Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 90 (2016) 37691-37694

ISSN: 2229-712X

Speech Generation using Kinect to Aid the Mute in Public Addressing

T. Shantha Kumar

Department of Computer Science Engineering, Alpha College of Engineering, Chennai, TamilNadu, India.

ARTICLE INFO

Article history:

Received: 20 April 2015;

Received in revised form:

8 January 2016;

Accepted: 13 January 2016;

Keywords

Kinect,
Gestures,
Speech,
Synthesizers,
Text to speech converter.

ABSTRACT

Various new technologies and aiding instruments are always being introduced for the betterment of the challenged. This project focuses on aiding the mute in expressing their views and ideas in a much efficient and effective manner thereby creating their own place in this world. The proposed system focuses on using various gestures traced into texts which could in turn be transformed into speech. The gesture identification and mapping is performed by the Kinect device, which is found to cost effective and reliable. A suitable text to speech convertor is used to translate the texts generated from Kinect into a speech. The proposed system though cannot be applied to man-to-man conversation owing to the hardware complexities, but could find itself very much of use under addressing environments such as auditoriums, classrooms, etc.

Introduction

The depth sensors finds its existence in utmost all possible fields of lately. The so considered pioneer of such depth sensor devices is the Kinect. Though initially was introduced with the sole purpose of satisfying the gaming community, but recently finds its intervention in all most every application possible. The Kinect's skeletal tracking system is used in mapping the gestured that are to be performed by the user.

In addition to the various general functions and tasks such as recognition of hand signs, identifying small activities and tracking movements, that are performed by the regular recognition devices, this paper aids in adding additional feature of tracking mechanisms which in turn helps in exploration of body movements.

The system could be classified unto three sections. The first elaborates on the functionalities of Kinect. The next section implies on our attempt to replicate gestures into alphabets. And the final section focuses on framing the words from the gestures that are performed. These words could be articulated using a simple text to speech converter.

Related Work

Much of the process that is to be implemented here uses Kinect. "Hand Gesture Recognition System" portrays the use of normal USB camera to detect gestures. These techniques involves basic steps in image processing, using which it identifies gestures based on variations in multiple frames of a video. This however restricted on the basis of the camera utilized. Other systems involved the process of using a markerless full body tracking system such as the system suggested by Belinda Lange, Skip Rizzo, Chien Yen Chang, Evan A. Suma and Mark Bolas [4] of The University of Southern California, which proves to be a virtual reality system which would have greater significance in simulations and gaming. Though the application varies the working process is still found to be similar in more than one way.

The Kinect comprises of programmable depth sensors which is applied to track hand gestures. This is evident in the work shown in "Using Hand Gesture Recognition in Natural way: A Survey" [6]. It has explored the various hand gesture recognition techniques which can be broadly classified as Appearance based approaches and Model based approaches. Among the appearance based approaches, Archana S. Ghotkar and Gajanan K. Kharate [7] use different color models i.e., the RGB and the HSV models to isolate the human hand by applying three different algorithms for detection and an edge detection algorithm for filtering out the noise.

A similar work would be the efforts made in recognizing the hand gestures to map them to Bangla Sign Language [12] for performing simple operations such as opening an application, raising the volume and so on. But the basic problem that exists in using hand based gestures is that they are limited. Since most of the gestures that had similar movements, their accuracy was limited to a dozen. In order to overcome this, "Skeleton-Based Action Recognition with Extreme Learning Machines" [8] have focused on action and gesture recognition using the skeletal tracking feature of Kinect to map gestures to actions. To improvise over the inefficiencies of using the Kinect SDK, Chanjira Sinthanayothin, Nonlapas Wongwaen, Wisarut Bholsithi [9] use OPENNI, NITE Primesense and CHAI3D open source libraries to develop the skeletal structure. Although it possesses advantages over the Kinect SDK, it poses difficulties in calibration and installation.

To classify these gestures, various techniques have been proposed. To start with, Georgios Th. Papadopoulos, Apostolos Axenopoulos and Petros Daras [10] track and recognize gestures based on the calculation of spherical angles between selected joints and the respective angular velocities. For all these techniques, overreliance on the depth sensor contributes a large portion to the misinterpretation of the gesture. For improvisations, a technique that merges the depth stream with the RGB input from the camera [11] has been proposed to eliminate the faulty calibrations of the Kinect to provide a more accurate feature recognition.

G Tao, PS Archambault and MF Levin's [13] technique involves the use of Kinect's skeletal tracking system to generate a virtual environment for curing Upper Limb Hemiparesis. To achieve this, they used the Kinect SDK to track the movement of the joints and adding a score system to know if the patient has

Tele:
E-mail addresses: shanmailrajni@gmail.com

achieved the required objective. This score system uses a static matching pattern. Virtual Reality Based Rehabilitation and Game Technology [14] effectively portrays the usage of Skeletal tracking to train or rehabilitate people with motor disabilities. It involves the usage of a live and interactive teaching methodology to rehabilitate people. To measure the movements, it utilized the variations in the X and Y coordinates of the subject.

