

Multi-camera systems for rehabilitation therapies: a study of the precision of Microsoft Kinect sensors^{*#}

Miguel OLIVER¹, Francisco MONTERO², José Pascual MOLINA²,

Pascual GONZÁLEZ², Antonio FERNÁNDEZ-CABALLERO^{†‡2}

(¹*Instituto de Investigación en Informática de Albacete, Universidad de Castilla-La Mancha, Albacete 02071, Spain*)

(²*Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete 02071, Spain*)

[†]E-mail: Antonio.Fdez@uclm.es

Received Oct. 21, 2015; Revision accepted Mar. 9, 2016; Crosschecked Mar. 17, 2016

Abstract: This paper seeks to determine how the overlap of several infrared beams affects the tracked position of the user, depending on the angle of incidence of light, distance to the target, distance between sensors, and the number of capture devices used. We also try to show that under ideal conditions using several Kinect sensors increases the precision of the data collected. The results obtained can be used in the design of telerehabilitation environments in which several RGB-D cameras are needed to improve precision or increase the tracking range. A numerical analysis of the results is included and comparisons are made with the results of other studies. Finally, we describe a system that implements intelligent methods for the rehabilitation of patients based on the results of the tests carried out.

Key words: Kinect sensor, Rehabilitation system, Capture precision, Multi-camera system

<http://dx.doi.org/10.1631/FITEE.1500347>

CLC number: TP391

1 Introduction

Assistive technologies have become an important research field in recent years. These systems aim to bring technology to the disabled and use specialized hardware designed specifically to solve an interaction problem. By using this new hardware, disabled people can actually interact with the system. One type of assistive technologies is associated with the computer-assisted rehabilitation domain, which


designs physical rehabilitation processes to help patients through the use of computers and specialized hardware.

The earliest of these proposals were designed with accelerometers, gyroscopes, and other sensors attached to the patient. These devices capture movements and send data to a computer for further processing. However, this approach involves the disadvantages that the sensors have to be put in place by qualified personnel and the resulting interaction is not entirely transparent to the user. The latest advances in cameras and computer vision algorithms have now replaced the sensors attached to the wearer by tracking cameras, which achieve a more transparent interaction with the system. In this field, the Kinect sensor, developed by Microsoft, stands out from the rest. This device allows the general public to access three-dimensional capturing technology at

[‡] Corresponding author

^{*} Project partially supported by Spanish Ministerio de Economía y Competitividad/FEDER (Nos. TIN2012-34003 and TIN2013-47074-C2-1-R) and FPU Scholarship (FPU13/03141) from the Spanish Government

[#] A preliminary version was presented at the 13th International Conference on Practical Applications of Agents and Multi-Agent Systems, June 3-4, 2015, Spain

 ORCID: Antonio FERNÁNDEZ-CABALLERO, <http://orcid.org/0000-0002-8211-0398>

© Zhejiang University and Springer-Verlag Berlin Heidelberg 2016

an affordable price. This sensor has been included in a multitude of proposals for the rehabilitation of patients and is currently one of the most widely used sensors in assistive technologies.

One of the rehabilitation systems that uses the Kinect sensor was developed by Chang *et al.* (2011). This system demonstrates its usefulness in a study of the rehabilitation of two young adults. Another example of a system that uses the Kinect device to capture the patient's movements during rehabilitation exercises was described in Freitas *et al.* (2012). This system also includes an entertainment component, designed to reduce the therapy's abandonment rate. The application that we developed and described in Oliver *et al.* (2014b) is another example of these systems that uses Microsoft's device, aiming at the elderly. There are also commercial rehabilitation systems based on this affordable sensor, such as KineLabs (<https://www.polyu.edu.hk/bme/kinelabs/>), Reflexion (<http://www.westhealth.org/resources/about-reflexion-the-rehab-measurement-tool/>), Toyra (<http://www.toyra.org/>), TeKi (<http://www.ilitia.com/>), and VirtualRehab (<http://www.virtualrehab.info/es/>).

The precision and accuracy of pattern recognition sensors have been tested in a number of studies (Khoshelham and Elberink, 2012; Gonzalez-Jorge *et al.*, 2013; Bonnechère *et al.*, 2014). These studies concluded that the use of a sensor like Kinect is suitable in many tasks due to its precision and accuracy. However, Regazzoni *et al.* (2014) analyzed the results of using two sensors and suggested that more than one sensor may degrade the recognition quality and that the data provided may thus not be useful for many applications. In these studies only one or two cameras were used and the configurations of the systems were not changed in the tests to account for different angles of incidence of infrared light, neither was the distance varied between the sensor and target, so the studies are not useful if the aim is to deploy several capturing sensors that share the same monitoring area.

In this paper we focus on the problems derived from infrared saturation when using more than one Kinect sensor in the same workspace. We also analyze the effect of varying the number of devices, angles of incidence of light, and the distance between sensors and the user.

2 Related work

Computer-based assistance and rehabilitation systems have now become popular. These systems typically use a depth camera that captures the user's figure and sends it to a computer. Two different types of camera use infrared imaging to detect objects or persons. One is based on pattern recognition solutions and the depth of the image is calculated by the deformation of an infrared pattern projected by a laser attached to the sensor. The Kinect v1 sensor operates in this way (designed initially for the Xbox 360 and then for PCs). The second type is based on time of flight (ToF), where the depth of the image is obtained by measuring the time in which a modulated infrared light takes to return to the sensor. The Kinect v2 operates in this way (designed for the Xbox One). In this paper we focus on the first type of sensor (pattern recognition cameras) and specifically on the Kinect v1 sensor.

The measurement errors of a depth image are called holes and can be caused by multiple factors. The first factor to consider is the occlusions caused by objects in the environment. This type of noise is produced by an object coming in front of the one we want to capture. These errors can be solved by using multiple cameras that capture the same scene from different angles, so that a part of the scene not captured by one camera can be captured by another. Another cause of holes is due to surfaces that do not reflect infrared light, as for instance glass or black surfaces. These errors are solved by using materials that correctly reflect the infrared beam that reaches them.

The disparity between the infrared emitter and the infrared receiver can also cause holes in the depth image. Although the sender and receiver are typically placed as close as possible to each other, there is always a disparity between what they see, which means that part of the space captured by the receiver is not illuminated by the infrared emission, so that the sensor cannot determine depth in that area. Infrared light saturation is another source of error which affects depth images. In environments in which there is an alternative source of infrared light, errors may be due to the infrared light emitted by the camera being lost. This can be avoided by using environments in which the infrared light emitted is

controlled, or ToF sensors can be used, since the ToF sensors use modulated infrared light on a frequency that the sensor is designed to capture. Very fast moving objects in the workspace can also result in holes in the depth map. This is because some sensors have to merge several image shots to form the final image depth, so that if an object moves or is moved, errors occur in the fusion of the images. This can be solved by building faster sensors or can be avoided by using pattern recognition sensors, since these errors affect only cameras that require the use of multiple shots.

All these problems can lead to the conclusion that the precision and accuracy of the Kinect depth sensor are not very high, and that the data collected can be useful only for gaming. However, Khoshelham and Elberink (2012) found that it is possible to successfully use these sensors in other areas. They determined that at distances ranging from 1 to 3 m between object and sensor there is an acceptable error for mapping applications, but the error is too large at greater distances. Gonzalez-Jorge *et al.* (2013) analyzed the error of measuring a number of spheres and cubes from different angles, using the Kinect sensor and the Xtion Pro Live device. This error is always less than or equal to 6 mm for distances up to 1 m, and is less than or equal to 12 mm for distances up to 2 m. Bonnechère *et al.* (2014) found that the data obtained from this sensor is accurate enough to be used in ergonomics, biometric analysis, and even in military applications. Finally, Fernández-Baena *et al.* (2012) compared the precisions of the Kinect and Vicon systems (<http://www.vicon.com/>), and concluded that Vicon is more precise than Kinect, but that Kinect is precise enough for developing rehabilitation exercises.

