Technological Progression and Procedures in Microsoft Kinect Sensor

Dr. N. Anandakrishnan
Assistant Professor
Department of Computer Science,
Providence College for women,
Coonoor-643104-Tamil Nadu.

Abstract: Kinect is a sensor technology capable of tracking any environmental phenomena by sensing it for identification. And it plays a unique and vital role in the study of identification as it recognizes every object using the sensor. However, Kinect sensors track and sense objects using their color and data with the enhancement of tracking various actions and postures. Improvisations are inhabited within the technology based on identity, digitalization, alpha channels, depth of color and sensors in Autism research. Major concepts of Kinect sensor and its enhancement features are surveyed in this paper.

***Keywords** -- Kinect, sensor, tracking, color data, depth data*

*****

I. INTRODUCTION

Multimedia is computer technology which incorporates all computer controlled integrations of all parts of media. In which data is passed and transmitted as information. Advancements in this field has brought the use of media in real time in a very large amount.

Interaction between humans and computers in a requirement and has very high effects. Kinect plays a vital role for this interaction using the study of human language understanding of signs and postures for communication.

In communication using Kinect human understanding is what required for interaction. Here, understanding the language means, the moves, signs, and postures and tracking each action as the media communication by the human body using the sensor of the Kinect. Understanding is a question of "what the human actually perform". It is necessary that both the system and the person understands what is essential before the answer is formulated or executed.

Kinect has been developed to overcome the simulative problems faced by the video cameras and sensors while tracking and in the process of interaction.

Kinect tracks and captures motions of any objects and humans using a sensor. These sensor implements three dimensional depth of data interpretation of all the moves and every action of the humans. Primarily Kinect saw its application and evolution in gaming industry as in figure 1 and 2. Has its efficiency and practicability was farther more wide it made its way into research, engineering, robotics and others [12].

Figure 1. Gaming using Kinect

Figure 2. Gaming using Kinect

The enhancements and impact it brought to the gaming industry covered a large area in and also apart from the games industry. Its efficiency, data depth and sustainability brought attention to researchers from various fields including Computer science, electronics, and robotic sensing

ISSN: 2454-4248
226 – 231

technology etc. due to its availability, its applications spread from autism research to multiple technology enhancements. Kinect sensor tracking depends on signs, gestures, moves and even tracking a part of the human body like face, hand or entire human body. In order to publish the best use of Kinect, the improvised and enhanced techniques are included such as enhancement in identification, digitalization, color and depth data recognition, segmentation, calibration, research and even more.

II. KINECT SENSOR

Kinect is builded on a software technology of a system that particulars in interpretation of any specific gestures. A touch or hand free control of the device using a projector, camera, or a microchip is incorporated in Kinect.

A depth sensor uses a monochrome complementary with a metal oxide semiconductor sensor. A dot pattern formed by deducing the geometry between the projector and the camera based on the distance covered. The values obtained from the Kinect sensor are at times inaccurate. Hence to affine transformation of depth values along with color data a recalibration technique is used as shown in figure 3

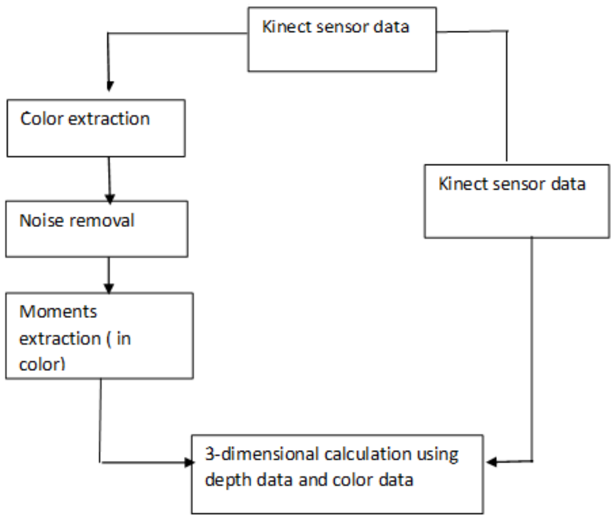


Figure2. Kinect sensing using color and depth data

Because of the calibration produced the depth values are inaccurate by the Kinect sensor. The recalibration toll providing the affine transformation of data is used to overcome it. RGB camera is used for recalibration given by the 3D coordinates which are true values in content. By minimizing the distance between two point classes the Kinect estimates the values and the transformations provided using recalibration.

