Real-Time Detection and Tracking with Kinect

Saket Warade, Jagannath Aghav, Petitpierre Claude, Sandeep Udayagiri

***Abstract*—**In this paper, we have proposed a system to keep track of human body movements in real time mode. The Kinect sensors are used to capture Depth and Audio streams. The system is designed by integration of two modules namely Kinect Module and Augmented Reality module. The kinect module performs Voice Recognition and captures depth images that are used by Augmented Reality module for computing the distance parameters. Augmented Reality module also captures real-time image data streams from high resolution camera. The system generates 3D module that is superimposed on real time data.

***Keywords*—**Kinect, Augmented Reality, Depth Imaging, Gesture Recognition.

I. Introduction

TRACKING and evaluating performance of a mechanic has been a very important issue in order to automate Maintenance and Repair work. Earlier, to perform such activity, expensive cameras and software have to be used. The sensors including RGB and depth cameras and audio capture system alone costs around $30,000.This provided a constraint over the research works in this domain. However with the launch of Kinect in last year, the same sensors bundled together are now available in just $150. Originally code named as 'Project Natal'[[17][19]] this amazing device provides 'Controller free' experience to users. Now, all of a sudden, the developers and designers are provided with a small, portable device capable of sensing with high accuracy rate. Using the Depth cameras installed on front panel, this device can get the much required 'z' parameter to calculate the distance between object and the sensor. The proposed system uses this 'z' parameter to keep track of the object in observation. The 'Identity Tracking Technique' [2] used by Kinect provides more accurate information in identification and Tracking.

Microsoft resolved the issue of expensive software requirements. They released a 'Software Development Kit' for Kinect based application development. This opened doors for developers to try their hand on this amazing device. It can be used to control systems such as TV, Radio etc. Even internet can be browsed with the help of spoken commands and gesture recognition. However Microsoft has been cleaver enough not to reveal the code and the algorithms used to track the body and face recognition. So if you are a developer, you are only provided with a standard set of APIs. Still it opens up lots of possibilities and thus created huge excitement in the world of hackers and developers.

This paper is organized as follows: Section 2 gives brief information about previous work on motion capture, depth sensing and augmented reality. Also few Kinect Identification techniques are also discussed. Section 3 throws light on Augmented Reality System design and Architecture. Also, the components and barriers for Augmented Reality system are discussed. Section 4 explains how the proposed system takes advantage of Kinect sensors and Augmented Reality to track the movements of mechanic working in workshop.

II. Literature Survey

Motion capture and depth sensing are two emerging areas of research in recent years. With the launch of Kinect in 2010, Microsoft opened doors for researchers to develop, test and optimize the algorithms for these two areas. J Shotton [1] proposed a method to quickly and accurately predict 3D positions of the body joints without using any temporal data. Key prospect of the method is they are considering a single depth image and are using a object recognition approach. From a single input depth image, they inferred a per pixel body part distribution.

Leyvand T [2] discussed about the Kinect technology. His work throws light on how the Identity of a person is tracked by the Kinect for XBox 360 sensor. Also a bit of information about how the changes are happening in the technology over the time is presented. With the launch of Kinect, Leyvand T expects a sea change in the identification and tracking techniques. The authors discussed the possible challenges over the next few years in the domain of gaming and Kinect sensor identification and tracking. Kinect identification is done by two ways: Biometric sign-in and session tracking. They considered the face that players do not change their cloths or rearrange their hairstyle but they do change their facial expressions, gives different poses etc. He considers the biggest challenge in success of Kinect is the accuracy factor, both in terms of measuring and regressing.

