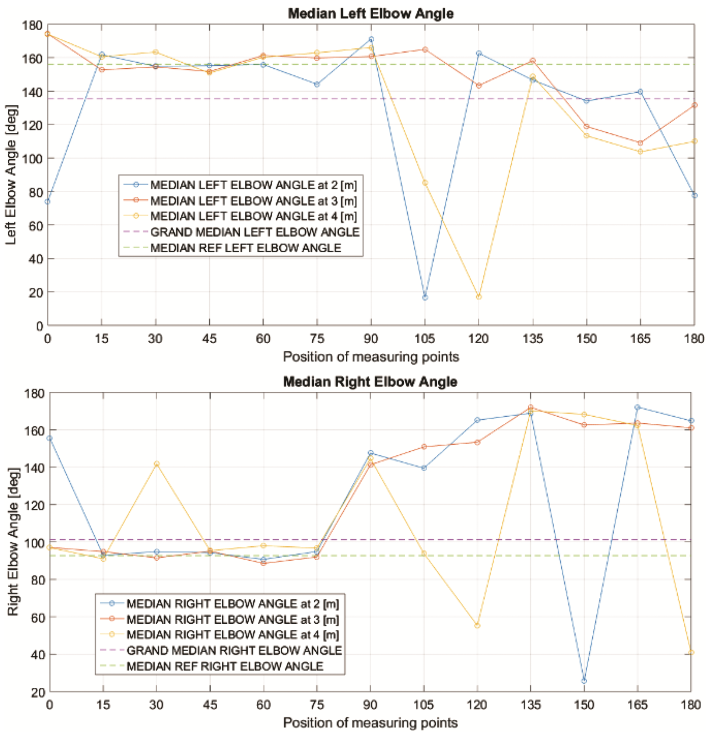

Fig. 5. Positions from where the subject was observed.

Errors are usually lower when the mannequin is observed from angles between 15° and 90°, where errors are lower than 10° except for three cases (frontal position, distance 2 m, left and right elbow and 30°, 3 m, right elbow angle) where the error is larger than 45°. In all these condition there are partial occlusions preventing from clearly observing the elbow.

The dynamic tests evidenced that, in general, the amplitude of the motion is estimated better when the motion occurs in planes orthogonal to the optical axis, although the behaviour is unpredictable and not related to the excitation frequencies. The next figures show three examples obtained with a lateral excitation; when the motion occurs along the Kinect optical axis, the amplitude of the motion is usually underestimated. Three examples are shown in Fig. 6; one can notice that at the frequency of 1 Hz the motion is underestimated both in lateral and fore-and-aft direction; conversely, at 2 Hz, the motion of the forehead is perfectly estimated in lateral direction. The limitations arose above 5 Hz, where a systematic underestimation occurred independently from the measurement direction.

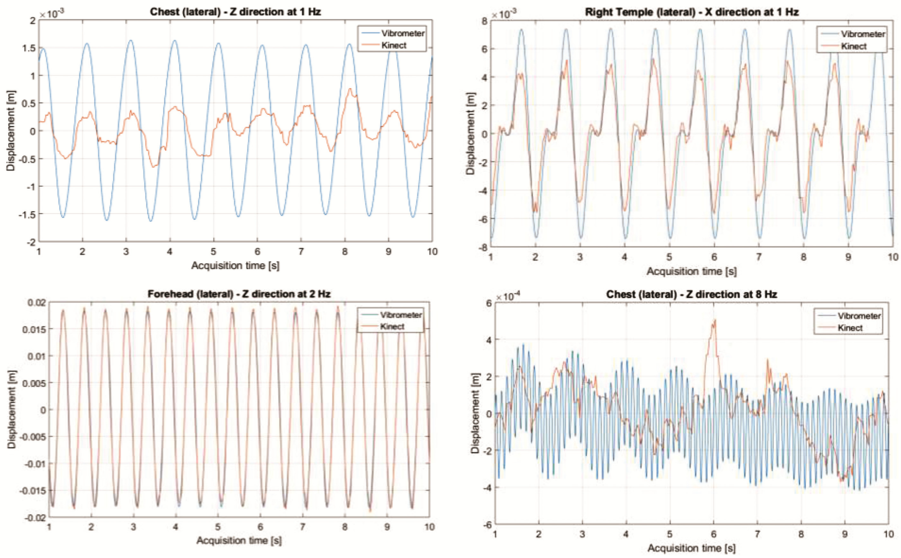

Fig. 6. Comparison between the signals measured by the Kinect (orange) and the ones measured by a vibrometer (blue). Lateral motion imposed by a shaker; Z direction is aligned with the Kinect axis, X direction is the direction connecting the mannequin shoulders. (Color figure online)