TRACKING

The innovation now in hand progress is tracking (i.e.) the basic skeleton of objects or human beings that's in the path. Tracking is influenced by the sought of object and the numbers. It also configures the atmosphere and surrounding environment within which tracking is to be performed. Basics of tracking depend on the parameters like size, shape, color, motion, speed, and the scares effects produced for example furry objects.

Kinect considers all the parts of an object for identification. Thus, in human body tracking is detected for head, neck, shoulders, arms, etc. by their corresponding coordinate values of the edges, thus executing the complete skeleton tracking mechanism.

Kinect tracking techniques are enhanced and varied for applications across.

III. SOME TRACKING METHODS

A. *Recognition of sign languages*

Sign language is detecting the sign and translating it to the particular media. Tracking the signs produced by humans are of a great challenge for the system to understand. It can be recognized using a set if learning examples as in figure 4. The differences are studied and classifiers are used to formulate the most discriminative classes. For depth data Heuristic distance measures are enforced.

Sign language, which is an excellent mode of communication for deaf and dumb people. It is a composition of hand and arm movements with the clear understanding of facial expressions. Recognition of these sign language in system leads to transferring it to text and other modes of media possibility. The main components of this sign language are extracting the information and the exact model of signs along with their differences and similarities noticeably differentiated with appearance values.

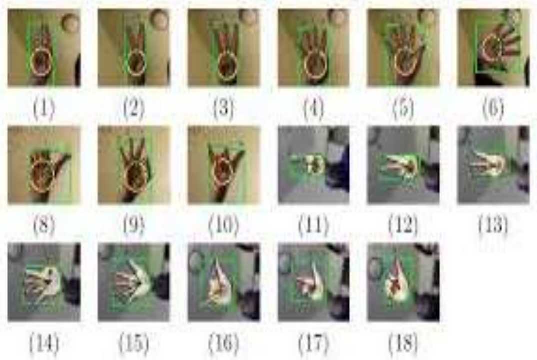


Figure 4. Sign language discrimination

All the frames when it comes to a video consists of a huge amount of difference for the same picture or a person [15]. But along with all the differences there exists some object key shapes from which they can be predicted. The key concepts are used as the representative object keys for the algorithm used for the recognition of the sign language.

B. *Hand Gestures*

Different Medias are involved with human computer interaction [1].Hand gesture is one amongst it. The performance of the hand gesture is highly affected by the environment, the background of the object and the color depths of all within the image in equal proportions as shown

 ISSN: 2454-4248
226 – 231

in figure 5. One possible way to avoid this is by using a different sensor that is not affected by lightning conditions and color [10]. Because of the low-resolution of Kinect depth map the ability to

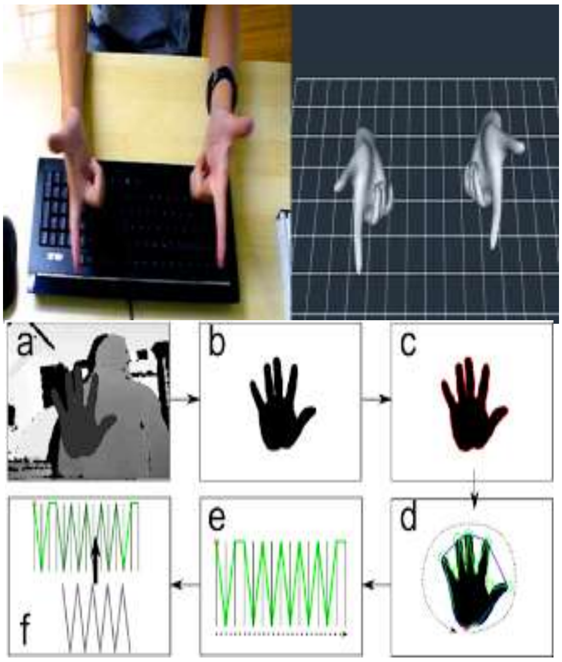


Figure 5. Hand gesture recognition

To recognize hand gesture is very low compared to the large objects. Kinect evolved a huge progress in tracking large objects within a range defined but it seems to produce limitations due to its size in robust hand recognition. Segmentation for these reduced size gestures is a difficult process. Dissimilarities are based on the distance measure using a new metric called Finger-Earth Mover's Distance that measure the dissimilarities within a finger gesture. Even the smallest of difference is calculated.