Jamie Shotton [1] took the advantage of the depth images for human pose recognition. The pixels in depth images indicate the depth in the image data and not any intensity or color information. This helps in calculating the 'z' (depth) parameter. They labelled the body parts according to body part position with respect to the camera. The body is

Saket Warade, Pusruing MTech(Computer Engineering), College of Engineering, Pune 411 014

Jagannath Aghav , Professor, Department of Computer Engineering and Information Technology, College of Engineering, Pune 411 014

Claude Petitpierre, Director, Computer Networking Laboratory, School of Computer and Communication Science, Swiss Federal Institute of Technology in Lausanne, CH-1015 Lausanne EPFL, Switzerland

Sandeep Udayagiri, Research Analyst , John Deere Technology Center India, Pune.

recognized as a set of 31 different labelled parts. They have recognized the body in the set of 31 different labelled parts. Machine Learning is performed by using classification techniques. With decision trees and forests, training is provided to machine.

Depth imaging refers to calculating depth of every pixel along with RGB image data. The Kinect sensor provides real-time depth data in isochronous mode[18]. Thus in order to track the movement correctly, every depth stream must be processed. Depth camera provides a lot of advantages over traditional camera. It can work in low light and is color invariant [1] The depth sensing can be performed either via time-of-flight laser sensing or structured light patterns combined with stereo sensing [9].The proposed system uses the stereo sensing technique provided by PrimeSense [21]. Kinect depth sensing works in real-time with greater accuracy than any other currently available depth sensing camera. The Kinect depth sensing camera uses laser beam to predict the distance between object and sensor. The technology behind this system is that the CMOS image sensor is directly connected to Socket-on-chip [21]. Also, a sophisticated deciphering algorithm (not released by PrimeSense) is used to decipher the input depth data. The limitations for depth cameras are discussed by Henry[9].

A motion capture system is a sensors-and-computers system that recovers and produces three-dimensional (3-D) models [7] of a person in motion. It is used in military[3], entertainment, sports etc. for validation purpose. In motion capture sessions, movements of one or more actors are sampled many times per second, although with most techniques motion capture records only the movements of the actor, not his/her visual appearance. To capture motion of a person, the very first step is to check identify the person. The motion capture algorithm given by J Gall [7] is very efficient for capturing and processing motion of an object. It has observed that the segment parameters of human body are indispensable to compute motion dynamic which causes inaccuracies. Kinect device used in our system is powered by both hardware and software. It does two things: generate a three-dimensional (moving) image of the objects in its field of view and recognize humans among those objects [16].

Christan Plagemann and Varun Ganpathi[11] proposed a feasible solution for identification and localization of human body parts while dealing with depth images. The greatest advantage of this method is that the output can be used directly to infer the human gestures. It can also be used to study and test other different algorithms which involve the detection of human body parts and depth images. The system identifies and localizes the body parts into 3D space. To obtain the results the machine is provided training data and classification technique is used to differentiate between two body parts and also between body part and other similar objects. The test results show that the system is able to identify body parts in different conditions and in different locations.

Henderson and Feiner [3] explored Augmented Reality system design for Maintenance and Repair work. They provided a state-of-the-art prototype that supports military mechanics conducting routine maintenance tasks inside a armoured vehicle, turret. They provided the interaction between system and the mechanic via Augmented Reality concept. The mechanic wears a special type a display glasses which are used to display the instructions. The system is controlled by a wrist-worn and-held device running on Android. An android application is written using open source Android APIs and Android SDK released by Google early in 2009. The application provides five forms of augmented reality content to assist mechanic. The content includes 3D and 2D arrows, Text instructions, labels, a close up view and 3D model of tools (e.g. a screwdriver). The arrows are used in such a fashion that it becomes denser when the mechanic is moving towards required tool and becomes fader if he is moving away from it. A small animation plays when mechanic reaches to the tool and arrow disappears.