Despite the above, Essmaeel *et al.* (2012; 2014) aimed to improve precision and accuracy with the help of filters and algorithms. In these studies it was found that it is possible to achieve higher precision and accuracy with proper processing of the raw data from the sensors. There were also solutions that require the addition of new hardware to improve the precision of the collected data; e.g., Mkhitarian and Burschka (2013) used an additional RGB camera to capture the scene from another angle to help the Kinect sensor calculate the depth.

The concept of increasing precision by using

additional hardware, such as another depth sensor, can also be used to solve occlusion problems. This can also increase the maximum number of users and enlarge the space-recognition work area. Due to all these advantages, researchers have proposed multi-camera systems which capture users from different angles. In our case, we have already implemented a system with three Kinect cameras located in a room for treating brain-injured patients (Oliver *et al.*, 2014a; 2015b). Increasing the number of sensors pointing at a certain area can solve the occlusion problem. However, using multiple pattern recognition sensors in the same workspace can create the problem of infrared light saturation.

Some studies have used multiple pattern recognition Kinect cameras. Regazzoni *et al.* (2014) performed experiments with two Kinect sensors and six PlayStation Eye cameras, and concluded that the error of these systems is always less than 100 mm, so they can be used in applications in which high precision is not required. Haggag *et al.* (2013) analyzed the accuracy of measurements from Kinect and the Xtion Pro Live sensors when a pair of these sensors share the same recognition surface. Olesen *et al.* (2015) focused on data interference when using up to three Kinect sensors pointing at the same space. Mallick *et al.* (2014) classified noise into three categories: spatial noise, temporal noise, and interference noise, the latter produced by superimposing the beam of infrared light emitted by two different sensors. To minimize this type of noise, three techniques can be applied when using multiple depth sensors: space division multiplex (SDM), time division multiplex (TDM), and pattern division multiplex (PDM).

However, the above studies did not make a thorough study of interference, since they did not test different sensor configurations. When the aim is to implement a computer-assisted rehabilitation system, a number of factors must be taken into account. First, we must consider how many patients can fit into one room, since the more patients, the higher the number of sensors needed. Another factor is the part of the patient's body that needs to be captured; if high precision is needed, then a large number of sensors must be used, and large rooms will need more sensors. Finally, the layout of the room plays an important role, since the presence of columns or other objects influences the placement of the sensors. Configuring the

rehabilitation room includes considering the number of sensors, the direction they face, and the distance between the sensors, to avoid them interfering with each other.

For the tests we implement a mono-camera assistive technology system for computer-assisted physical rehabilitation (Oliver *et al.*, 2014b), in which a physiotherapist performed exercises that were saved for the patient to imitate. Although the system did manage to achieve its objectives, we found that more sensors were needed to improve the results. In some of the exercises more than one camera was required to increase precision and sometimes the space to be monitored could not be controlled by a single camera.

As a result, we have conducted a series of experiments (Oliver *et al.*, 2015a), which attempt to determine the effect of multiple sensors in user's recognition. This paper is the application of such experiments.

3 Experimental setup

As mentioned in the previous section, different factors come into play when designing a computer-assisted rehabilitation room in the real world. These factors change from one implementation to another, making it impossible to define a one-fits-all deployment. The purpose of these experiments is to determine the optimal setup for each case, conditioned by real world factors, and allowing the rehabilitation of several patients at the same time. The factors to consider are: the number of sensors, the distance between the sensors and the patient, the distance between the sensors themselves, and the angle of incidence of infrared light.

The experiments consist of measuring the position of the user in the workspace and determining how the number and position of the capture devices affect the precision of the data achieved. The depth sensors emit a pattern of infrared light that bounces off nearby objects and calculate the distance to them by identifying the deformation of the infrared pattern. The devices have a field of view of 57.5° in the horizontal direction and 43.5° in the vertical, with a maximum viewing distance of about 4 m and a minimum distance of 80 cm from the sensors.

In this study we employed a tailor's dummy

to avoid the problem of involuntary human movements. The dummy was placed on previously established marks on the floor and the distance from the sensor to the dummy's hip was measured. The data collected by different Kinect sensors by Microsoft's skeleton tracking algorithm was transferred to a global coordinate system. Fig. 1 details the grid of the space used in the experiments. In each experiment three positions were used. For each of these positions 4 measuring points were used for the first, 9 for the second, and 16 for the third, and each of the points was 1 m away from the next. The position of each sensor was set to test how the relative position of the sensors and distance between the sensors affect interference noise.

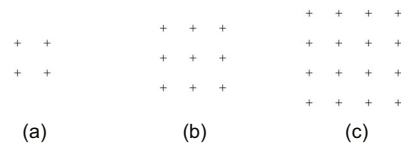


Fig. 1 Measuring points at which the dummy was positioned: (a) Grid 1 with 4 measuring points; (b) Grid 2 with 9 measuring points; (c) Grid 3 with 16 measuring points

Fig. 2 shows how the dummy was positioned in Grid 1. First, it was placed over the first mark and data was collected. The dummy was then moved to each mark successively and data was collected at each one.



Fig. 2 Movement of the dummy in Grid 1

Precision was obtained by calculating the number of erroneous pixels from the sensors' depth map and the standard deviation (SD) of the position of the user. The number of erroneous pixels refers to the positions of the depth image in which the distance cannot be obtained by the sensor. At each measuring point, 100 samples of erroneous pixels were collected plus 100 samples of the dummy's position. The tests were performed in a space free of infrared light to increase accuracy and avoid interference from other sources. A photograph of the setup is shown in Fig. 3.

These tests were done with different numbers of sensors, different sensor positions, different distances to the target, and different distances between sensors. The test parameters and the corresponding values are as follows:

1. The number of sensors used could be 1, 2, or 3, which is enough in most cases. In the rehabilitation room there may be more sensors but they did not share the same interaction space.
2. The sensors were positioned at the front, side, and back of the dummy to simulate front, side, and back images of the patient, respectively.
3. The distance between sensor and patient was from 1 to 4 m in steps of 1 m because the operating distance of the sensor was from 80 cm to 4 m. The intermediate locations could be inferred from the basic cases.
4. The distance between sensors varied in each test to suit the distance and angle between user and sensor.

The Z (vertical) component provided by the sensors was not considered and thus all the sensors were at the same height as the dummy's hip and parallel to the floor. In Fig. 4, the camera field of view and the directions of the X and Y axes are shown.

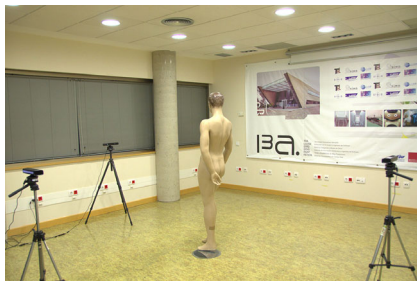


Fig. 3 Photograph of the experiment performed with three sensors. Sensors 1 and 2 were positioned perpendicular to each other, sensors 2 and 3 were also perpendicular to each other, and sensors 1 and 3 were facing each other

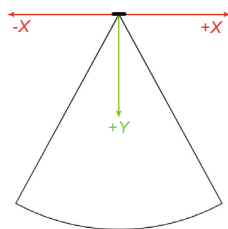


Fig. 4 X and Y axes of the sensors. The X -axis is parallel to the sensor and the Y -axis is perpendicular to it

3.1 Experiments

The following information was given for each experiment: the formulas used to unify the position of the user, an image showing the arrangement of sensors and user, and an explanation of the expected data.

