

# Providing The Natural User Interface(NUI) Through Kinect Sensor In Cloud Computing Environment

**Mr. Muneshwara M.S**  
*Assistant Professor*  
*Department of Computer Science & Engineering*  
*BMS Institute of Technology & Management*  
*Avalahalli, Yelahanka, Bangalore -560064*  
*Karnataka – India.*

**Mrs. Swetha M.S**  
*Assistant Professor*  
*Department of IS&E*  
*BMS Institute of Technology & Management*  
*Avalahalli, Yelahanka, Bangalore -560064*  
*Karnataka – India.*

**Mr. Anil G.N**  
*Associate Professor*  
*Department of Computer Science & Engineering*  
*BMS Institute of Technology & Management*  
*Avalahalli, Yelahanka, Bangalore -560064*  
*Karnataka – India.*

## Abstract

Cloud computing has continued to evolve and advance over the ensuing years. Cloud computing is the practice of using a network of With the Advancement of technologies, the low-cost Microsoft Kinect Sensor revolutionized the field of 3D Vision. Microsoft Kinect Sensors gives eyes, ears and brain to the computers by simple hand gesturing and speaking. The Microsoft Kinect Sensor has brought new era of Natural User Interface (NUI) based on gaming and the associated SDK provided access to its powerful sensors, which can be utilized especially in Research purposes. Thousands of people around the world are playing with built-in multimodal sensors, but still a complete kinect system lacks, thus requiring a physical device to fulfill its work. The Kinect Sensors recognizes each and individual users when they talks and what they speak. The information provided by the Kinect gears up new opportunity to fundamental problems in Computer Vision.The Kinect Sensors incorporates several advanced sensing hardware's. Most notably, it contains depth sensor, a color camera, and a four-microphone array that provides full-body 3D motion capture along with facial recognition, and voice recognition capabilities. The Kinect has robust 3D sensors for face recognition, using Microsoft Kinect sensors we can build an effective Rehabilitations system .Apart from the gaming applications, the Microsoft Kinect has lot of applications in all fields like clothing, medical imaging, used in many organizations for effective presentations. This innovation behind Kinect hinges on advances in skeletal tracking.

**Keywords: Kinect Sensor, Natural User Interface, Rehabilitations, Skeletal Tracking.**

## I. INTRODUCTION

Kinect is an RGB-D sensor providing synchronized color and depth images. It was initially used as an input device by Microsoft for the Xbox game console. With a 3-D human motion capturing algorithm, it enables interactions between users and a game without the need to touch a controller [9]. More specifically, the reviewed topics include object tracking and recognition, human activity analysis, The Kinect sensor lets the computer directly sense the third dimension (depth) of the players and the environment[2], making the task much easier. It also understands when users talk, identifies who they are when they walk up to it, and can track their movements and translate them to such a format that developers can use to build new experiences. Kinect's impact has moved on to other fields far beyond Gaming Industry. Kinect's wide availability, low cost, researchers and practitioners in computer science, and robotics are leveraging the sensing technology to develop creative new ways to interact with machines and to perform other tasks, from helping children to learn and assists doctors in operating rooms [10].

Recently, the computer vision society discovered that the depth sensing technology of Kinect could be extended far beyond gaming and at a much lower cost than traditional 3-D cameras such as (stereo cameras and time-of-flight).Additionally, the complementary nature of the depth and visual (RGB) information provided by Kinect bootstraps potential new solutions for classical problems in computer vision.