A couple of similar works is that of Jiann-Der Lee, Chung-Hung Hsieh and Ting-Yang Lin's [15] proposal of a system to create an interactive system to encourage motor disabled people to take up Tai Chi exercise using the Kinect's Skeletal tracking system and the other one being Helten T. Muller M. Seidel H. P. and C. Theobalt's technique [17] to identify postures by superimposing or comparing the new data with data present in the database. Both the systems used a static posture matching algorithm. To classify entire actions though, Human activity recognition using body pose features and support vector machine [16] proposed a system that utilizes the skeletal tracking feature of the Kinect to map activities (such as running, sleeping) performed by the subject. Using a Support Vector Classifier, he successfully identified twelve different actions performed by ten different people of varying physical structure.

Proposed Work

System Overview

The entire model comprises of three steps. The first step focuses on feature extraction which involves the tracking of skeletal structure. This initial step is then followed by feature classification which involves various operations performed using the postures. This final input is then fed into a regular speech converter which yields the specified word.

Feature Extraction

Skeletal tracker

The first instalment of Kinect i.e. version 1 of Kinect utilizes a 20 joint tracking system. The 20 joint comprises of the following: head, shoulder center, shoulder right, shoulder left, elbow right, elbow left, wrist right, wrist left, hand right, hand left, spine, hip center, hip right, hip left, knee right, knee left, ankle right, ankle left, foot right, foot left.

Fig 1. The Kinect's Skeletal Tracker

The Fig.1 replicates a typical skeletal tracker which is used. The Kinect makes use of structured light technique to generate discrete measurements of the three dimensional physical environment. When this is achieved it tracks points which vary upon alternative co-ordinates based on their depths. The joints that appear to be concealed the Kinect tracks these joints on the basis of 'inferred joints' identification routine. Through such a process the device assists in identifying and plotting the hidden or partially hidden joints of the user. Through the identification of inferred joints the rest of the body of the user could possibly be tracked as far as some portions of the body are visible. This possible avoids any unwanted obstacles which might possible interfere.

Kinect packages which could be used to track the joints and access the state of the joints and their corresponding change in values of the states are available commercially that can be accessed in C#, Visual C++ or Visual Basic.

Listing 1. Generating the Skeleton

1: for each skeletal_data u ∈ depth map Di do
2: if state (u) ==tracked
3: generate line with last tracked u
4: if state (u) ==tracked (positiononly)
5: generate line with tracked u
6: repeat till skeletal_data==max limit

Listing 1 generates the skeletal data. The joints in the same depth are detected. Once detected, the tracking of joints take place. These tracked joints are constantly monitored for their co-ordinates. Assumptions are made based for the obscured joints and conclusions are made when they are visible again.

Listing 2. Rendering the Skeletal Data

1: for each skeletal_data u ∈ head Hi do
2: DrawBone (Head, Shoulder_Center)
3: DrawBone (Shoulder_Center, Shoulder_Right)
4: DrawBone (Shoulder_Center, Shoulder_Left)
5: for each skeletal_data v ∈ l_arm li do
6: DrawBone (Shoulder_Left, Elbow_Left);
7: DrawBone (Elbow_Left, Wrist_Left);
8: DrawBone (Wrist_Left, Hand_Left]);
9: repeat till skeletal_data==max limit

Listing 2 generates the skeletal structure. A line is generated which replicates the skeletal structure of the user. This line is capable of varying dynamically which in turn results in varying co-ordinate points.

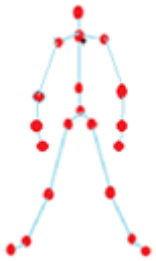

Fig 2. Output from the Skeletal Tracker

The Fig.2 depicts the output from the skeletal tracker that was generated using the Listings 1 and 2.

1) Tracking modes of the Kinect: The Kinect can track the user in two modes namely Seated Mode and Default mode

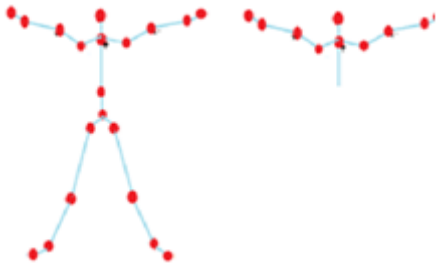

Fig 3. The Kinect's Default and Seated modes

The Fig.3 depicts skeletal tracking modes deployed by the Kinect. In seated mode, only the upper 10 joints out of the twenty are tracked. One of the important differences between the two modes is that in default mode, the user's position is tracked based on the distance between the background and the user whereas in the latter, the detection is based upon movement. Hence, seated mode requires the user to move back and forth to facilitate detection.