3.1.1 Experiment 1

The first experiment was with a single Kinect sensor pointing at the front of the workspace (Fig. 5). As in all the experiments, the dummy was facing the sensor. The aim was to set up a baseline for comparison with the experiments with two Kinect sensors. A low number (but not zero) of erroneous measurements was expected. Even if there was no interference between the sensor beams, there were always erroneous measurements due to improper infrared reflection from surfaces. The difference of the view from the camera sensor and infrared emitter also involves errors. The dummy must be correctly placed with relatively small SD.

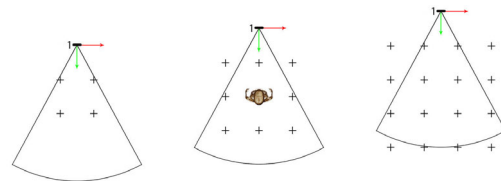


Fig. 5 First experiment, with a single sensor. The three different grids with their measurement points are shown. Grid 2 shows the user's orientation. The dummy is parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left. Reprinted from Oliver et al. (2015a), Copyright 2015, with permission from Springer

This arrangement is typically used in rehabilitation systems. The patient is looking directly at the only sensor that captures him/her. The user's position is determined by the following formulas, where x_1 and y_1 are the coordinates of the user, with respect to the sensor's reference axes:

$$\begin{cases} x_{\text{user}} = x_1, \\ y_{\text{user}} = y_1. \end{cases}$$

3.1.2 Experiment 2

The second experiment was with two sensors pointing in parallel directions (Fig. 6), to determine

the performance of two overlapping beams of infrared light with the same recognition surface and pointing in the same direction. Fairly high interference was expected, since the sensors shared the same reflection area on the dummy. The internal Kinect pattern recognition algorithm was assumed to fail when identifying patterns. The number of erroneous pixels and the SD were expected to be higher than in Experiment 1.

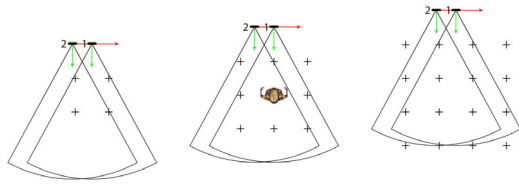


Fig. 6 Second experiment, with two sensors pointing in the same direction. The three different grids with their measurement points are shown. Grid 2 shows the user's orientation. The dummy is placed parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left of each sensor. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

In this type of deployment, we added an additional sensor to better capture the front of the subject. The main interest in this configuration was on rehabilitation exercises in which the data obtained from the front of the user was of vital importance. This required higher precision than that provided by a single sensor.

The transformation of the relative positions of each sensor to global positions is made by applying the following formulas, where x_1 and y_1 are the user's coordinates with respect to the reference axes of the first sensor, and x_2 and y_2 are the user's coordinates relative to the coordinate axes of the second sensor. Finally, t_x is the separation in the X-axis between the sensors, with respect to the coordinate axes of the first sensor:

$$\text{Translation: } \begin{bmatrix} x_{1 \leftarrow 2} \\ y_{1 \leftarrow 2} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix},$$

$$x_{\text{user}} = \frac{x_1 + x_{1 \leftarrow 2}}{2}, \quad y_{\text{user}} = \frac{y_1 + y_{1 \leftarrow 2}}{2}.$$

3.1.3 Experiment 3

The third experiment was also with two Kinect sensors located orthogonally to each other (yaw angle difference equals 90°), as described in Fig. 7. The

aim was to determine the effect of superimposing two infrared beams perpendicularly on the same object. Less interference was expected than in Experiment 2. Although two sensors were used in both cases, the surface reflecting the infrared pattern was almost entirely separate, so that most of the patterns emitted by one sensor were not captured by the other.

This type of deployment is useful to focus on the side of the patient. With just one camera occlusions caused by the patient appear. However, if we add a second camera that captures the patient's side the occlusions disappear. With this configuration, more precise data can be obtained on rehabilitation exercises that focus on the limbs, since the movements not captured by one sensor will be captured by the other.

In this case, the X and Y coordinates of the second sensor are the opposite of the first sensor. This is represented in the following formulas, where x_1 and y_1 are the user's coordinates with respect to the reference axes of the first sensor, and x_2 and y_2 are his/her coordinates relative to the reference axes of the second sensor. Finally, t_x is the separation in the X-axis between the sensors, with respect to the coordinate axes of the first sensor, and t_y is the separation in the Y-axis between the sensors with respect to the coordinate axes of the first sensor:

$$\text{Rotation: } \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix},$$

$$\text{Translation: } \begin{bmatrix} x_{1 \leftarrow 2} \\ y_{1 \leftarrow 2} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix},$$

$$x_{\text{user}} = \frac{x_1 + x_{1 \leftarrow 2}}{2}, \quad y_{\text{user}} = \frac{y_1 + y_{1 \leftarrow 2}}{2}.$$

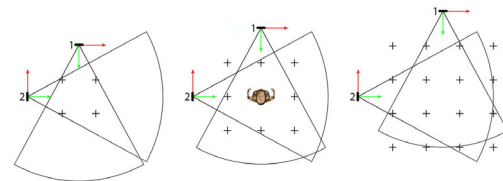


Fig. 7 Third experiment, with two perpendicular sensors. The three different grids are shown with their measurement points. Grid 2 shows the user's orientation. The dummy is parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left of each sensor. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

3.1.4 Experiment 4

This experiment was also with two Kinect sensors, but facing each other (yaw angle difference equals 180°), as shown in Fig. 8. The aim was to determine how the user's position is affected when an infrared source receives infrared light directly from the other sensor. High interference was expected.

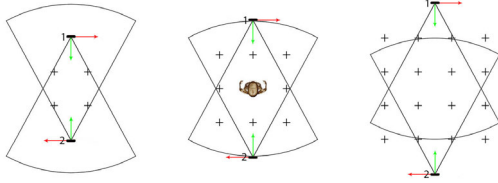


Fig. 8 Fourth experiment with two sensors. The three different grids are shown with their measurement points. Grid 2 shows the user's orientation. The dummy is placed parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left of each sensor. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

This deployment obtains data from the user's back. Capturing data from this position is a problem in existing rehabilitation systems, because the user hampers the sensor's line of sight. If the sensor is placed behind the patient, this problem disappears. The user's position is calculated by the following formulas, where x_1 and y_1 are the positions of the user with respect to the coordinate axes of the first sensor, x_2 and y_2 are the positions of the user with respect to the reference axes of the second sensor, and t_y is the separation in the Y-axis between the sensors:

$$\begin{aligned} \text{Rotation: } \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix} &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}, \\ \text{Translation: } \begin{bmatrix} x_{1 \leftarrow 2} \\ y_{1 \leftarrow 2} \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix}, \\ x_{\text{user}} &= \frac{x_1 + x_{1 \leftarrow 2}}{2}, y_{\text{user}} = \frac{y_1 + y_{1 \leftarrow 2}}{2}. \end{aligned}$$

3.1.5 Experiment 5

Experiment 5 was with three Kinect sensors pointing in the same direction (Fig. 9) to determine, in conjunction with Experiment 2, how parallel overlapping infrared beams affect the user's position. High interference was expected than in

Experiment 2. The aim was to obtain more precise data on the patient's front, in the same way as in Experiment 2. However, in this case more accurate results were expected, since there were three sensors involved.