III. Augmented Reality System Architecture and Design

Augmented Reality is the concept of implementing virtual Reality that duplicates the real world's environment. The system is able to provide the combination of real scene as observed by user and a virtual scene that is generated by the user. This superimposition of one view over the other provides a lots of comfort too users. They can visualize the things that are not there in existence but are very much relevant to understand the concept. The virtual scene generated by the computer is the representation of Augmented Reality concept. The success of any Augmented Reality system lies in the way such that the user should not be able to differentiate between real world scene and the one which is generated by the computer. Augmented Reality concept is used in many fields like engineering, entertainment, military training, manufacturing etc. Implementing Augmented Reality consists of generating 3D graphics on a plain surface called marker. The surface could be a paper, a file or even your hand. According to Henderson and Feiner [4] if Augmented Reality is applied in Maintenance and Repair work, this could automate the whole process and a lots of resources (e.g. Paper) can be saved. If the maintenance and repair work is connected with Augmented Reality then it saves a lot of time for a mechanic to perform the tasks. All the information he needs is directly available in front of him on the display provided thus he does not have to switch the focus of attention between the parts and the documentation.

Azuma [6] formally defined the Augmented Reality system as a system that "supplements the real world with virtual (computer generated) objects that appear to co-exist in the same space as the real world". The Augmented Reality systems are of two types, fixed and mobile[8]. The fixed AR systems are non-movable and have a attached graphics card to process input data. Mobile systems can be moved from one place to other. According to the need, the system design needs to be chosen. Our system is an approach to take advantage of fixed AR system in a mobile AR system.

A. Augmented Reality System Components

The Augmented Reality system consists of input devices, a

CPU, tracking devices and a display. Each of the devices along with their functionalities is discussed in this section.

1) Input Devices: The input devices are the most important part of an AR system. They are chosen according to the type of AR application. For fixed AR system, they consist of high definition cameras to capture input data. Whereas mobile AR system uses movable/detachable input devices .The proposed system uses Depth and Audio data obtained from Kinect sensors.

2) Display Devices: An Augmented Reality system can be designed by using 3 types of displays namely Head Mounted displays, Hand held displays and Spatial displays. The head mounted displays enables users to see both real-time images and superimposed 3D images on the screen. These displays are fixed near head of user, hence the name Head Mounted displays. Hand-held displays are the displays attached to a Hand-held device like iPhone/iPad or mobile device running on Android. They typically use the camera in hand-held device to get input data. The Spatial displays make use of projectors and other tracking devices display graphical data. Our system uses Head-Mounted displays. The mechanic will wear specially designed goggles to get input instructions from AR system.

3) Tracking Devices: These are the devices used to track the system. They consists of GPS, Optical camera, WiFi, Accelometer, RFID etc. The proposed system uses Kinect system for tracking purpose. The depth camera in Kinect system module takes care of this operation.

4) CPU: A powerful CPU is needed in order to process the data in any Augmented Reality system. There are also some minimum RAM requirements for the system to work properly. High amount of RAM is needed to process the camera data and to generate 3D model in real-time. Mobile AR systems can process the data by attaching a smart phone or laptop whereas the fixed AR systems use CPU supported by powerful graphics card. Our system works in fixed environment and hence we use a high end intel i5 Processor with nVidia Quadro FX 1700 graphics card. The system is also provided with 4 GB of main memory.

IV. Proposed System Design

Our system implements Augmented Reality using processing capabilities of Kinect. The system consists of 4 major components as Tracking Device, Processing Device, Input Device and Display Device. We use Kinect as a Tracking device as shown in figure 1. It contains three sensors for processing of depth images, RGB images and voice. Depth camera and Multi-Array Mic of Kinect are used to capture Real-Time image stream and audio data respectively. Depth sensor is used to obtain the distance between sensor and tracking object. The input device to our set-up is a high definition camera which is used to get input image stream and run as the background to all Augmented Reality components. On this background stream, we superimpose event specific 3D models to provide virtual reality experience. The processing Device, consisting of Data Processing Unit, Audio Unit and software associated with it takes care of which model to superimpose at which time. Processing Unit passes the input video stream and the 3D model to display device for visualization purpose.

