Kinect based Real Time Gesture Recognition Tool for Air Marshallers and Traffic Policemen

Anusha Prakash, Swathi R, Stuti Kumar, Ashwin T S and G Ram Mohana Reddy
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, Mangalore, India
{anusha.blr, swathirajanna94, stutikumar06, ashwindixit9, profgrmreddy}@gmail.com

***Abstract*—The Microsoft Kinect which is a motion sensing input device presents a very straightforward and affordable approach to facilitate real-time user interaction. Although a lot of research has been conducted on the application of Kinect to gaming and virtual reality environments, its relevance to real-world scenarios has not been explored much. The features provided by the driver platforms such as OpenNI and Microsoft Kinect Software Development Kit (SDK) for development using Kinect coupled with the motion sensing ability of Kinect, presents a unique opportunity for extending the scope of the Kinect sensor. This paper proposes a system for automatically recognizing the road traffic control gestures of police officers and air marshalling commands by ground personnels. This system is aimed for self-learning, training and testing these officers to equip them with the skills to tackle real-world situations. Since these applications are very crucial and performing accurate gestures are of at most importance, this system will prove to be very essential. Experimental results also demonstrate that our system is robust and effective and is suitable for real-time application.**

***Keywords*—Gesture Recognition, TTS, Assistive Technology, Human Computer Interaction**

I. Introduction

During the past decade, the interaction between human and computer, also known as Human Computer Interaction (HCI) has gained a lot of popularity especially due to the advent of the Microsoft Kinect. It has emerged as a pioneering field with the aim to bridge the communication gaps between humans and computers [1]. Among these interactions, the Hand Gesture Recognition has become a prominent field of research as some situations demand silent communication employing gestures. Gesture and Gesture recognition are very significant and heavily encountered terms in HCI. Gestures are the movement of the body or physical actions typically constructed by the combination of hand movements and orientations of the user meant to convey some useful information [2]. Gesture recognition is a process where the system is able to successfully identify the gestures performed by the user. This is very relevant in situations demanding silent communication, where visually transmitted gestures are employed in the place of acoustic sounds for communication.

The Microsoft Kinect along with the drivers for Kinect such as OpenNI and Microsoft Kinect SDK provide an affordable and easy way for real-time user interaction using gestures. Gestures are widely used for communication in day-to-day life usually for exchanging information, ideas, thoughts, providing commands etc. It is employed as a means of communication among dumb and deaf people in the form of sign language. The Kinect has been employed extensively to develop applications for sign language recognition to aid communication between dumb and deaf people and people who are not well acquainted with sign language, thus helping reduce social limitations of the disabled. However, one particular direction where Kinect has not been explored to its limits is the usage of Kinect to develop control gesture training and testing applications for real-world scenarios. In this paper, we mainly focus on two such applications : 1) Road traffic control gestures by police officers and 2) Air marshalling commands by ground personnels.

Human traffic control is preferred for developing nations because of the relatively few cars and major intersections, and cheap human traffic-controllers [2]. In a human traffic control environment, police officers perform traffic control command gestures using either or both of their hands. The vehicle drivers are expected follow the commands of the traffic policeman and drive accordingly. A human traffic control system outperforms an automated traffic control system in case of event-related emergencies like accidents, road blocks, etc. due to the dynamic decision-making ability of the police officers who are physically present in the scene and can witness the situation around to make apt decisions to divert the traffic. The decision made by the police officer to handle the traffic is very crucial and and the traffic command gestures performed by him needs to be unambiguous and authentic. To ensure the safety of the drivers and the smooth moving of the road traffic, it is very essential that the police officer is well aware of the various traffic control commands and is able to perform them accurately in the real-world scenario. This paper proposes a system for training these policemen for performing proper traffic command gestures and testing them to guarantee error free traffic control.

Air mashalling is visual signalling between ground personnel and pilots on an airport, aircraft carrier or helipad [3]. The marshallers give take-off, keep turning, slow down, stop, shut down engines and landing clearances to aircrafts and helicopters, leading them to parking stand or to the runway. This human air marshalling approach servers as a better choice when compared to radio communications in situations where there is very limited space and time between take-offs [4].

978-1-5090-6115-0/16 $31.00 © 2016 IEEE
DOI 10.1109/T4E.2016.14

To ensure glitch free aircraft ground handling, these ground personnels performing air marshal commands needs to be trained well as even a small blunder may lead to huge loss of life and resources. The system proposed caters well to this need of training the air marshallers to perform accurate air marshalling command gestures.

The rest of the paper is divided into 4 sections. In section II we discuss about the various gesture training and testing approaches that already exist for these two applications. Section III outlines the details of the proposed gesture training and testing system. In section IV we discuss about the experimental results demonstrating our approach's performance. Finally we conclude with the challenges associated with the proposed approach and few possible future work in section V.

