Gestural Recognition by a Four-Kinect Module in a CAVE “Le SAS”

Saleh Salous saleh.salous@citu.fr

Julien Newton soosaine@ece.fr

Laure Leroy Laure.Leroy@citu.fr

Safwan Chendeb safwan.chendeb@citu.fr

CiTu paragraphe, Paris 8 university

2 Rue de la Liberté, 93200 Saint-Denis, France

http://www.citu.info/, http://www.univ-paris8.fr/

ABSTRACT

Among the various types of interactions in a virtual reality (VR) room, physical gestures such as moving one’s hands are one of the ways to communicate with the virtual environment. It is possible to capture those gestures with sensors, in our case the Microsoft Kinect. In the context of a large CAVE, not one but four Kinects are used to track the user and capture its gestures. However, the question of the detection of specific gestures by this system comes into play as the CAVE requires recognizing the user’s gestures in order to interpret them into an input. In this paper we focus on the gestural recognition ability of our CAVE applied to three specific gestures: Raising the right hand, raising the left hand and short hopping. Moreover, we provide experimental results showing that a Four-Kinect system provides more effective gesture recognition than a single sensor.

Author Keywords

Multi-Kinect; gestural interaction; gesture recognition; virtual reality; CAVE.

INTRODUCTION

In the field of Virtual Reality, one of the most important aspects is the interaction between the human user and the computers processing the virtual environment. In order to create a fluid and seamless user experience, it is necessary to properly track the user, recognize its inputs and successfully apply them to the VR scene.

In our context, we have a CAVE named “Le SAS” with a two-screen set-up that includes a front screen and a ground screen. (see Figure 1)

We discussed the topics of multi-sensor configurations, user tracking and camera calibration in [7][8]. The goal of this article is to propose a method to recognize specific gestures from the user in order to translate them into inputs.

Indeed, the number of potential interactions is high, and this step is essential to provide proper interaction between the users and the VR environment.

Figure 1 : Representation of “Le SAS”

This paper is organized as follows. Section 0 provides a literature review. Section 0 will outline our approach and explain the data collection and posture recognition processes. Section 4 will focus on experiments designed to test our approach. The test results are also available for reference. Finally, Section 0 will conclude on the acquired data.

STATE OF THE ART

Gestural recognition is a known topic in computer vision and analyses of gestures in a VR context have already been conducted [1, 3, and 4].

Several notions have previously been established in this field, such as types and categories of gestures.

For instance, in his survey of gestural interactions, [4]defined types of gestures as well as a large amount of specific movements and explained their use in various fields such as medical science or entertainment.

[3] Provided an extensive depiction of the state of the art in 3D gesture recognition, and explained the upsides and drawbacks of the technology.

Other researchers such as [1] chose to focus on the gestures in themselves and analyze the potential risks of physical exhaustion, muscular fatigue and pain that some of the gestures can provide if done repeatedly. This kind of analysis is useful in order to determine the least uncomfortable gestures to use in an interactive environment such as the SAS.

Gestural recognition was successfully used in other experiments; for instance [5] developed a gesture recognition application that can use a low-cost webcam to recognize hand gestures and translate them into interactions with digital objects in a virtual environment.

[6] are also able to track gestures, in the context of Augmented Reality applications. However, their approach focuses on using finger-mounted tracking targets to provide a higher level of accuracy and individual finger tracking.

Among the fields in which gestural recognition can be used, history is the one chosen by [9] with a VR application that is implemented in a museum and allows visitors to explore a 3D re-creation of an ancient ship called Vrouw Maria using a hand gesture-based interface.

In [2], the authors created and presented a complete system with a stereo camera and a microphone to receive inputs from the user, a projector to display the virtual world with which the user interacts, as well as software developed to link recognized gestures and speech with the VR system's visual output.

PROPOSED APPROACH

In this section, we will explain the methodology we follow to recognize inputs in our CAVE. This part will first explain the collection of joint data and then discuss the posture detection technique in the CAVE.

Posture data collection and posture detection

The environment in which the data is collected is a CAVE named "Le SAS" with two screens. Each is 3 meters high and four meters wide. Data about the user's location and its joints is measured by four Kinects in a multi-sensor system that was notably discussed in [7].

The objective of this system is to cover the maximum amount of space in the "SAS" and track the user's gestures regardless of its position inside the CAVE.

As the user is supposed to be able to move around the "SAS", the multi-sensor system tracks the user's coordinates for each frame and an algorithm [8] is used to determine in real-time which of the Kinects have the more accurate joint data.

The actual algorithm used by our module to detect gestures can be observed in Figure 2. Each Kinect sends joint data to the gesture recognition module in real-time. As all Kinects may not be very close to the user, the module applies a filtering mechanism discussed in [8], where only the most relevant Kinects for the desired measured joints are taken into account.

In our case, the specific joints required for the detection of the three gestures we wish to test (cf Section 0) are both elbows, both hands and both knees.