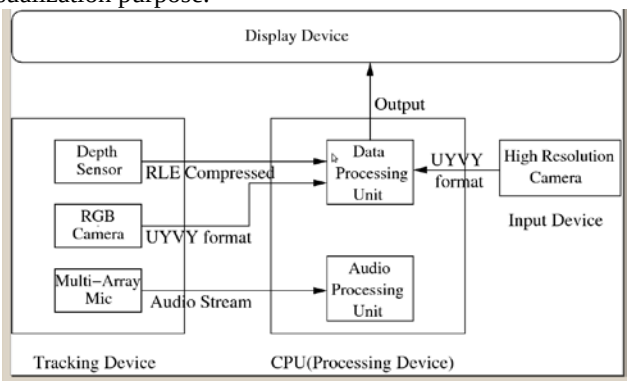

Fig. 1 Augmented Reality system

The proposed system tracks the movements of mechanic by processing the skeleton data. First, the body joints are identified and later, bones are drawn by joining appropriate joints. Using Kinect we identify 20 joints of human body and track their positions. We detect a motion by considering the difference between two consecutive frames. We identify the motion in a particular direction by taking difference between particular direction parameter(x, y or z).

The system is controlled with the help of audio commands as well as gesture inputs. Table 1 gives list of audio commands and their details. The system flow is given in System Flow:

System Flow

1: **Start**
2: Identify Position of operation
3: Locate mechanic in specified area
4: Guide mechanic to reach to position of operation by audio commands
5: Make sure that mechanic is ready
6: Display (Visualize) next Instruction
7: Wait for signal from mechanic
8: Require More Details?
9: Run pre-captured animation
10: goto Step 6
11: Repeat Instructions?
12: Repeat Instruction
13: goto Step 6
14: Done?
15: goto Step 516: Repeat 6-9 until all activities are performed
17: Verify the result
18: **Stop**

While performing maintenance work, if the mechanic is going out of range of the specified area, the system sounds an alarm. This enables the supervisor to check whether mechanic is moving out without completing the allocated work.

The system checks whether the mechanic is performing with correct tool or not. Since co-ordinates of every tool is fixed, We can obtain the difference of two depth images, one taken before start of operation and the other one while performing. The difference data is then compared with the shape of the tool the mechanic is supposed to use. If both the shapes are found to be equal, we conclude that correct tool is selected. If correct tool is selected, the mechanic is notified by green signal by the system. If wrong tool is selected, he is notified by Red signal. When the mechanic signals the system, he will visualize the next step on the display provided. The system understands the signals with the help of pre-defined gestures and audio commands. When the mechanic says "Next Command" and waves his right hand from rightmost position to leftmost position, the Kinect system understands that the user wants to move on to next instruction. All the visual effects are processed by AR system. Once all plug-ins are loaded, AR system adds particular event specific model to the screen-graph and provides Virtual Reality experience. After that, according to signal received, the event specific model is loaded and unloaded.

There are pre captured sessions involving the experts for every maintenance and repair activity. Animations or light-weighted (compressed) videos are prepared according to experts actions performed in centralized workshops. These Animations/Videos are played if mechanic chooses for "More Details" option. The system also "Repeat" the instructions if mechanic wants to visualize the step information again. Also the mechanic can visualize "Previous Instructions" if he wants to cross check the work done. The system keeps track of this movement and marks that activity as "Completed" or "Current" or "Yet to start".

Motion or movement is detected by considering difference between two frames. The Kinect system is very efficient in tracking skeletal of human body. The tracking is done by identifying different body parts and Joints. For the tracking purpose, the Kinect system considers the fact that human body is capable of giving enormous range of poses.

V. Conclusion

In this paper, we have discussed how the Kinect sensor is used for Detection and Tracking. We are using Kinect as a tracking device as well as input device for Augmented Reality System. Our work is a step towards automation of maintenance and repair activities for Tractors and other vehicles. The proposed system helps reduce the burden on experts to look into few regular activities. Instead, they can use our system for such activities. Also, the work simplifies the documentation process. The supervisor can keep track of current status of activity from his desk. Also, stepwise verification is possible as the system keeps track of each step. Through the introduction of our system, we will bring new opportunities for mechanical engineering based companies to use Augmented Reality for simplification of their complex tasks. This will add new dimensions to the conventional way of maintenance and Repair activities.