Fig. 1: Hardware Component of The Kinect Device

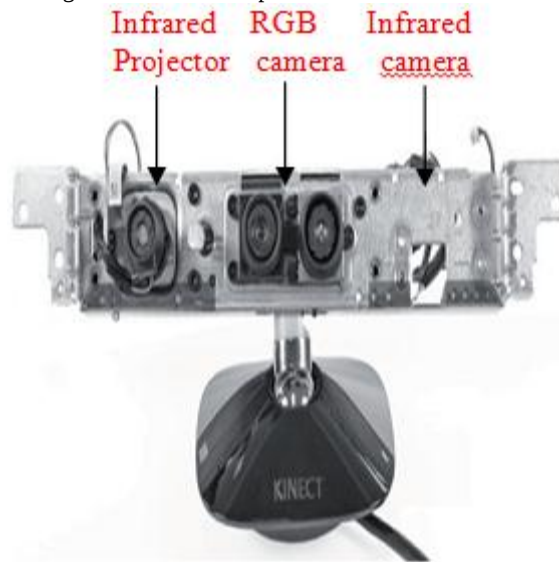


Fig. 2: The Camera

## II. DISCUSSION ON SYSTEM ARCHITECTURE & ITS CONSEQUENCES

Kinect architecture consists of 3 entities the Kinect sensor array, NUI library and Application. The following figure shows the Kinect architecture diagram and its actions [7]. The Sensor array sends all the streams of data such as Image stream, Depth stream data of images it receive along with audio stream. The NUI library contains all pre-defined hand gestures and also recognizes new gestures and finally applied on the application.

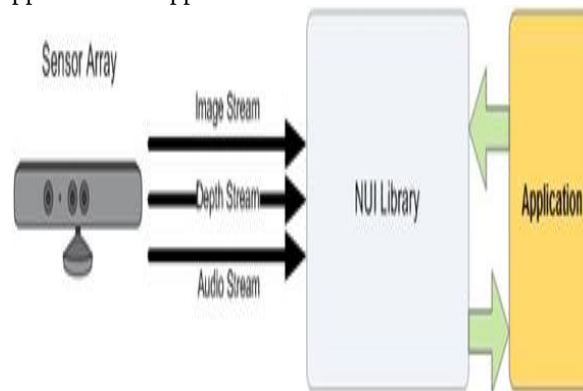


Fig. 3: Existing Architecture of Kinect

The architecture of Kinect Sensor comprises of 3 data streams and 3 data frames. The 3 data streams comprises of Color, Depth and Skeleton streams that are traced by the Kinect Sensor for any object. The Color stream gives the colorImage stream, the Depth stream gives the DepthImage stream and Skeletal stream gives the SkeletalImage stream of the traced image [7]. The 3 data frames gives the width or height, tracking mode, skeletal Array length, the pixel length and so on.

### A. Face Recognition

Facial recognition has been an effective and active research area in computer vision it has attracted many research interests in both security and surveillance [4]. Sometime facial recognition can be performed non-intrusively, without user's knowledge or explicit co-operation. However, Facial images captured in an uncontrolled environment can vary in poses, facial expressions, illuminations and disguise. Kinect Sensor allows tracking of facial expressions along with hand gestures using performance-driven facial animations. Kinect sensor also allows 3D facial scans by fitting morphable models [8]. The figure shown below tells how the facial recognition takes place in Kinect sensor.

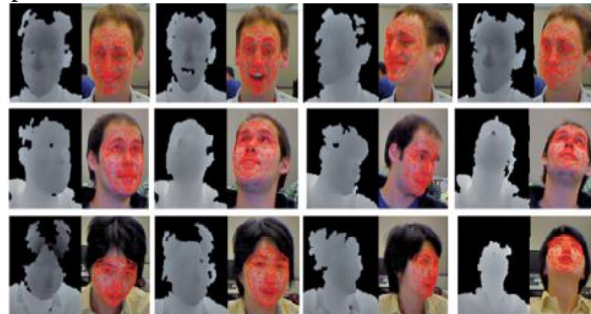


Fig. 4: The Facial Recognition And Tracking In Kinect Sensor

### B. Skeletal Tracking

The innovation behind kinect hinges on skeletal tracking. The skeletal tracking works identically for every human being without any kind of calibrations. In skeletal tracking, a human body is represented by a number of joints representing body parts such as head, neck, shoulders, and arms as shown in figure, and each joint is represented by its 3D coordinates[6].

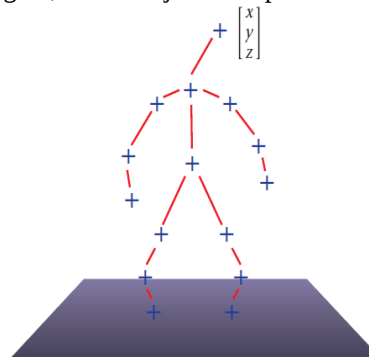


Fig. 5: Skeletal Tracking Joints

Skeletal Tracking allows Kinect to recognize people and follow their actions. Using the infrared (IR) camera [5], Kinect can recognize up to six users in the field of view of the sensor. Out of these, up to two users can be tracked in detail at a time. An application can locate the joints of the tracked users in space and track their movements over time [5].