Once the user joint data is properly recorded by the CAVE's multi sensor-system, the potential gestures reflected by those coordinates have to be detected as such before being translated into inputs to interact with the VR environment. The second part of the algorithm is tasked with analyzing the filtered joint data and detecting gestures.

In order to recognize a raised hand, the algorithm compares the X and Y coordinates of the hand and elbow joints. If the hand's two coordinates are higher than the elbow's, the module returns a gesture notification.

The method used to detect the short hopping movement is to compare the differences in the X and Y coordinates of both knees between a previous frame and the current frame. If both knees' joint data provide a difference lower than a chosen threshold, the system signals a short hop.

It is also possible to develop other specific posture detection calculations related to new gestures we may intend to implement in our CAVE. The reason why the hand-raising gestures have been chosen for this experiment is related to another input device previously implemented on the "SAS". Interaction with the "SAS" is also possible with a joystick that allows the users to move on the left or the right. As a consequence, the tested gestures aim at replicating those movements.

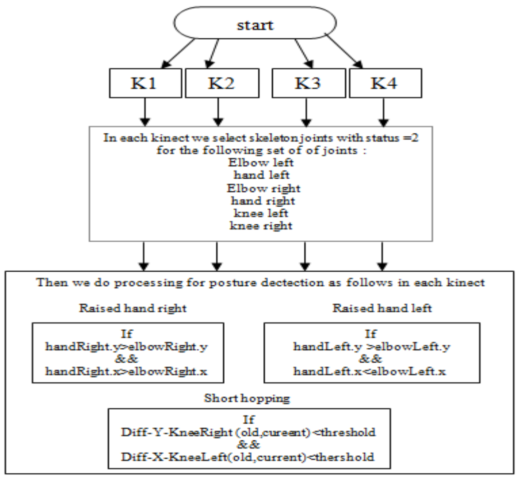

Figure 2: Posture detection algorithm

The technical context related to the data transmission and the software libraries is as follows. Accessing the Kinect joint data from the multi-sensor system and using the values to perform recognition operations was possible thanks to the Microsoft Kinect SDK.

The data transmission itself used the UDP protocol. This protocol was chosen as to send and receive data in real-time.

Constraints of the CAVE

While gesture recognition on its own requires complying with a few constraints, the environment in which we wish to deploy the gestural recognition abilities of our sensors provides new challenges.

Indeed, the gestures are used to interact with the VR environment. Therefore, the user should do movements that will not interfere with his or her experience within the CAVE. For instance, asking the user to move backwards in order to make his virtual self-move backwards is potentially dangerous as the user can fall off the ground screen if he or she is not cautious. A similar issue can happen with moving forward as the user may collide with the front screen.

Those constraints have to be taken into account when designing what gestures will be required for our VR applications. However, it is still possible to test gestural recognition with gestures that may not be implemented in future works.

EXPERIMENT

Posture detection

Experiment methodology

Our starting hypothesis is that four Kinects working together are more accurate and cover more potential user locations than a single sensor and are therefore more suited to capture gestures in a CAVE.

In order to obtain proper feedback on the effectiveness of our methodology and to provide conclusive data on our hypothesis, we set-up experiments to test whether the four Kinects system properly tracked the users' gestures and whether it was more reliable than a single Kinect.

Participants

The experiment was conducted as follows: 23 participants were invited to the "SAS" and were not informed beforehand about the specific objective of the experiment. This was done to avoid creating bias in the users' behavior during the tests. All of the participants are adults. 12 of them are male and 11 are female.

Detected Postures

Three gestures were tested in this experiment: raising the right hand, raising the left hand, and doing a short hop as seen in Figure 3. We asked each of the participants to do each of these gestures four times at random locations inside the "SAS" while a single Kinect was tracking them. This procedure was repeated with the four Kinects system instead.

Overall, each person made 3 different types of gestures. Each specific type of gesture was repeated four times by each participant for both Kinect set-ups (single Kinect and four Kinects). As 23 people took part in the experiment, the total number of recorded gestures is 276 for each Kinect configuration inside the "SAS" for a total of 552 gestures.

Figure 3: Participants in the middle of the experiment performing the three gestures a) Raising the left hand, b) Raising the right hand, c) Short hopping

The system analyzed the data in real-time to recognize the participants' gestures using the algorithm discussed in Section 0. We recorded this data to provide the results of our experiment.

False detection

As our approach aims at providing an accurate motion sensing system, there is another variable that has to be taken into account before interpreting the experiment results.

Indeed, the phenomenon known as "False Detection" can occur and provide a layer of inaccuracy to the data captured from the user's movements.

False detection happens whenever one of the sensors registers a motion that was not actually performed by the user, thus losing in overall accuracy. In our case, we measured the false detection issue regarding our three tested gestures within our SAS.

The testing protocol is as follows. A random user chosen from the 23 participants is asked to move as he or she wishes inside the SAS during 15 minutes without performing any of the three gestures: raising either hand or short hopping. We then analyze the data provided by our gesture recognition system and find the number of false detections.

The aforementioned experiment was conducted twice: once with a single Kinect, and once with the Four-Kinect set-up.