C. Face Recognition

A low resolution 3D sensor tracks the face under different atmospheric conditions with the use of pre-processing algorithm that obtains a canonical view and shape of the face as shown in figure 6. This algorithm is used mainly to smooth the noisy depth data given by the sensor. The depth map and texture of the shape is studied and the sparse value is approximated from training data. That texture is converted to a discriminate color before sparse coding.

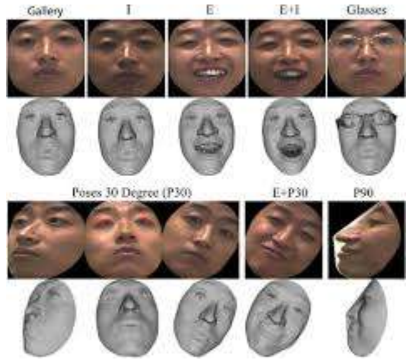
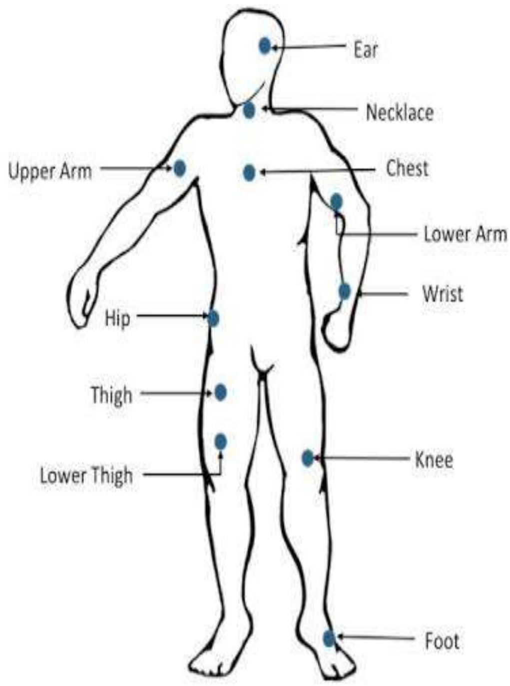


Figure 6. Face recognition using 3D sensor

Facial expressions are tracked while recognizing in order to apply in animation and in accordance with the human computer interaction. They are also influenced by the background and lightening. Applications with this facial recognition using 3D facial scanning for the purpose of morphing images produce high quality outcomes.

D. Full Human Body Recognition

In this part the entire human body is scanned using multiple sensors as shown in figure 7. Without interference and overlapping there are certain positions captured from a human body. Various parts are captured using a rough mess template that deforms the frame followed by a global alignment in order to distribute the errors, handling closure problems and misalignment [8].



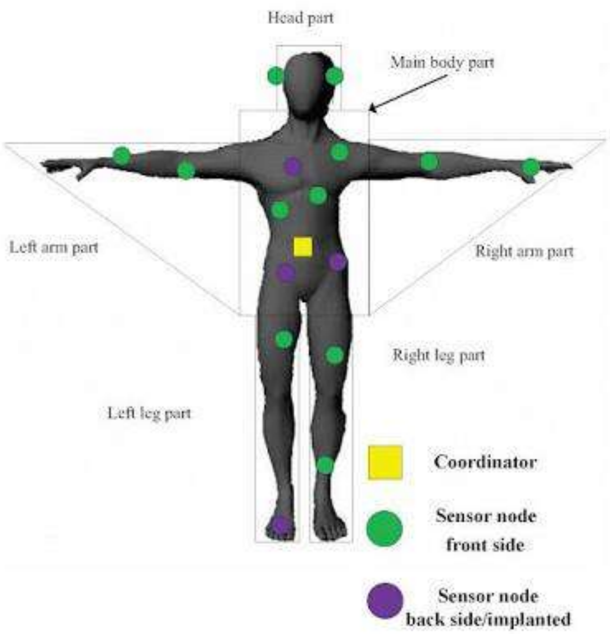


Figure 7. Human body recognition using multiple sensors

As the full body scanning is used a rough template of the scanning results is to be obtained from the recognized parts. The scan results produced from the sensor are compared to the recognition of the template. Thus the alignment process is totally an involvement of the occlusion process.