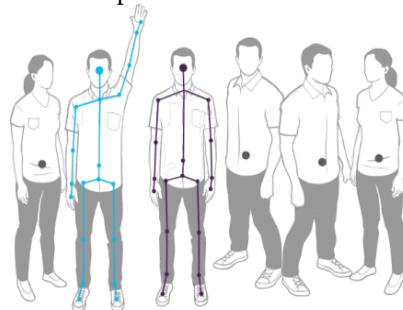


Fig. 6: Kinect Can Recognize Six People And Track Two

Skeletal Tracking is optimized because it recognizes users standing or sitting positions [9]. When facing the Kinect sideways, it poses some challenges regarding the part of the user that is not seen to the sensor.

To be recognized, users simply need to face the sensor, making sure that the sensor can track their head and upper body. No specific pose or actions needs for a user to get tracked [2].

### C. 3D Depth Camera.

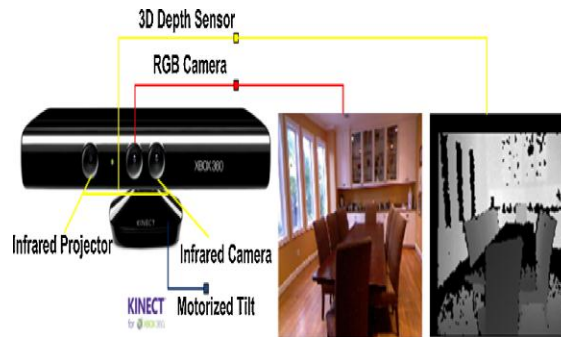


Fig. 7: Depth Image S Captured

Kinect sensor consists of an IR laser projector and an IR camera. Together, the projector and the camera create a depth map, which provides the distance information between an object and the camera. Figure 3 shows the depth map produced by the Kinect sensor for the IR image in Figure 2. The depth value is encoded with gray values; the darker the pixel, the closer the point is to the camera [8]. The black pixels indicate that no depth values are available for those pixels. This might happen sometimes if the points are too far (and the depth values cannot be computed accurately), are too close (there is a blind region due to limited fields of view for the projector and the camera), are in the cast shadow of the projector (there are no IR dots), or reflect poor IR lights (such as hairs or specular surfaces). The depth values produced by the Kinect sensor are sometimes inaccurate because the calibration between the IR projector and the IR camera becomes invalid. This could be because of heat or vibration during transportation or a drift in the IR laser [3].

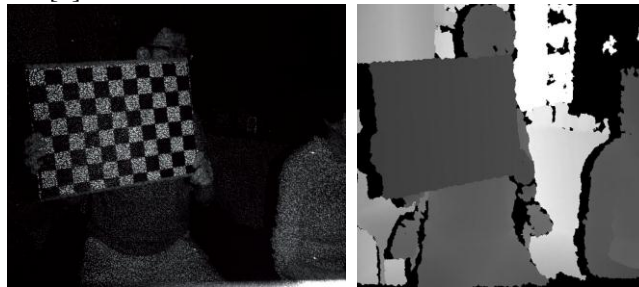


Fig. 8: The Depth Image Stream

### D. Hand Gesture Recognition

There is always a need to communicate using sign languages, such as chatting with speech and hearing challenged people. Additionally, there are situations when silent communication is preferred: for example, during an operation [9], a surgeon may gesture to the nurse for assistance. It is hard for most people who are not familiar with a sign language to communicate without an interpreter. Thus, software that transcribes symbols in sign languages into plain text can help with real time communication [8], and it also provides interactive training for people to learn more about sign language. Gesture recognition has become an important topic in research field with the current focus on interactive emotion recognition and HGR. Traditionally, gesture recognition requires high quality Stereoscopic cameras and complicated computer vision algorithms to recognize hand signals; the systems often turn out to be expensive and require extensive setup [10]. Microsoft Kinect provides an inexpensive and easy way for real-time user interaction [8]. Kinect, originally designed for gaming on the Xbox platform, uses a depth sensor to capture color (RGB) images and the associated depth (distance) data. It allows the algorithms that classify and perform recognition of the image data. Hand Gesture Recognition is an important research topic because some situations require silent communications with large sign languages. Computation Hand Gesture Recognition systems assist silent communications, and help people to learn sign languages [5]. Hand Gesture Recognition using Kinect provides way for Natural User Interface.