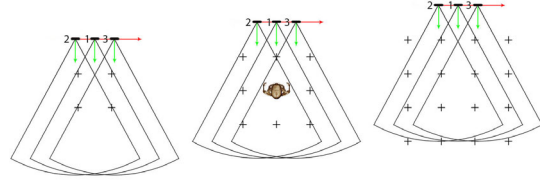


Fig. 9 Fifth experiment, with three sensors focused in the same direction. The three different grids are shown with their measurement points. Grid 2 shows the user's orientation. The dummy is parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left of each sensor. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

The position of the user is obtained from the following formulas, where x_1 and y_1 are the positions of the user, with respect to the coordinate axes of the first sensor, x_2 and y_2 the positions of the user with respect to the reference axes of the second sensor, and x_3 and y_3 the positions of the user with respect to the coordinate axes of the third sensor. t_{x12} is the separation in the X-axis between the first and second sensors and t_{x13} the separation in the X-axis between the first and third sensors:

$$\begin{aligned} \text{Translation: } \begin{bmatrix} x_{1 \leftarrow 2} \\ y_{1 \leftarrow 2} \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & t_{x12} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}, \\ \text{Translation: } \begin{bmatrix} x_{1 \leftarrow 3} \\ y_{1 \leftarrow 3} \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & t_{x13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ y_3 \\ 1 \end{bmatrix}, \\ x_{\text{user}} &= \frac{x_1 + x_{1 \leftarrow 2} + x_{1 \leftarrow 3}}{3}, \\ y_{\text{user}} &= \frac{y_1 + y_{1 \leftarrow 2} + y_{1 \leftarrow 3}}{3}. \end{aligned}$$

3.1.6 Experiment 6

This experiment was the combination of Experiments 3 and 4 (Fig. 10). We therefore expected the results to combine those of Experiments 3 and 4. One sensor captured the patient frontally for a global view. Another captured the patient perpendicularly, which was of particular interest for arm and leg exercises, in which occlusions occur. The third

sensor captured the back of the patient, for additional information.

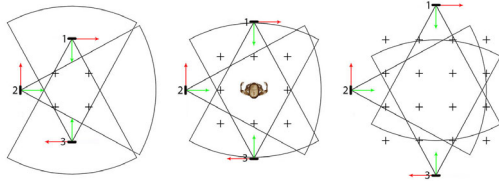


Fig. 10 Sixth experiment, with three sensors with 90° of separation. The three different grids are shown with their measurement points. Grid 2 shows the user's orientation. The dummy is located parallel to the normal vector of the first sensor in each grid. The sensor identifier for this test is shown on the left of each sensor. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

The user's position is calculated by the following equations, where x_1 and y_1 are the positions of the user with respect to the coordinate axes of the first sensor, x_2 and y_2 the positions of the user with respect to the reference axes of the second sensor, and x_3 and y_3 the positions of the user with respect to the coordinate axes of the third sensor. t_{x12} is the separation in the X-axis between the first and second sensors, t_{y12} the separation in the Y-axis between the first and second sensors, and t_{y13} the separation in the Y-axis between the first and third sensors:

$$\begin{aligned} \text{Rotation: } \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix} &= \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}, \\ \text{Translation: } \begin{bmatrix} x_{1 \leftarrow 2} \\ y_{1 \leftarrow 2} \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & t_{x12} \\ 0 & 1 & t_{y12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_2 \\ y'_2 \\ 1 \end{bmatrix}, \\ \text{Rotation: } \begin{bmatrix} x'_3 \\ y'_3 \\ 1 \end{bmatrix} &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ y_3 \\ 1 \end{bmatrix}, \\ \text{Translation: } \begin{bmatrix} x_{1 \leftarrow 3} \\ y_{1 \leftarrow 3} \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & t_{y13} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_3 \\ y'_3 \\ 1 \end{bmatrix}, \\ x_{\text{user}} &= \frac{x_1 + x_{1 \leftarrow 2} + x_{1 \leftarrow 3}}{3}, \\ y_{\text{user}} &= \frac{y_1 + y_{1 \leftarrow 2} + y_{1 \leftarrow 3}}{3}. \end{aligned}$$

4 Data and results

This section contains an explanation of how the data was obtained and how each configuration could

be applied in real situations.

The following information was given for each of the measuring points:

1. The SD of the user's position in the 100 samples collected.
2. The average number of erroneous pixels in the 100 samples collected. An erroneous pixel is a pixel in the captured image for which the sensor was not able to determine the distance at which it was located.
3. The heat map of the SD of the user's position in the 100 samples collected. The SD ranges from 0 to 10 cm, which is shown by different colors: blue means a small error, green small-medium, yellow medium-large, and red a large error.

4.1 Results in Experiments 1, 2, and 5

This section analyzes the combined information from Experiments 1, 2, and 5, as all these experiments tracked the dummy in parallel directions. Starting from the basic single sensor case, one or two more were added to extend the monitored area.

In all the three experiments there are measurement points that have not been captured by any sensor (Figs. 5, 6, and 9). Also, measuring points can be found over 4 m away from the sensor, as in Grid 3, which means that the user is recognized, but the position error is very large due to the sensor's limits. In Figs. 11–13 the SDs in precision of the dummy's position and the numbers of erroneous pixels captured at the scene are shown. The SD heat maps in precision of the dummy's position are shown in Figs. 14–16.

In Experiment 1, there is a deterioration proportional to the distance between the user and the sensor. Almost all the SDs are less than 1 mm for distances between 1 and 2 m. For distances of 3 m, SD ranges between 1 and 5 mm, and at a distance of 4 m the SD reaches 10 mm. This indicates that the Kinect provides adequate results for most applications at short distances. In Experiment 2, the user's position is similar to that obtained in Experiment 1 and stays within the previous value range. At some points the results are better than in Experiment 1, but this does not necessarily imply better tracking. On the other hand, at a distance of 4 m, the measures become significantly worse; the values obtained are between 34.4 and 55.1 mm (in

Grid 1				Grid 1			
1.0	0.2			58	43		
0.3	0.6			51	53		
Grid 2				Grid 2			
-	0.3	-		-	53	-	
0.2	0.3	1.3		51	49	49	
1.2	0.8	1.8		49	48	50	
Grid 3				Grid 3			
-	0.2	0.5	-	-	63	46	-
-	0.3	0.3	-	-	54	55	-
2.2	1.3	1.2	3.6	53	53	53	53
11.4	47.1	13.1	32.2	52	53	53	53

(a) (b)

Fig. 11 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 1. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

Grid 1				Grid 1			
0.7	0.3			52	46		
0.7	0.7			49	46		
Grid 2				Grid 2			
0.3	0.6	-		47	46	-	
1.0	0.5	2.2		49	49	52	
1.5	2.1	2.2		46	47	46	
Grid 3				Grid 3			
-	0.9	0.3	-	-	64	42	-
0.5	0.6	0.7	-	41	51	47	-
2.1	1.4	2.2	2.5	49	49	49	55
34.4	37.9	55.1	46.9	48	49	49	48

(a) (b)

Fig. 12 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 2. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

Experiment 1 the values are between 11.4 and 47.1 mm). In Experiment 5 the position data is worse than in Experiment 2 with two sensors. However, this data follows the same pattern as in Experiments 1 and 2. At distances between 1 and 2 m, the SD is less than 1 mm at most of the measurement points. For distances of 3 m, the SD falls between 1 and 6.1 mm, and at a distance of 4 m the SD is similar to that of Experiment 2.