II. Literature Survey

In the past, a few methods for gesture recognition of traffic control have been developed in the literature. Wenjie et al. [5] applies the relative relationship between human body skeletons to classify different actions. It is favorable for solving the problem resulted from relative movements between Kinect and the traffic police. Bang Le et al. [6] calculate the palm center's coordinates based on the moment of hand contour feature and present a new method of detection and identification, called Python programming environment, which realizes the gesture track recognition based on the depth image information obtained by the Kinect sensor. This method proved very effective to achieve interactive features. Police gestures from the complementary body parts were recognized by Fan Guo et al. [7] on the colour image plane. The traffic police officer in the complex landscape is identified by extracting the reflective vest using colour thresh-holding hence; the detection results of this approach were densely influenced by background and outdoor brightness. Yuan Tao et al. [8] attached on body sensors onto the back of the officer's hand to obtain the gesture data. This accelerometer-based sensor approach yields accurate hand position but it needs an exclusive communication protocol for the vehicles. D. Kumarage et al. [9] carried out the process of breaking down motion gestures to sub components for parallel processing and mapping motion data into static data representation. This provides for minimal image processing and helps to reduce latency and convert it to text in real time in order to make the communication to run smoothly. Stephen Witherden et al. [10] at Beca Applied Technologies explored whether the integration of Microsoft Kinect and VBS2 (militarised version of the commercially available game) simulation environment can be used to build a model for personnel training in aircraft marshalling. The Dynamic time warp algorithm was very effective in identifying gestures involving movement as well as stationary ones (poses). However, there was an interplay between hover and other signals as all signals include the hold signal as part of their possible movement. This ambiguity lead to confusion in the algorithm and hence resulted in false positives.

III. Proposed Methodology

The basic idea is to build an application which acts as a simple hand gesture recognition system, which is capable of detecting static as well as dynamic hand gestures. We have considered Kinect as our standard input peripheral to capture movements and gestures. The choice of our capturing device is based on ease of installation in various places including homes, training centres, labs etc., hence helping to get most out of the built system at lower costs. Kinect camera is highly efficient due to its features such as skeletal tracking, raw sensor streams, depth sensors and advanced audio capabilities. It allows us to capture and analyse gestures in real time.

A basic architecture of the proposed system is depicted in Figure 1. Our system is designed to track user's movements. The kinect camera is fixed at a certain height in order to track the user's entire body structure within a practical ranging limit of 1.2 to 3.5m. It is also capable of tracking the user movements while the user is seated, where only the upper body of the user is captured. It first extracts the manual features and detects and maps the joints to render a skeletal equivalent illustration of the user. All the movements of the user's joints are continuously tracked and the movements are put forth on the User interface of the system. To implement a system capable of deducing Traffic Signal and Air Marshalling commands, an algorithm to capture gestures incessantly has been proposed to ensure that each small movement that is tracked is broken into segments and the combination of segments then put together and equated to a particular gesture. Every gesture is divided into three components, the starting image, the movement and the final image. These three features help deduce the complete gesture. It is possible for more than one set of hand signals to have the same first joint position. The system is designed to interpret the movements in real time, thus keeping track of the movements made up to a certain point and matching it to the most suitable gesture.

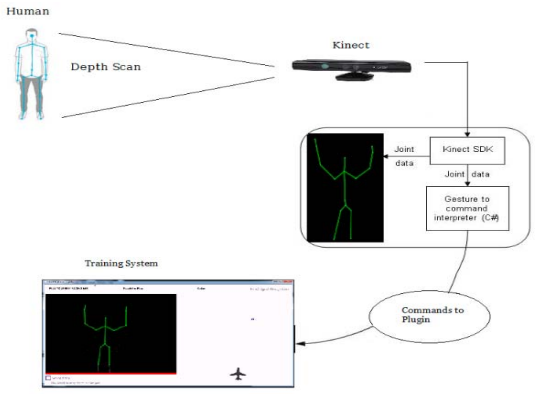

Fig. 1. System Architecture

Kinect also provides us with a coordinate system, where the X, Y, Z coordinates of the user joints are measured. If a specific joint moves to a specific relative position for a certain amount of time, then the system detects a gesture. The

amount of time is based on the number of frames captured at that relative position. Using the same logic the hand signal gesture recognition algorithm can be designed for training traffic police officers and Air marshals. Suppose a wave has to be detected. Generally people wave by moving their arm back and forth with their palm facing front held on almost the same level as their shoulder, while keeping their elbow completely still acting as the pivot. The hand wave can either be the right hand or the left hand, and for each the coordinates are specified at the beginning and the end, and tracked the entire movement in the middle.