IV. ENHANCEMENT AND IMPROVISATION IN THE KINECT

A. *Kinect Identity*.

The Kinect sensor has two different techniques basically i.e. biometric sign in and the session play. In the biometric sign in technique the user or the object is identified every single time the appearance is made. In the session play technique the users appearance is recognized for a particular period of time and it can change which is majorly used in gaming [11].

This observed technique consists of the recognition of the face, color, clothes and the estimation of height which are given to each individual user. In the session play concept, the characteristics are valid for a mentioned session and beyond it would be invalid leaving way to next user session to begin. The face micro structure characteristics play an important role in the recognition process.

B. *Object Digitalization*

Kinect provides the capability to identity and recognize the 3D objects within the range of the sensor. A 3D model is created for the objects to add them as a character within the process using object digitalization, solving the problem of constructing a 3D model based on the inputs. Lasers are used to measure the overall surface and the orientation structure to focus the object. From the time taken for the retransmission of the light to the object is used as measurement [13]. The technique the models in real life are measured and digitalized to form a virtual real life model. This method is used prominently for people who think traditional methods are ineffective or costly to implement for the object recognition.

C. *Color and Depth Data*

In most of all concepts used in reconstruction and rendering the technique color and depth data information id prominent. The calibration of the cameras and the initialization of the calibration is achieved by bringing the color and depth data together. The display distortion correction applied in involves more accuracy ranges.

Scene reconstruction is an outcome for the calibration process of the system involved in relative calibration of the cameras and the calibration of each phenomena. Independent calibration does not produce optimal results whereas the combined calibration creates the required absolute output as it uses the information of the color and depth data [4]. Depending upon the camera quality, as certain cameras produce the depth and intensity at the same time which simplifies the calibration required other than some are not known with such characteristic so the image with original intensity is not obtained.

D. *Alpha Channels in Depth Maps*

Active or passive sensing methods are used in depth estimation. The disparity is calculated from two images of the active and passive methods. But the estimation of the image values is tedious process [7]. In certain technique detection is high but the noise and low resolution reduces the effectiveness of the technique used. In other words the saturated pixels are not detected. Usage of this passive and active sensing technology has its own limitations but it can be overcome by using the combination of both the sensing technology together forming a single alpha channel for sensing.

At present the depth map consists of the estimation of the single alpha channel including the advantages of using both the active and passive sensing technology. The degradation can be overcame through the use of single alpha channel in sensing. Here the depth map is created using the combination of color and alpha images, resulting in high clarity and naturalness.

E. *Autism research*

The Kinect based systems help in development of special assessments tools for the welfare and better communication of autism individuals. These autism research have increased over time to help them in better communication process. These researches .lead to the development of the avatar that is embedded in virtual scenes involving both the individual and a computer model involved [2].

The improvement is very less in helping such individuals as the assessment tools development and sensing is still in need. The possibility of an individual examiner to absorb the problems in less assessment can make better understanding. But this not the case in all circumstances, hence the entertainment technology that has enhanced and developed assessment tools for understanding, documenting and more realistically communicating.

F. Brain-Computer Interface

The real time brain-computer interface is used in control of the action to be performed by the person to the system. That particular action which the person is intentioned to be performed is decoded using the Kinect based vision system with the interface that controls the actions [3].

The most common impairment patients are affected is the stroke so the presence over the time marks the limitation. In this case, the patient is supported with a rehabilitation of robot. So this interface avoids the limitation created by the human support like the necessity of the person's presence at every time of process, so to repeat the action every time.