In Experiment 1, there are between 43 000 and 63 000 erroneous pixels. The data shows that

Grid 1				Grid 1			
0.6	0.7			56	63		
0.8	0.8			65	63		
Grid 2				Grid 2			
0.4	0.9	0.5		54	59	62	
0.5	1.2	2.8		61	66	71	
1.3	2.2	3.8		66	67	66	
Grid 3				Grid 3			
-	0.9	1.8	-	-	69	64	-
0.6	2.7	0.6	0.9	51	68	65	78
1.9	2.2	2.6	6.1	62	67	67	76
39.6	33.1	53.1	54.8	67	69	68	68

(a) (b)

Fig. 13 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 5. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

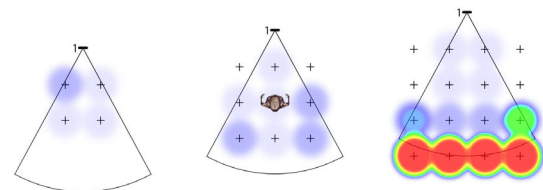


Fig. 14 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 1, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

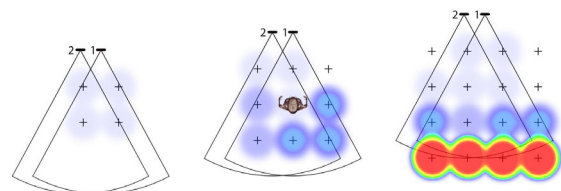


Fig. 15 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 2, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

errors are usually higher at 1 m than at greater distances. This is mainly due to one of the factors mentioned above, the disparity between the infrared transmitter and the infrared receiver (binocular disparity). As already noted, the emitter projects an infrared beam on a surface and the receiver captures the scene slightly differently. There are therefore areas that are captured by the sensor but not reached

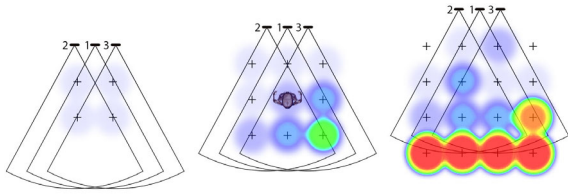


Fig. 16 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 5, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

by the infrared beam. These areas are represented as holes in the depth map, for which the distance to the transmitter cannot be determined. This is inversely proportional to distance, so that at 1 m there should be more erroneous pixels than at 2 m or more.

In Experiment 2, there are on average fewer erroneous pixels than in Experiment 1. The explanation for this is simple; the first sensor captures more erroneous pixels due to the interference from the second sensor. Also, the second sensor captures the scene better than the first. This does not necessarily mean that the data is more reliable than that captured by the first sensor, but that there are some areas of the scene that the first sensor does not capture properly and the second sensor does. This indicates that the number of erroneous pixels in the scene is not directly proportional to the error in the user's position since, in this experiment there are fewer erroneous pixels, but the SD of the user's position is higher. In Experiment 5, there are more erroneous pixels than in Experiments 1 and 2. This may suggest that the user recognition is worse, but, as in the other experiments, the fact that there are more erroneous pixels is not a clear indication that the tracking error is greater.

Figs. 17–19 show graphs which indicate the SDs in Experiments 1, 2, and 5 respectively, according to the distance between the sensors and the dummy. It can be seen that the correct positioning of the user decays with distance. The positioning error becomes exponential when the distance reaches 4 m. This means that when the user is closer to the sensor, the data obtained is more accurate. With the data, we can predict the limits where this happens and thus take steps to control it. In Grid 1 (Fig. 17), at a distance of 2 m, the error associated with Experiment 1 is lower than those with the other two experiments. In Grid 1 (Fig. 17), at a distance of 1 m, the error

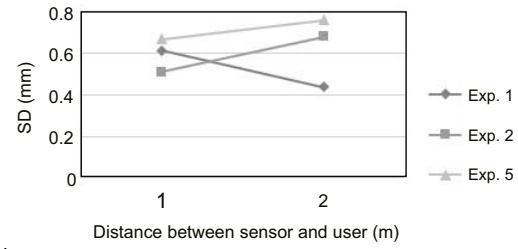


Fig. 17 Comparison of one (Experiment 1), two (Experiment 2), and three (Experiment 5) sensors focusing in the same direction (Grid 1). Reprinted from Oliver et al. (2015a), Copyright 2015, with permission from Springer

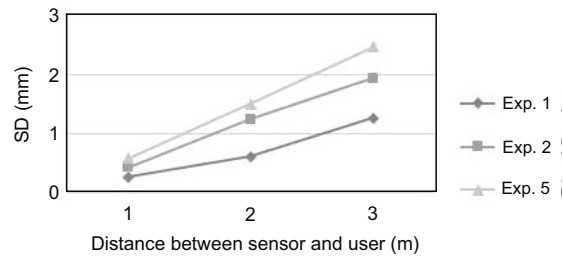


Fig. 18 Comparison of one (Experiment 1), two (Experiment 2), and three (Experiment 5) sensors focusing in the same direction (Grid 2). Reprinted from Oliver et al. (2015a), Copyright 2015, with permission from Springer

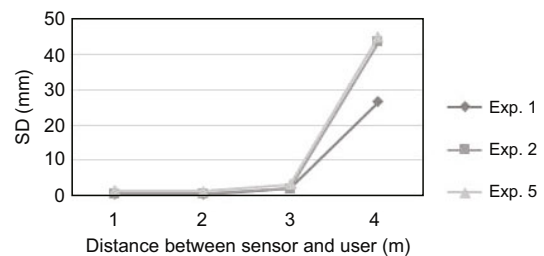


Fig. 19 Comparison of one (Experiment 1), two (Experiment 2), and three (Experiment 5) sensors focusing in the same direction (Grid 3). Reprinted from Oliver et al. (2015a), Copyright 2015, with permission from Springer

associated with Experiment 1 is larger than that with Experiment 2. This is due to a measurement error of the sensor, since the SD is very small (less than 1 mm).

From Experiments 1, 2, and 5, it can be seen that placing sensors in a row and pointing them in parallel directions is not a good idea. The

reason for this is that the workspace is not increased significantly. We cannot obtain another perspective of the user, and the collected data is no better than that from a single sensor. Nevertheless, the maximum number of users can be actually increased, as in Experiment 2, because one sensor handles a group of users and another can deal with another group.

4.2 Results in Experiment 3

Experiment 3 was with two sensors which covered the same area but perpendicularly. In addition to the sensor of Experiment 1, we added another sensor perpendicular to the workspace. This means that the first sensor captures the user frontally, and the second captures the user's side. As in the previous experiments, there are points that cannot be collected by any sensor and therefore these points are empty in Fig. 20. The heat map of the user's positioning results is presented in Fig. 21.

Grid 1		Grid 1	
0.4	1.4	42	31
3.1	0.4	39	35

Grid 2			Grid 2		
-	0.6	2.3	-	42	44
0.2	0.8	1.2	49	41	40
1.4	3.1	2.3	49	39	39

Grid 3				Grid 3			
-	0.2	2.6	47.2	-	64	37	27
5.5	0.5	3.2	20.4	38	43	43	28
1.4	2.6	1.0	5.2	45	43	42	41
10.2	62.5	20.1	20.5	52	53	41	40

(a)
(b)

Fig. 20 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 3. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

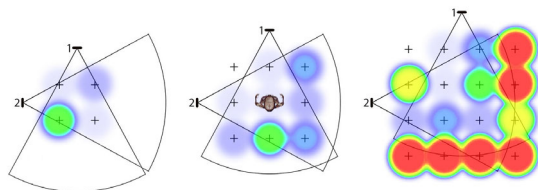


Fig. 21 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 3, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

The SD in this experiment is larger than that in the previous ones. This means that the user's positioning is worse. The data was not consistent with our expectations, since two perpendicular emitters were used and theoretically the beams should not interfere with each other. The explanation for this is simple; the light beams do not interfere with each other, but the second sensor does not capture the user correctly. That is, the sensor captures the user's side, and the occlusions produced by his/her own body lead to errors. Another factor is that while the distance between a sensor and the user can be small for one of the sensors, at the same time it can be larger for the other sensor. This means that the measures are more precise for the first sensor, but not so good for the second one, and vice versa.