4 Discussion and Conclusions

In this work, we described the analyses performed to identify the accuracy and repeatability in the identification of posture of single subject in dwellings using the Microsoft Kinect V2. The first series of tests was performed observing a mannequin from different positions and in different experimental conditions. Results evidenced that, in general, the measurement error does not depend from the light condition nor from the presence of clothes. The main source of errors are the occlusions, that basically prevent from detecting the position of subjects in a room using a unique Kinect; in general, when one of the limbs is not perfectly visible, the posture is not correctly recognized and the error magnitude is often so large that the arithmetic average between measurements performed in different frames does not reduce the measurement uncertainty.

Dynamic measurements evidenced that below 5 Hz the main limit arises as the consequence of the resolution of the instrument; if the displacement is larger than 5 mm the motion reconstruction is usually correct. On the contrary, when the displacement is lower and/or the frequency is larger, the amplitude of the motion is not estimated correctly. This may represent a limitation in case of reconstruction of very fast movements of subjects, preventing from instance the adoption of the Kinect to track the motion of athletes in sports. Given the adoption of the machine learning algorithm to reconstruct the skeleton starting from the measured cloud of points, this limitation should be endorsed to the algorithm itself and not to the limited dynamic performances of the sensor.

References

1. Zhou, H., Hu, H.: Human motion tracking for rehabilitation—a survey. Biomed. Sig. Process. Control **3**, 1–18 (2008)
2. Tarabini, M., Saggin, B., Scaccabarozzi, D., Moschioni, G.: The potential of micro-electro-mechanical accelerometers in human vibration measurements. J. Sound Vibrat. **331**, 487 (2012)
3. Moschioni, G., Saggin, B., Tarabini, M.: 3-D sound intensity measurements: accuracy enhancements with virtual-instrument-based technology. IEEE Trans. Instrum. Meas. **57**, 1820–1829 (2008)
4. Moschioni, G., Saggin, B., Tarabini, M., Hald, J., Morkholt, J.: Uncertainty of array-based measurement of radiated and absorbed sound intensity. Appl. Acoust. **78**, 51–58 (2014)
5. Plantard, P., Auvinet, E., Pierres, A.L., Multon, F.: Pose estimation with a kinect for ergonomic studies: evaluation of the accuracy using a virtual mannequin. Sensors **15**, 1785–1803 (2015)
6. Asteriadis, S., Chatzitofis, A., Zarpalas, D., Alexiadis, D.S., Daras, P.: Estimating human motion from multiple kinect sensors. In: Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications, p. 3 (2013)
7. Lun, R., Zhao, W.: A survey of applications and human motion recognition with microsoft kinect. Int. J. Pat. Recognit. Artif. Intell. **29**, 1555008 (2015)
8. Stone, E.E., Skubic, M.: Fall detection in homes of older adults using the microsoft kinect. IEEE J. Biomed. Health Inform. **19**, 290–301 (2015)
9. Giancola, S., Giberti, H., Sala, R., Tarabini, M., Cheli, F., Garozzo, M.: A non-contact optical technique for vehicle tracking along bounded trajectories. J. Phys. Conf. Ser. **658**, 1–13 (2015)
10. Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S., Menegatti, E.: Performance evaluation of the 1st and 2nd generation kinect for multimedia applications. In: 2015 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2015)
11. Corti, A., Giancola, S., Mainetti, G., Sala, R.: A metrological characterization of the kinect V2 time-of-flight camera. Robot. Auton. Syst. **75(**Part B), 584–594 (2016)
12. Munaro, M., Basso, F., Menegatti, E.: OpenPTrack: open source multi-camera calibration and people tracking for RGB-D camera networks. Robot. Auton. Syst. **75**, 525–538 (2016)
13. 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**, 116–124 (2013)