Another type of false detection also exists. False Detection can also occur when the system exhibits erroneous gesture recognition and does not detect the proper gesture when a user performs it. This type of false detection can be measured from the same data pool than the one used to determine the sensors' gesture recognition accuracy.

Experiment results

Detected postures

Figure 4 shows the results for the three gestures (raising the right hand, raising the left hand and short hopping). As each gesture was repeated four times by each of the 23 participants, the data pool from which the experiment results were extracted contains 92 recordings for each combination of gesture and Kinect set-up. The results of our experiment are as follows. While the right hand raising gesture was successfully recognized by the four-Kinect system 93% of the time, only 77% of the gestures were detected by the single Kinect. The success rate for the left hand raising gesture is similar to the right hand one, with the four-Kinect system accurately capturing 91% of the gestures and the single Kinect only 75% of them. Short hopping proved to be a slightly more difficult gesture to detect according to the results. Indeed, the four-Kinect system's success rate is 87% and the single Kinect's is 68%.

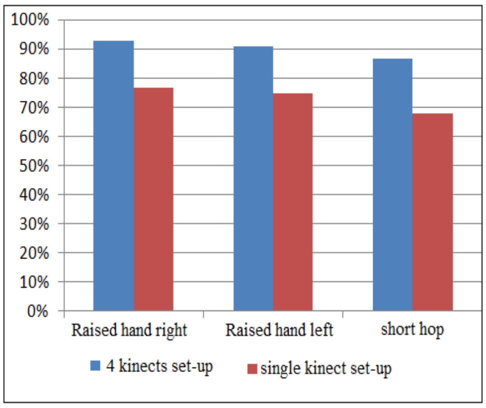

Figure 4: Level of accuracy of the gesture recognition module for each of the three tested gestures and both Kinect configurations

False detection simulation

As explained in Section 0, our experiment also measured the risk of a false detection occurring while using the SAS.

The results are as follows. The single Kinect set-up returned 17 false detections during the 15 minute-long test. The Kinect detected a short hop 9 times, recognized a raised right hand 5 times and returned a raised left hand 3 times.

However, data fusion has proved to significantly decrease the number of false detections in a multi-Kinect set-up. Once the 4 Kinects' joint data was fused into a single skeleton, only 6 false detections were returned by the system. 3 short hops, 2 raised right hands and 1 raised left hand were detected, thus providing a lower amount of false detections than a single Kinect.

CONCLUSION

Gestural recognition is a technical field with several constraints and its implementation in a Virtual Reality environment adds more issues that have to be considered when developing gestural recognition solutions.

Our experiment demonstrated that the use of a multi-sensor system provides more accuracy than a single sensor. As shown by the experiment results, the use of a multi-Kinect system increased the accuracy of the gesture detection by 16 percentage points for arm-raising gestures and 19 percentage points for short hopping. Furthermore, the four-Kinect set-up is suited to the size of a CAVE and complies with the specific constraints of the "SAS" as it provides a larger detection range.

ACKNOWLEDGEMENTS

This research and its results are made possible thanks to the members of the CiTU-Paragraphe lab. Thanks to Sobhi AHMED in particular for his assistance on the subject and his support and advice.

REFERENCES

1. Ahmed, S., Leroy, L., Bouaniche, A., (2014). Approches de Réduction des Impacts Négatifs des Interactions Gestuelles sur le Corps Humain. Reims Image 2014, p113-121.
2. Demirdjian, D., Ko, T., & Darrell, T. (2005). Untethered gesture acquisition and recognition for virtual world manipulation. Virtual Reality, 8, 222–230
3. LaViola Jr., J. J. 3D Gestural Interaction: The State of the Field. ISRN Artificial Intelligence 2013, pp. 1 – 18.
4. Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate Wiki Users : Results of a Survey. Wall Street Journal, pp. 99–104.
5. Rautaray, S. S. (2012). Real Time Hand Gesture Recognition System for Dynamic Applications. International Journal of UbiComp, 3(1), 21–31. doi:10.5121/iju.2012.3103
6. Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., & Rigoll, G. (2007). Static and dynamic hand-gesture recognition for augmented reality applications. Human-Computer Interaction, Pt 3, Proceedings; 4552, 728–737.
7. Salous, S., Ridene, T., Newton, J., and Chendeb, S,. "Study of geometric dispatching of four-kinect tracking module inside a Cave," In Proc. 10th International conference on Disability, Virtual Reality and Associated Technologies, Gothenburg, 2014, pp. 369-372.
8. Salous, S, Newton, J., Leroy, L., and Chendeb, S,. "Dynamic sensor selection based on joint data quality in the context of multi-kinect module inside the CAVE "Le SAS".IJCTE, V9N1, Feb. 2017.
9. Şen, F., Díaz, L., & Horttana, T. (2012). A novel gesture-based interface for a VR simulation: Re-discovering Vrouw Maria. Proceedings of the 2012 18th International Conference on Virtual Systems and Multimedia, VSMM 2012: Virtual Systems in the Information Society, 323–330.