The application can detect several hand gestures defined within the application in order to manipulate an object in the virtual environment. It can currently detect Left hand wave, Right hand wave, folded hands, right hand swiped up, right hand swiped down, left hand swiped down and left hand swiped up. For the traffic recognition system, it recognizes right and left hand stop gestures which can be deduced to stop all vehicles in every road directions, to stop all vehicles in front of and behind the traffic police officer and to stop all vehicles on the right of and behind the traffic police officer. For the Air Marshalling system, The gestures, Hold/Stand by, Pull yourself facing me, Pull yourself facing me right and Pull yourself facing me left can be detected by the system and it converts the inferred result into text and audio output in real time while simultaneously moving an object in the virtual environment according to the commands interpreted.

The gestures have been used to build an application for manipulation of objects in virtual environment. This entire system has been implemented in C# with the use of Visual studio. The hardware requirements of the implemented system includes a computer with 32-bit (x86) or 64-bit (x64) processors, Dual-core, 2.66-GHz or faster processor, USB 2.0 bus dedicated to the Kinect, minimum 2 GB of RAM and a Graphics card that supports DirectX 9.0c. The Kinect camera used in the application setup captures image sequences at the resolution of 640 x 480 pixels at 30 Hz.

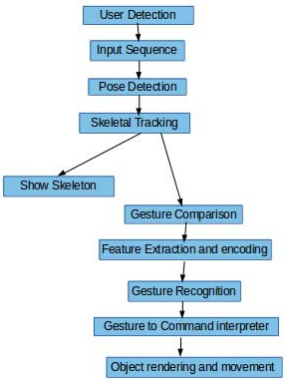

Fig. 2. Flowchart for Gesture Recognition

Figure 2 is a flowchart which depicts the entire process of gesture recognition. Initially, the user and user height is detected, followed by detection of the user's input sequence and pose. Based on this, the skeleton tracking occurs and the skeleton of the user is displayed on the screen. The gesture performed by the user is compared with a standard gesture data-set after extracting the gesture features. In this way the gesture performed by the user is recognized. After gesture recognition, the gesture is interpreted as a traffic or air marshalling command. Thus this command controls the object rendering and movement.

IV. Results and Analysis

The gesture recognition happens in real-time which involves breaking down the motion into sub-components. In the case of dynamic gestures, the frames of motion are composed into segments, the depth recognition is made, and finally the gesture is recognized. A text output as well as an audio rendering of the recognized gesture is enabled to benefit the disabled.

A. Air Marshalling Command Recognition

In case of Air Marshalling, a graphic gaming interface consisting of an animated aeroplane and a flag-post is created. The task at hand for the air marshallers is to navigate the aeroplane using marshalling commands and ensure that it reaches the flag-post. The air marshalling commands implemented include *stand by, hold, pull yourself forward, left accelerate* and *right accelerate*. Thus, through this system the air marshallers can easily learn and practice the commands. Figure 3 depicts the *left accelerate* command which will accelerate the plane in the left direction. Similarly, it can be seen in Figure 4 that the aeroplane is being accelerated in the forward direction by performing the *pull* command.

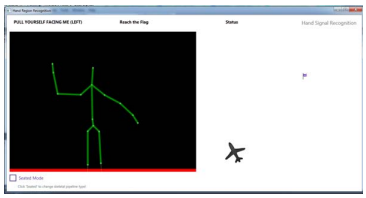

Fig. 3. Air Marshalling - Left accelerate

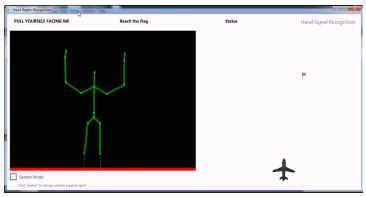

Fig. 4. Air Marshalling - Pull Command

B. Road Traffic Command Recognition

The Road traffic Command Recognition System is build to recognize road traffic hand signals performed by the traffic policemen. The Road Traffic Commands in general are meant

to stop the vehicles to the left, right, behind and in front of the policeman. Initially, when the commands are performed by the officers, the 3D human model is simplified to a 2D skeleton composed of body joints. Kinect by default provides mapping for 20 body joints including *head, spine, shoulder left, wrist right, hand right,* etc. The relative angles between these joints are calculated, based on which a distinction is made to recognize corresponding traffic signal gestures. Figure 5 and Figure 6 show how the real-time traffic command recognition occurs. Thus our system acts as an effective tool for training traffic police officers.