References

[1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1297 –1304, June 2011.

[2] T. Leyvand, C. Meekhof, Y.-C. Wei, J. Sun, and B. Guo. Kinect identity: Technology and experience. Computer, 44(4):94 –96, April 2011.

[3] S. Henderson and S. Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. Visualization and Computer Graphics, IEEE Transactions on, 17(10):1355 –1368, Oct. 2011.

[4] S. Henderson and S. Feiner. Augmented reality for maintenance and repair (armar). Technical report, AFRL-RH-WP-TR-2007-0112, United States Air Force Research Lab,, Jul 2007.

[5] R. Azuma. Tracking requirements for augmented reality. Commun. ACM, 36:50–51, July 1993.

[6] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. Mac Intyre. Recent advances in augmented reality. Computer Graphics and Applications, IEEE, 21(6):34 –47, nov/dec 2001.

[7] J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn, and H.P. Seidel. Motion capture using joint skeleton tracking and surface estimation. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1746 –1753, june 2009.

[8] R. Grasset, A. Mulloni, M. Billinghurst, and D. Schmalstieg. Navigation Techniques in Augmented and Mixed Reality: Crossing the Virtuality Continuum. In B. Furht, editor, Handbook of Augmented Reality. Springer, 2011.

[9] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgbd mapping: Using depth cameras for dense 3d modeling of indoor environments. In In RGB-D: Advanced Reasoning with Depth Cameras Workshop in conjunction with RSS, 2010.

[10] T. Caudell and D. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on, volume ii, pages 659 –669 vol.2, jan 1992.

[11] C. Plagemann, V. Ganapathi, D. Koller, and S. Thrun. Real-time identification and localization of body parts from depth images. In Robotics and Automation (ICRA), 2010 IEEE International Conference on, pages 3108 –3113, May 2010.

[12] R. Y. Wang and J. Popovi´c. Real-time hand-tracking with a color glove. In ACM SIGGRAPH 2009 papers, SIGGRAPH '09, pages 63:1–63:8, New York, NY, USA, 2009. ACM.

[13] A. Webster, S. Feiner, B. Macintyre, W. Massie, and T. Krueger. Augmented reality in architectural construction, inspection, and renovation. In In Proc. ASCE Third Congress on Computing in Civil Engineering, pages 913–919, 1996.

[14] Kinect sdk and api developer resources & faq — kinect for windows http://kinectforwindows.org/resources/faq.aspx.

[15] Nokia research center, augmented reality and michael j fox—nokia conversations http://conversations.nokia.com/2010/04/06/nokia-research-center-augmented-reality-and-michael-j-fox/.

[16] T. Carmody. How motion detection works in xbox kinect http://gizmodo.com/5681078/how-motion-detection-works-in-xbox--inect, Nov 2011.

[17] D. Clayman. E3 2010: Project natal is "kinect" - xbox 360 news at ign http://xbox360.ign.com/articles/109/1096876p1.html, August 2011.

[18] Microsoft. Microsoft research kinect for windows sdk beta http://research.microsoft.com/en-us/collaboration/kinect-windows.aspx, Oct 2011.

[19] MicrosoftCorporation. 'kinect for xbox 360' is official name of microsoft's controller-free game device: Formerly called "project na tal", kinect was revealed sunday evening in a cirque du soleil performance on the eve of the electronic entertainment expo in losangeles. http://www.microsoft.com/presspass/features/2010/jun10/06-13kinectintroduced.mspx, Jun 2010.

[20] Pocket-Lint. Five biggest barriers for augmented reality-pocket-lint http://www.pocket-lint.com/news/38882/5-biggest-barriers-augmented-reality

[21] PrimeSense. http://www.primesense.com/

[22] Qualcomm-http://www.qualcomm.co.uk/products/augmented-reality.