This robot provides with a high intensity motor for movements and actions for any task specified for the human body under the specific conditions. This helps the affected individual to perform movements by initiating it by the robot technology executed in replace of him. Impedance assistance, BCI based on the electroencephalographic that has the reality bases or robot assistance, EMG. Direct feedback with simulation etc. has been the various versions created for the improvement of this brain-computer interface.

G. Scene Segmentation

The process of identifying different parts in a particular scene describes scene identification. Scenes are characterized by the image concepts which have its own depth of understanding in it. Thus segmentation process in scene has various methods to undergo of the various implementations such as graph theory, clustering, region matching, and many others for better scene segmentation process [9].

The major limitation seen in the image given is not sufficient to represent the entire scene. Thus to overcome the complexity of scene segmentation the geometric data is used. Geometric data, main details of the image and the scene in combination are taken into consideration. To represent the color and the depth data multidimensional vectors are used and so the segmentation of the scenes are performed using spectral clustering in particular.

Presently both structured and unstructured scene segmentation is possible with tools that are developed. Those tools use both the geometric and the color values to obtain a fused framework of the scene. Mainly geometric data provides a lot of knowledge in partitioning and the features present in the image that helps in forming the scene.

H. Data Compression

The traditional methods involves the capturing and recognizing with the effective use of sensors while the scene is performed [5] [6]. Such scenes where digitalized and was used for other purposes such as matching and merging in data.

The essential need of transmitting such data after digitalization and the necessity to transfer a large collection of such data's at a particular time lead to the study of data compression. Changes where brought to the camera features to reduce the size of data but because of its bulky size and the cost involved in bringing the changes it was unachievable.

Kinect facilitated this possibility to meet the requirement of data compression with the depth sensors which was considered bulky with data. So it was given with the related applications implementation for easy transmission over the network.

Depth data in Kinect is of a very large data range. Its instability causes problems in inhabiting the compression process. Thus a bilateral filter is used to suppress the noisy data and padding is performed to avoid the invalid data calibration with coding while compressing the depth data of the Kinect sensor scenes. Once the data reaches the other end after transmission it is once again reconstructed.

I. Spinal Loading

In order to sort out the problem of low back disorder, trunk kinematics is used for the purpose of supporting biomechanics. The methods used in this requires instrumentation, thus resulting in the usage of videos, tracking with the use of single camera etc. but the manual operation in joint detection was their limitation and only a few frames were developed in every task proceeded. This affects the acceleration and finally the spinal loading [16].

The successful development of human skeleton extraction using the Depth sensors lead to the growth of tracking individual objects in the frame that is based on spinal loading. Kinect was used in biomechanical process for detecting the joints which was efficient and effective. And thus, Kinect is used for the assessment of the spinal loading. Here the joints in skeleton system are predicted and taken advantage of. The movements are captured as from the parts of a rigid body. And the measure of linear and angular acceleration for movement of each part is estimated using kinematics in Kinect.

J. Tele collaboration

Most of collaboration tolls where of symmetric view point example video conferencing etc. the participants of the conference have a face to face communication, making the process robust. But the symmetric view in number of collaborative tools was proven ineffective [17].

The study in the concept indicated that the lack of representing the entire 3D space. One such method is VR to represent the entire 43D data but it requires instrumentation. Another method was to use the beaming project that could be used in any of environment at the influence of limitations like misuse, transformations etc. now the use of depth camera and other assessment tools of Kinect have led to the use of asymmetric tool is that the telepresence system that allows the recognition of all the objects within the region as the 3D model enhances the capabilities of teleconferencing. It is still to be enhanced in order to support the multimodal representation of the source and destination.

K. *Out-of-Box Experience*

Kinect applications spread over wide range in usability. Thus to visualize the use if this product a digital out-of-box experience tool was developed along with the Kinect that served to be a huge improvisation for a number of digital devices that uses sensors. OOBE provided the steps and details for the usage of Kinect in various applications [14].

V. CONCLUSION

Thus from all the applications and uses of Kinect it plays a vital in tracking objects and recognition. The depth and color data sensors combine together to provide the best scene segmentation that can be framed for detection. Digitalization of every object tracked and the interfaces created to incorporate the high standard of human computer interaction is clearly maintained here. This paper thus is a clear study of Kinect, its applications and enhancements for understanding the importance, power, usability and Kinect at present generation.