Regarding the number of erroneous pixels, just as in Experiment 2, it is better than that collected in Experiment 1. This is due to the same reason as explained above; the second sensor captures the scene better than the first, and therefore the average number of erroneous pixels is lower. This proves again that the number of erroneous pixels is not directly proportional to the error in the user's position.

From the data obtained it seems that two sensors placed perpendicularly that capture the same area is a bad idea, but only at first. This is because we obtain worse results in the user's position than if we use a single sensor but, on the other hand, the recognition space is greater than with a single sensor. Also, the maximum number of users that the system is able to recognize is increased, and most importantly, a different part of the user's body is captured (his/her side). Thus, this arrangement may be useful in a few cases:

1. If we need to capture the side and the front of the user at the same time, this configuration can help solve this problem. This way, one of the sensors can focus on the front and the other on his/her side. In addition, post-processing would be useful in the reconstruction of the user's body, selecting the data of the sensor that best captures the user.

2. This arrangement can also be useful when we want to let the user move freely within the workspace, without the need to look at the sensor. Using two perpendicular sensors gives the user more freedom, since each sensor captures him/her from a different perspective. As in the previous case,

post-processing of the user's data could be useful.

4.3 Results in Experiment 4

In Experiment 4 we placed two sensors facing each other, with one sensor capturing the user frontally and the other his/her back. In this case, there are no points that the sensors cannot capture, and the entire workspace is covered. All the measuring points provide information (Fig. 22), and the heat map of the user's position results is given in Fig. 23.

Grid 1				Grid 1			
1.0	0.4			43	36		
1.3	0.8			38	45		
Grid 2				Grid 2			
2.7	1.0	6.6		43	48	28	
0.8	0.2	4.5		46	48	48	
1.1	1.0	1.7		49	53	50	
Grid 3				Grid 3			
7.5	3.6	2.4	17.4	65	64	55	63
3.2	1.0	0.6	17.2	65	59	60	63
2.3	1.5	1.1	2.5	55	60	61	53
23.1	29.2	8.3	31.4	52	51	60	53

Fig. 22 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 4. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

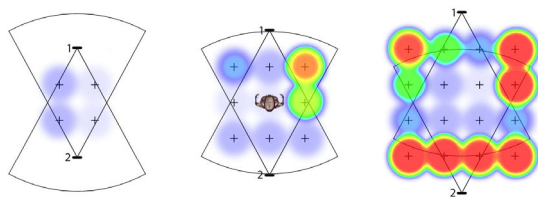


Fig. 23 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 4, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

In this experiment, we obtained worse results as regards the user's position than from Experiment 1. The data obtained, in the vast majority of positions, is worse than in Experiment 1. This deterioration is evident in the first row of Grid 3, because the results increase from less than 1 mm to over 17 mm. In this case there are two factors that cause this

deterioration. First, as in the previous experiment, there are positions where, although a sensor captures a user at a short distance, the other sensor captures him/her further away, and therefore the data decays greatly. Second, the sensor beams are pointing directly at each other, which also deteriorates the tracking information.

In the same way as in previous experiments, the data shows that the number of erroneous pixels does not indicate very clearly what is happening in the user's position. While in Grid 3 the number of erroneous pixels increases considerably, reaching an average of about 60 000, this may not be clearly related to interference between the two sensors. As before, using this deployment may not be advantageous after looking at the results, which are definitely worse than in Experiment 1. However, the recognition space is larger than with a single sensor. In addition, this arrangement can be considered if the user is allowed to move freely around the room, or if we need to capture his/her back.

4.4 Results in Experiment 6

Experiment 6 aimed at determining how a series of Kinect sensors, which captured the user from different perspectives, affect the user's position. In this case, the sensors capture the user from the front, back, and side simultaneously. The data generated is given in Fig. 24 and the heat map of the user's position results is given in Fig. 25.

Grid 1				Grid 1			
0.4	4.5			39	30		
3.5	1.2			35	36		
Grid 2				Grid 2			
2.8	0.9	7.7		42	42	37	
0.6	1.0	3.5		46	44	42	
1.3	1.2	3.4		49	45	39	
Grid 3				Grid 3			
6.7	6.3	3.5	15.8	66	65	46	47
3.3	0.8	2.6	8.9	50	52	51	47
1.1	0.8	1.6	20.7	46	52	51	42
21.4	24.4	18.7	22.5	53	51	50	41

Fig. 24 Standard deviation in precision of mannequin position (mm) (a) and the number of erroneous pixels (thousands of pixels) (b) in Experiment 6. Reprinted from Oliver *et al.* (2015a), Copyright 2015, with permission from Springer

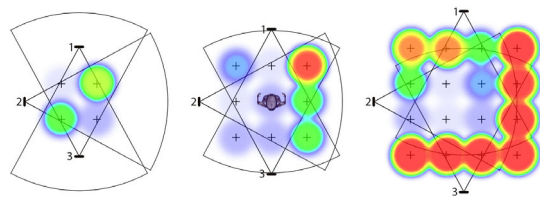


Fig. 25 Heat map of the standard deviation (SD) in precision of dummy's position in Experiment 6, considering an SD of 10 cm as the upper limit (references to color refer to the online version of this figure)

With regard to the tracked position, we can see that the data collected in this experiment is worse than in all the previous ones. In this experiment there are positions that are close to a sensor, and therefore with a small SD, but this position is far away from another sensor. The data shows that the central areas of Grid 2 and Grid 3 have a lower SD, due to its exponential growth. The number of erroneous pixels in this deployment is quite similar to other measurements taken previously. Once again, it is demonstrated that a high number of erroneous pixels may indicate a problem but a low number is not a measure of quality.

The use of this deployment can be useful if we need to capture the user's entire body. If this deployment is complemented by a fourth sensor that captures another side, we would obtain an interaction space with freedom of movement.

5 Discussion

In this section the results obtained in the experiments are compared with the results obtained by other researchers.

Because the probability distribution of the collected data follows a normal distribution, the data is distributed uniformly over the line of real numbers, and the error in 99.7% of cases will always be less than $\mu \pm 3\sigma$ (μ is the mean of the distribution and σ is its SD).

5.1 Comparison with the results in Bonnechère et al. (2014)

Bonnechère et al. (2014) focused on the SD of the data collected from a single user. This study used only one pattern recognition sensor. The user was positioned facing the sensor and the sensor collected data from his/her upper and lower extremities. The

user was placed at a distance of 1.5, 2.0, or 2.5 m from the sensor.

For this comparison we have chosen Experiment 1. The number of sensors (which is 1) and the angle between the user and the sensor are the same in both experiments. The measuring distances from both experiments are similar, so it should not be a problem in the comparison. The only remarkable difference is that while Bonnechère et al. (2014) measured the distance to the subject's extremities, we measured from the hip. The comparison data can be found in Table 1.

We detected that the SD collected by Bonnechère et al. (2014) largely exceeds those found in Experiment 1. This is due to the fact that the measured points differ in the two experiments. While they measured from the limbs, we measured from the hip. We think that the data of the user's extremities is always worse than the hip data, mainly due to two reasons: involuntary movements are always greater in this area, and the Kinect sensor reconstructs the user's body starting from the hip, which makes the accumulated error higher at the extremities.