Feature Classification

Now that the skeletal structure tracking and accessing the skeletal structure's data has been elucidated upon, the next step involved is mapping the gestures. To map gestures, a tool was developed that used changes in coordinates of the joints and the states of the joints to detect gestures. This tool tracks every single joint and uses a simple UI for assigning the action that has to be performed for changes in state of the joints.

1)Tracking mechanisms: The tool measures changes in the X and Y coordinates. Also, it looks for the duration of a posture to activate the triggers. Using the changes in X and Y coordinates, detect activities of jumping and crouching. Angular variations measured using these coordinates are useful in differentiating between outstretched and mildly stretched joints. Distance between joints calculated similarly help in finding the proximity of the joints to another joint. The duration for which the joint remains at a point is mainly used to activate triggers. When a joint is no longer visible, or has been obscured, the state of the joint is accessed.

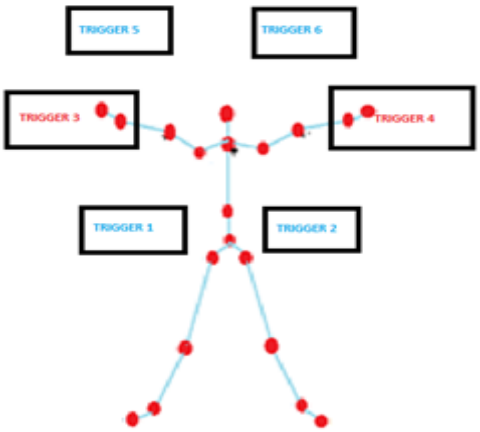

Fig 4. The Tracking tool

Fig.4 displays the output of the tracking tool having 6 triggers. The triggers get activated when the joints are located in its proximity for more than two seconds.

For mapping these gestures, a static method was used where each of the alphabets were printed on a simple text editor when the posture matched. Since the usage of a skeletal matching algorithm required more samples to train the machine to have an acceptable level of accuracy, a coordinate change technique was implemented. For actions that involved rotation, the action was recognized when the joints couldn't be tracked. This technique was chosen since the upper body always hid the other joints that were behind it during rotation. The efficiency was far better than those that involved measuring the angle of rotation. For linear movements such as stretching the hands horizontally or vertically, the change in the coordinates was far better criterion since the rise or fall in the coordinates of the tracking points was high and the possibilities of misinterpretation was less. For actions that involved bending the knee or jumping, the changes in the coordinates of tracking points of the head, the knee and the hip center were tracked simultaneously. A rise in both coordinates meant a jump and a fall in their values signified a bend. Angular measures were utilized for actions that involved using the triggers. A mild angular change specified the use of a nearby trigger and a large angular change activated a faraway trigger. This technique helped in specifying a maximum of 12 triggers which were activated when the joint remained at that trigger for a period of more than 2 seconds. It was a simple and intuitive method which could be programmed to perform more complex tasks. For actions that involved joining hands together, the proximity of the two points were calculated. When the distance between them fell below the threshold value, the gesture was recorded as having been performed. These methodologies ensured a satisfying level of accuracy was achieved.

Results and Discussion

The tracking tool built using the proposed method was used to map gestures to alphabets.

The Fig.5 represents how gestures are mapped to alphabets. The alphabet A involved turning the shoulders to the left. When this action was done, the left was completely obscured by the upper body. Hence, this action was chosen. The same is the case for B; but in the opposite direction. For the alphabets C and D, there was a change in the coordinates of all the joints when the subject moves forward or backward. Hence, the accuracy was fairly good and was not ambiguous.

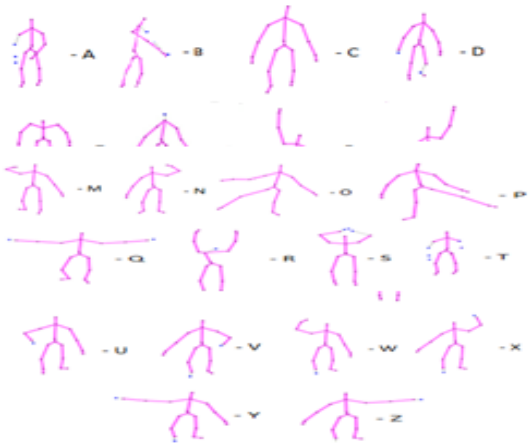

Fig 5. Mapping gestures to alphabets

E and F required tilting the head forward and backward. The difference in directions were relatively easy to identify. When there was a rise in coordinate values followed by a fall, the tilting was backward. For forward tilting, there is only a fall in the values. G and H were relatively easier since it involved only tracing the arm coordinates. Sidestepping left or right showed large changes in X coordinates and relatively less changes in Y coordinates. This comparison helped in differentiating it from jump.