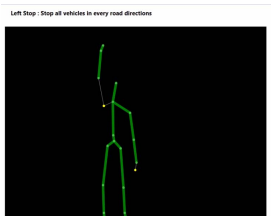

Fig. 5. Traffic Signal Command Recognition - Left Stop

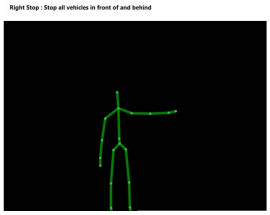

Fig. 6. Traffic Signal Command Recognition - Right Stop

Figure 7 is an Accuracy Graph which compares the time required to capture a user's movements for different actions and of users of varying heights. The graph depicts the time taken to interpret a gesture for three different gestures, the wave left, wave right and folded hands of users of varying heights ranging from 155 to 178 cm. Different points on the graph represent time taken to obtain the accurate result after gesture comparison.

Table I is a representation of the same comparison in tabular format. As we can observe, the detection time is maximum around the average and is less on either side. Though the values are a little inconsistent, we can observe that the time taken for gesture recognition increases with the user height initially and then drops.

V. Conclusion

We conclude by saying that the system we have developed has a lot of scope in real-world applications and can be employed easily to train and test air marshallers and traffic policemen. This will help ensure hassle free training process for the officers and also let them have a first-hand real-time experience before they actually start working. The developed gaming system is also very interactive and hence is more adaptable to use.

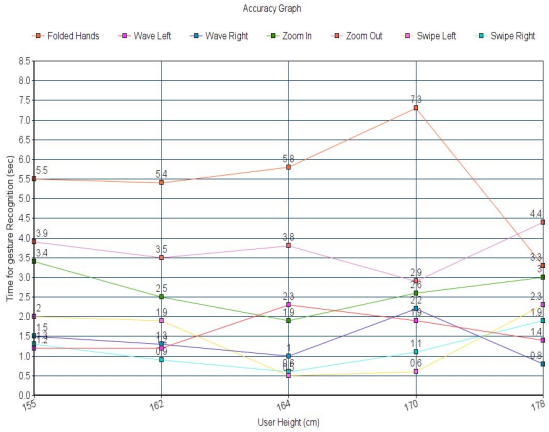

Fig. 7. Time Graph for various gestures and users

TABLE I
Time taken for Gesture Recognition

	User Height(cm)				
Gestures	**155**	**162**	**164**	**170**	**178**
Folded Hands	5.5	5.4	5.8	7.3	3.3
Wave Left	1.2	1.2	2.3	1.9	1.4
Wave Right	1.5	1.3	1.0	2.2	0.8
Zoom In	3.4	2.5	1.9	2.6	3.0
Zoom Out	3.9	3.5	3.8	2.9	4.4
Swipe Left	2.0	1.9	0.5	0.6	2.3
Swipe Right	1.3	0.9	0.6	1.1	1.9

References

[1] Omer Rashid, Ayoub Al-Hamadi and Bernd Michaelis. "A Framework for the Integration of Gesture and Posture Recognition Using HMM and SVM", Intelligent Computing and Intelligent Systems, 2009. ICIS 2009, 20-22 Nov. 2009, Shanghai

[2] Meenakshi Panwar, "Hand Gesture Recognition based on Shape Parameters", International Conference on Computing, Communication and Applications (ICCCA), 22-24 Feb. 2012, Dindigul, Tamilnadu

[3] Shahriar Khan,"Automated versus human traffic control for Dhaka and cities of developing nations",10th International Conference on Computer and information technology, 2007. iccit 2007, 27-29 Dec. 2007, Dhaka

[4] "Air Marshalling", http://en.wikipedia.org/wiki/AircraftMarshalling

[5] Song Wenjie, Fu Mengyin ,Yang Yi , Chen Yao. "Recognition method of traffic police and their command action based on Kinect", Control Conference (CCC), 2014 33rd Chinese,28-30 July 2014,Nanjing.

[6] Van Bang Le, Anh Tu Nguyen, and Yu ZhuHand. "Detecting and Positioning Based on Depth Image of Kinect Sensor", International Journal of Information and Electronics Engineering, Vol. 4, No. 3, May 2014,Shanghai.

[7] Fan Guo, ZixingCai, Jin Tang, "Chinese Traffic Police Gesture Recognition in Complex Scene", Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE 10th International Conference.

[8] Yuan Tao, Wang Ben. "Accelerometer-based Chinese traffic police gesture recognition system", Chinese Journal of Electronics, pp. 270-274, 2010.Article (CrossRef Link)

[9] D. Kumarage, S. Fernando, P. Fernando, D. Madushanka and R. Samarasinghe,"Real-time Sign Language Gesture Recognition Using Still-Image Comparison and Motion Recognition",2011 6th International Conference on Industrial and Information Systems, ICIIS 2011, Aug. 16-19, 2011, Sri Lanka

[10] Stephen Witherden, Air Marshalling with the Kinect, Beca Applied Technologies