5.2 Comparison with the results in Khoshelham and Elberink (2012)

In Khoshelham and Elberink (2012), SD was measured in a theoretical geometric plane by a Kinect sensor and the technical characteristics were studied. The number of sensors was only one, the angle between the sensor and the geometric plane was 0° , and the distances ranged between 1 and 5 m in steps of 1 m.

For this comparison we have chosen, again, Experiment 1 as the number of sensors, measuring distances, and angle between the user and the sensor are the same in both. The only parameter that varies is the measurement point. While we measured a tangible point in the capture space from the hip of a dummy, Khoshelham and Elberink (2012) measured a theoretical geometric plane. This comparison determines whether the experiment is consistent with the expected theoretical data. The comparison data can be found in Table 2.

The results obtained theoretically by Khoshelham and Elberink (2012) and the results obtained in Experiment 1 are broadly consistent. While there are small differences between the data obtained by

Table 1 Comparison with the data collected by Bonnechère et al. (2014)

Reference	Number of sensors	User-sensor angle (°)	Measuring point	Distance (m)	SD (mm)
Bonnechère et al. (2014)	1	0	Upper limb	1.5	36.0
	1	0	Upper limb	2.0	35.0
	1	0	Upper limb	2.5	40.0
	1	0	Lower limb	1.5	100.0
	1	0	Lower limb	2.0	80.0
	1	0	Lower limb	2.5	79.0
Current study	1	0	Hip	1.0	1.0
	1	0	Hip	2.0	1.3
	1	0	Hip	3.0	3.6

Table 2 Comparison with the data collected by Khoshelham and Elberink (2012)

Reference	Number of sensors	User-sensor angle (°)	Measuring point	Distance (m)	SD (mm)
Khoshelham and Elberink (2012)	1	0	Theoretical plane	1	2.5
	1	0	Theoretical plane	2	7.5
	1	0	Theoretical plane	3	15.0
	1	0	Theoretical plane	4	25.0
	1	0	Theoretical plane	5	40.0
Current study	1	0	Hip	1	1.0
	1	0	Hip	2	1.3
	1	0	Hip	3	3.6
	1	0	Hip	4	47.1

each, the data follows the same pattern. At a distance of 1 m, SD is less than 2.5 mm in both cases, and at 4 m SD reaches in both cases 25 mm or more. This exponential behavior of the SD with respect to the measuring distance is consistent in both experiments, which indicates that measuring a flat surface and measuring the user's body are close to the same values.

5.3 Comparison with the results in Gonzalez-Jorge et al. (2013)

Gonzalez-Jorge et al. (2013) focused on the error in the measurement instead of SD. However, as previously mentioned, our data is normally distributed and we can calculate the precise error.

Gonzalez-Jorge et al. (2013) used a Kinect sensor and an Xtion Pro Live sensor. Both sensors rely on the method of pattern recognition to provide the depth of the scene. In this case, the errors in the recognition of a series of cubes and spheres, at 1 or 2 m from the sensor, were studied. Various angles were tested (45°, 90°, and 135°) but they did not provide data on the error made at each angle, but

only the overall error was given.

For this comparison we have chosen, once again, Experiment 1. The number of sensors and the measuring distances are equal in both experiments. However, the angle of measurement and the measured objects differ. While in our case distance was measured from the dummy's hip, in Gonzalez-Jorge et al. (2013) spheres and cubes were used. The comparison data can be found in Table 3. The results of both experiments are quite similar and both are found in the same margins. However, in Gonzalez-Jorge et al. (2013) there is an evidently worsening of the distance between the sensor and the object measured.

5.4 Comparison with the results in Regazzoni et al. (2014)

As in the previous comparison, Regazzoni et al. (2014) focused on the error in measurement accuracy instead of SD. However, our data was normally distributed and we can calculate the error with precision. They presented the precision data collected by a multi-camera system. First, they used a series of RGB cameras to collect data from one person and

these were compared with the data collected by two Kinect sensors pointing at the same person, on which we will focus. These two sensors were at an angle of 54° and -54° with respect to the front of the user and at a distance of 3.5 m. All the points of the user were measured.

For this comparison we have chosen Experiments 2–4. The number of sensors used and the measuring distance are similar in all the tests. However, the measured points and the angles between the measured object and the sensor are different. The comparison data can be found in Table 4.

In this comparison, we can see how Regazzoni *et al.* (2014) showed a maximum error of 100 mm at a distance of 3.5 m. These results were verified by our experiments, with a similar deployment and results between 7.5 at 3 m and 187.5 mm at 4 m.

5.5 Our data

As we mentioned in the previous sections, our data is consistent in most cases with the results obtained in other experiments by other researchers. Also, the data is consistent with the theoretical results discussed in other studies. Besides, the experiments performed provide data of the interference of multiple sensors depending on the number of sensors used, the incidence angle of infrared light,

the distance to the target, and the distance between sensors.

As for the number of experiments carried out, we identified the interference of one, two, and three sensors, with three different angles between them and the dummy. Therefore, we performed a total of six experiments, each of them with three distance grids, making a total of 29 measurement points in each of the experiments, which leads to a total of 174 measurement points, of which 14 have not supplied data. However, for the 160 with data, we obtained the standard deviation of the precision in the user's positioning.

This will be of great help in the future when we need to carry out a real deployment in a rehabilitation room.

6 Developed system

The data collected and the conclusions concluded can be used to predict the effect of assigning each sensor a specific location in the rehabilitation room. We designed a new system that allows specialists to define the location of different RGB-D sensors to cover the interaction space in which the patient will carry out a specific therapy and determine the effect of this choice. This system allows a specialist to define the whole interaction space.

Table 3 Comparison with the data collected by Gonzalez-Jorge *et al.* (2013)

Reference	Number of sensors	User-sensor angle ($^\circ$)	Measuring point	Distance (m)	Precision error (mm)
Gonzalez-Jorge <i>et al.</i> (2013)	1	45, 90, and 135	Cubes and spheres	1	6.0
	1	45, 90, and 135	Cubes and spheres	2	12.0
Current study	1	0	Hip	1	3.0
	1	0	Hip	2	3.9

Table 4 Comparison with the data collected by Regazzoni *et al.* (2014)

Reference	Number of sensors	User-sensor angle ($^\circ$)	Measuring point	Distance (m)	Precision error (mm)
Regazzoni <i>et al.</i> (2014)	2	54 and -54	Full body	3.5	100.0
Current study	2	0 and 0	Hip	3.0	7.5
	2	0 and 0	Hip	4.0	165.3
	2	0 and 90	Hip	3.0	15.6
	2	0 and 90	Hip	4.0	187.5
	2	0 and 180	Hip	3.0	7.5
	2	0 and 180	Hip	4.0	94.2

Thus, one sensor can be placed in one location and then the system provides information on the effect of this choice in the quality of the measurements. To visualize the data about the precision of the measurement obtained, the information on theoretical accuracy was shown by means of a colored area in which blue means good accuracy and red indicates possible problems with the precision of the location. Due to the fact that not all the therapies need the same accuracy when measuring the coordinates of body parts, the specialist can choose the permissible error and define the range of values, assigning blue to small errors and red to large ones.

With this information, a specialist can decide to include more sensors and see the effect of combining these sensors on the precision of the measured area. In addition, the system offers the possibility of defining the sensor to be used. If the specialist does not define the sensor, the system can assign one, taking into account the data collected in the previous experiments. As the current sensors limit the number of users recognized, the system also displays the number of possible users for each area, allowing more sensors to be added to control the users' movements. Finally, as the system knows in which areas the error exceeds the established limits, the users can be informed when they are inside these areas and indicate a more precise area.