REFERENCE

[1] Antonio Frisoli, Claudio Loconsole, Daniele Leonardis, Filippo Banno, Michele Barsotti, Carmelo Chisari and Massimo Bergamasco, “A New Gaze-BCI-Driven Control of an Upper: Applications and Reviews, Vol. 42, No. 6, November 2012, 1094-6977.

[2] Antony Steed, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, Jan Kautz, Daran Friedman, Angelika Peer, Massimiliano Soiazz, Franco Tecchia, Massimo Bergamasco and Mel Slater, “Beaming: An Asymmetric Telepresence System” IEEE Computer Society, November/December 2012, 0272-1716.

[3] Carlo Dal Mutto, Pietro Zanittigh, and Guido M.Cortelazzo, “Fusion of Geometry and Color Information For Scene Segmentation” IEEE Journal of Selected Topics in Signal Processing, Vol. 6, No. 5, September 2012, 193-4553.

[4] Chao Sun, Tianzhu Zhang, Bing-Kun Bao, Changsherg Xu and Tao Mei, “Discriminative Exemplar Coding for Sign Language recognition with Kinect” IEEE Transaction on Cybernetics, Vol. 43, No. 5, October 2013, 2168-67.

[5] Danial Herrera C., Juho Kannala, and Janne Heikkila, “Joint Distortion Correction” IEEE transactions and Pattern Analysis and Machine Intelligence, Vol. 34, No. 10, October 201, 0162-8828.

[6] Digan Um, Dongseok Ryu, and Myungjoon Kal, “Multiple Intensity Differentiation for 3D surface reconstruction with mono-Vision Infrared Proximity Array Sensor” IEEE sensor journal, Vol. 11, No. 12, December 2011 1530-434X.

[7] Dimitrios S. Alexiadis, Dimitrios Zarpalas, and Petro Daras, “Real Time, Full 3D Reconstruction of Moving Foreground Objects from Multiple Consumer Depth Cameras” IEEE Transaction on Multimedia, Vol. 15, No. , February 2013, 1520-9210.

[8] Jeff Munson and Phillip Pasqual, “Using Technology in autism research: The Promise and the Perils” IEEE Computer Society, June 2012, 0018-9162.

[9] Ji-Ho Cho, Kwan H, Lee, and Kiyoharu Aizawa, “Enhancement of Depth Maps with Alpha Channel Estimation for 3D Video” IEEE Journal of Selected Topics in signal Processing Vol. 6, No. 5, September 2012, 1932-4553.

[10] Jing Tong, Jin Zhonu, Ligang Liu, Zhigeng pan and Hao Yan, “Scanning 3D Full Human Body Using Kinects” IEEE Transactions on Visualizations and Computer Graphics, Vol. 8, No. 4, April 2012, 10777-2626.

[11] Jingjing Fu, Dan Miao, Weiren Yu, Shiqi Wang, Yan Lu, and Shipeng Li, “Kinect-Like Depth Data Compression” IEEE Transaction on Multimedia, Vol. 5, NO. 6, October 2013, 1520-9210.

[12] John Solaro, “The Kinect Digital Out of Box Experience” IEEE Computer Society, June, 2011, 0018-9162.

[13] Justin Clark “Object Digitalization for Everyone” IEEE Computer Society, October, 2011, 0018-916.

[14] Tommer Leyvand, Casey Meekhof, Yi-Chen Wei, Jian Sun and Baining Guo “Kinect Identity: Technology and Experience” IEEE Computer Society, 2011, 0018-9162.

[15] Wenjum-Zeng, “Microsoft Kinect Sensor and its Effect” IEEE Computer Society, 2012, 1070-986X.

[16] Xiaopeng Ning and Guodong Guo, “Accessing Spinal Loading using the Kinect Depth Sensor: A Feasibility Study” IEEE Sensors Journal, Vol. 13, No. 4, April 013, 1530-437X.

[17] Zhou Ren, Junsong Yuan, Jingjing Meng and Zheng Zhnag, “Robust Part-Based Hand Gesture Transactions on Multimedia, Vol. 5, No. 5, August 2013, 1520-910.