I and J were mapped to sidestepping. A knee bend and a jump were mapped to K and L respectively. These gestures involved monitoring the changes in the Y coordinates. M and N were tracked by measuring angular variations between the elbow and the hand. O and P tracked the changes in the coordinates of the leg joints. Since the leg movements were performed with a straight leg, the coordinate changes of all the leg joints were proportional.

Q and R were identified and mapped when all the hand joints were on a straight line. The variations between the horizontal raise and vertical raise were differentiated by measuring the angular variations. S and T involved measuring the proximity of the joints to each other. The rest were mapped to triggers. The entire technique had an overall accuracy of close to 90 percentage when tested on 12 different test subjects of varying physical structure.

Mapping gestures to words

To proceed further, a dynamic system that identified a gesture as a whole was designed. This was then followed by mapping 10 gestures to complete words. The gestures chosen were arm rotations, one at a time, and both at a time and variations in direction. Some of the other gestures include a

jump followed by a knee bend and vice-versa. The results were encouraging and a success rate of 84.5 percentage was attained when the tests were performed on 12 different subjects. Using a text to speech converter the gestures were converted to speech. This thus produced a translator system which was capable of producing simultaneous speech simulations for the gestures that are tracked by the Kinect. This work requires a complete knowledge of the sign language. Since the amount of gestures were abundant, this is currently a work in progress.

Conclusion

Most of the works which are in practice or being in progress focuses on just mapping gestures to actions. The main objective of the so-proposed system is that by coalescing hand gesture recognition with skeletal tracking we are defining a whole new possibility there-by paving way for new techniques to be nourished. The system also aids in manipulating greater number of gestured as compared to regular hand tracking methods and also equally helps in overcoming the prospect of perplexing hand gestures.

The only biggest challenge however lies in misjudging each action which would be made. And as the process tends to be continuous, a lack of divider to identify the pauses and break of each word makes this system less vulnerable to making this system from being produced on a practical notion.

References

[1] Mauro dos Santos Anjo, Ednaldo Brigante Pizzolato, Sebastian Feuerstack "A Real-Time System to Recognize Static Gestures of Brazilian Sign Language (Libras) alphabet using Kinect", IHC 2012 proceedings.

[2] K. K. Biswas and Saurav Kumar Basu "Gesture Recognition using Microsoft Kinect", 2012, unpublished.

[3] Yi Li "Hand Recognition using Kinect", 2012, unpublished.

[4] Belinda Lange, Skip Rizzo, Chien Yen Chang, Evan A. Suma, Mark "Markerless Full Body Tracking: Depth-•Sensing Technology within Virtual Environments", Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), 2011.

[5] Mohamed Alsheakhali, Ahmed Skaik, Mohammed Aldahdouh and Mahmoud Alhelou "Hand Gesture Recognition System", 2011, Intelligent Approaches to interact with Machines proceedings.

[6] Ankit Chaudhary, J. L. Raheja, Karen Das, Sonia Raheja, "Using Hand Gesture Recognition in Natural way: A Survey", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.1, Feb 2011.

[7] Archana S. Ghotkar and Gajanan K. Kharate, "Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction", International Journal of Human Computer Interaction (IJHCI), Volume (3): Issue (1), 2012.

[8] Xi Chen and Markus Koskela, "Skeleton-Based Action Recognition with Extreme Learning Machines", ELM2013 23 October 2013.

[9] Chanjira Sinthanayothin, Nonlapas Wongwaen and Wisarut Bholsithi, "Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene", International Journal of Advancements in Computing Technology(IJACT), Volume4, Number11, June 2012.

[10] Georgios Th. Papadopoulos, Apostolos Axenopoulos and Petros Daras, "Real-time Skeleton-tracking-based Human, Action Recognition Using Kinect Data", 2013.

[11] Abhishek Kar, "Skeletal Tracking using Microsoft Kinect", 2013, unpublished.

[12] Najeefa Nikhat Choudhury and Golam Kayas,"Automatic Recognition of Bangla Sign Language",2012, unpublished.

[13] G Tao, PS Archambault and MF Levin, "Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis", Virtual Rehabilitation (ICVR), 2013.

[14] Alessandro De Mauro, "Virtual Reality Based Rehabilitation and Game Technology", EICS4Med, 2011.

[15] Jiann-Der Lee, Chung-Hung Hsieh and Ting-Yang Lin, "A Kinect-based Tai Chi exercises evaluation system for physical rehabilitation", Consumer Electronics (ICCE), 2014.

[16] Bengalur, "Human activity recognition using body pose features and support vector machine", Advances in Computing, Communications and Informatics (ICACCI), 2013.

[17] Helten T. Muller M. Seidel H. P. and Theobalt, C., "Real-Time Body Tracking with One Depth Camera and Inertial Sensors", Computer Vision (ICCV), 2013.