To sum up, the developed system supports the process of designing a rehabilitation space, offering therapists the possibility to design the rehabilitation rooms and locate all the sensors, taking into account the accuracy of the measures achieved by each sensor. It also controls the number of users to be recognized and indicates when a user is not in a suitable area to carry out a therapy, guiding him/her to a good location. Finally, it helps identify the most suitable sensor to monitor each patient in the rehabilitation process.

7 Conclusions

Although mono-camera rehabilitation systems work well enough in most situations, if we need to monitor large interaction areas or need to supervise more than one user, these systems have several limitations. As multi-camera rehabilitation systems can be the solution, we studied multi-camera devices and

how they interfered with each other.

To our knowledge, there are only a few publications focusing on the placement of sensors of this kind, and most are based on only one device. We therefore need to get more information on the effect of adding several sensors which monitor the same area. In this paper, we have studied how three sensors can interfere with each other, according to the distance between them, distance to the object to be measured, angle of incidence of the projected infrared light, and the number of sensors used.

The data obtained is consistent with Bonnechère *et al.* (2014), who concluded that the results obtained with the Kinect sensor can be used in many applications, due to the fact that sensor's data has a precision high enough to be used without additional processing. Our data is also consistent with the results obtained by Khoshelham and Elberink (2012), who found that the Kinect sensor provides sufficiently precise data at distances between 1 and 3 m. However, beyond this threshold the data becomes less precise and the error becomes too large.

In this regard, we also concluded that the data obtained by Gonzalez-Jorge *et al.* (2013) agrees with ours. Both experiments reveal an error that increases exponentially, relative to the distance between the measured object and the sensor. This also happens with other pattern recognition sensors, such as Xtion Pro Live. Our data is also in agreement with the results obtained by Regazzoni *et al.* (2014), who used two Kinect sensors to calculate the position error; however, they tested only a single grid, which means the cause of the error cannot be ascertained.

Finally, the following conclusions can be extracted from our experimental data:

1. At points measured 4 m away from the sensor the error can also be very large. At these distances the precision error must be considered. If the precision required is higher than that collected by the sensor, rehabilitation exercises should not be performed with it.
2. With a sensor whose distance from the user is between 1 and 2 m, almost all the SDs are less than 1 mm. This accuracy is sufficient for almost all rehabilitation exercises.
3. With a sensor whose distance from the user is 3 m, almost all the SDs collected are between 1 and 5 mm, which is sufficient for almost all rehabilitation

exercises.

4. With two or three sensors pointing in the same direction, the result is slightly worse. However, at a distance between 1 and 3 m the results are valid for almost any rehabilitation exercise.

5. A sensor that captures the patient's side may give unreliable results, due to occlusions arising from his/her own body. If a sensor is placed in this way, for instance, to monitor the patient's limbs, this limitation must be kept in mind.

6. A sensor that projects its beam into another may spoil the results of this sensor, due to the interference in the capture of infrared light. Therefore, if the focus is on the patient's back, and a set of opposing sensors is necessary, this interference should be taken into account.

7. When multiple sensors are used pointing in the same area, and the point to be measured is between 1 and 3 m from all the sensors, the precision at this point is sufficient.

8. A high number of erroneous pixels may indicate a problem, but a low number of these is not necessarily an indication of quality.

References

- Bonnechère, B., Jansen, B., Salvia, P., *et al.*, 2014. Determination of the precision and accuracy of morphological measurements using the Kinect™ sensor: comparison with standard stereophotogrammetry. *Ergonomics*, **57**(4):622-631.
<http://dx.doi.org/10.1080/00140139.2014.884246>
- Chang, Y., Chen, S., Huang, J., 2011. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. *Res. Deve. Disabil.*, **32**(6):2566-2570.
<http://dx.doi.org/10.1016/j.ridd.2011.07.002>
- Essmaeel, K., Gallo, L., Damiani, E., *et al.*, 2012. Temporal denoising of Kinect depth data. Proc. 8th Int. Conf. on Signal Image Technology and Internet Based Systems, p.47-52. <http://dx.doi.org/10.1109/SITIS.2012.18>
- Essmaeel, K., Gallo, L., Damiani, E., *et al.*, 2014. Comparative evaluation of methods for filtering Kinect depth data. *Multim. Tools Appl.*, **74**(17):7331-7354.
<http://dx.doi.org/10.1007/s11042-014-1982-6>
- Fernández-Baena, A., Susín, A., Lligadas, X., 2012. Biomechanical validation of upper-body and lower-body joint movements of Kinect motion capture data for rehabilitation treatments. Proc. 4th Int. Conf. on Intelligent Networking and Collaborative Systems, p.656-661.
<http://dx.doi.org/10.1109/INCoS.2012.66>
- Freitas, D., da Gama, A., Figueiredo, L., *et al.*, 2012. Development and evaluation of a Kinect based motor rehabilitation game. Proc. SBGames, p.144-153.
- Gonzalez-Jorge, H., Riveiro, B., Vazquez-Fernandez, E., *et al.*, 2013. Metrological evaluation of Microsoft Kinect and Asus Xtion sensors. *Measurement*, **46**(6):1800-1806.
<http://dx.doi.org/10.1016/j.measurement.2013.01.011>
- Haggag, H., Hossny, M., Filippidis, D., *et al.*, 2013. Measuring depth accuracy in RGBD cameras. Proc. 7th Int. Conf. on Signal Processing and Communication Systems, p.1-7.
<http://dx.doi.org/10.1109/ICSPCS.2013.6723971>
- Khoshelham, K., Elberink, S.O., 2012. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, **12**(2):1437-1454.
<http://dx.doi.org/10.3390/s120201437>
- Mallick, T., Das, P.P., Majumdar, A.K., 2014. Characterizations of noise in Kinect depth images: a review. *IEEE Sens. J.*, **14**(6):1731-1740.
<http://dx.doi.org/10.1109/JSEN.2014.2309987>
- Mkhitarayan, A., Burschka, D., 2013. RGB-D sensor data correction and enhancement by introduction of an additional RGB view. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, p.1077-1083.
<http://dx.doi.org/10.1109/IROS.2013.6696484>
- Olesen, S.M., Lyder, S., Kraft, D., *et al.*, 2015. Real-time extraction of surface patches with associated uncertainties by means of Kinect cameras. *J. Real-Time Image Process.*, **10**(1):105-118.
<http://dx.doi.org/10.1007/s11554-012-0261-x>
- Oliver, M., Molina, J.P., Montero, F., *et al.*, 2014a. A multisensor system for positioning of multiple users. Proc. XV Int. Conf. on Human Computer Interaction, Article 59.
<http://dx.doi.org/10.1145/2662253.2662312>
- Oliver, M., Molina, J.P., Montero, F., *et al.*, 2014b. Wireless multisensory interaction in an intelligent rehabilitation environment. Proc. 5th Int. Symp. on Ambient Intelligence, p.193-200.
http://dx.doi.org/10.1007/978-3-319-07596-9_21
- Oliver, M., Montero, F., Molina, J.P., *et al.*, 2015a. How many Kinects should look at you? A multi-agent system approach. Proc. 13th Int. Conf. on Practical Applications of Agents and Multi-Agent Systems, p.105-112.
http://dx.doi.org/10.1007/978-3-319-19629-9_12
- Oliver, M., Montero, F., Fernández-Caballero, A., *et al.*, 2015b. RGB-D assistive technologies for acquired brain injury: description and assessment of user experience. *Expert Syst.*, **32**(3):370-380.
<http://dx.doi.org/10.1111/exsy.12096>
- Regazzoni, D., de Vecchi, G., Rizzi, C., 2014. RGB cams vs RGB-D sensors: low cost motion capture technologies performances and limitations. *J. Manuf. Syst.*, **33**(4):719-728.
<http://dx.doi.org/10.1016/j.jmsy.2014.07.011>