There are two different scenarios for recognition that is popular gesture with nine gestures, and the numbers with nine gestures [7]. The systems allows the users to select a scenario ,and it is able to detect hand gestures made by users ,to identify fingers, and to recognize meaning of gestures and also to display the meaning of pictures seen on screen [7]. Because of the depth sensor present in kinect which is an infrared Camera, the lighting conditions signer's skin colors and clothing .The background have little impact on performance of the system, The accuracy and Robustness provided by this system is a versatile component that can be integrated in a variety of applications in daily life [8].

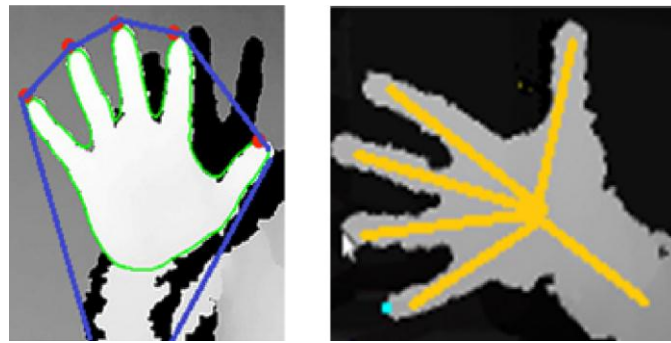


Fig. 9: Hand Gesture Recognition

### E. Hand Gesture Recognition System

The HGR system can be divided into three parts according to its processing steps: finger identification, gesture recognition and hand detection. The system has two advantages. First is, it is highly modularized [5], and each of the three steps is related from others; second, the edge detection of hand as well as gesture recognition is an add-on layer, which can be easily moved to other applications [8]. Depth data is generated and converted from the raw image data of Kinect sensor by an open-source framework called OpenNI (Natural Interaction), with an open-source driver called Sensor Kinect by Prime Sense, which makes Kinect for Xbox compatible with Microsoft Windows 7.

This system has several key features:

- Is Capable of capturing images in the dark.
- Identifying fingers of up to two different hands, under all reasonable and possible rotations of the Hands.
- displaying gestures and translating them in real time.
- Allowing user to choose different scenarios.

This system is able to accomplish its task in the dark because Kinect uses an infrared camera for depth image. In addition, as the frame rate for Kinect sensor output at about 30Hz, the process of gesture recognition can be considered as finished in real-time. A practical sensing range for Kinect is 1.2 – 3.5m when the raw data IS processed by the Xbox software. For the purpose of Hand Gesture Recognition, the hands have to be closer than that in order to resolve the details of fingers. Therefore, the effective range of detecting hands and gestures is set to be between 0.5m and 0.8m [3].

### F. Glove-based Gesture Recognition

At the finger-spelling level of Alphabet Source Language, several signs of letters are similar to each other [5]. Figure 1 shows the gestures of the ASL alphabet. As an example, letters 'A', 'E', '1', 'N', 'S', and 'I' are all formed by a closed fist with a little variation in thumb placements; as another example, 'K' and 'V' both use index finger and middle finger with the same angle, and the only difference is again in the thumb placements [4]. The overlaps of fingertips make gesture differentiation a difficult task for 2D video based recognition systems. Accurate data of each finger information will be needed. Therefore, glove-based sensing systems have been studied for decades to solve such problem. Although it seems inconvenient for users to wear extra for the purpose of recognition, glove-based systems do have the advantage to make up the 'cumbersomeness' by largely increasing the accuracy in finger-spelling recognition [7].



Fig. 10: Alphabet-Sign Language Symbols

### III. THE PROPOSED SYSTEM

The Kinect currently has fled its applications beyond the computer vision, In this paper we have proposed a system where in the Kinect sensor switched to a mode in which it can monitor the heart rate of a person standing in front of it using the color cameras to measure how flush the skin was and the infrared cameras to track blood flow underneath the skin. This could ostensibly allow a developer to determine whether a user was scared, or even lying, and could also have health monitoring implications and other diagnostics. When the blood flows inside body the skin tone color changes very slightly which is captured but the kinect depth Camera based on that images the sensor enables the heart beat of the user.

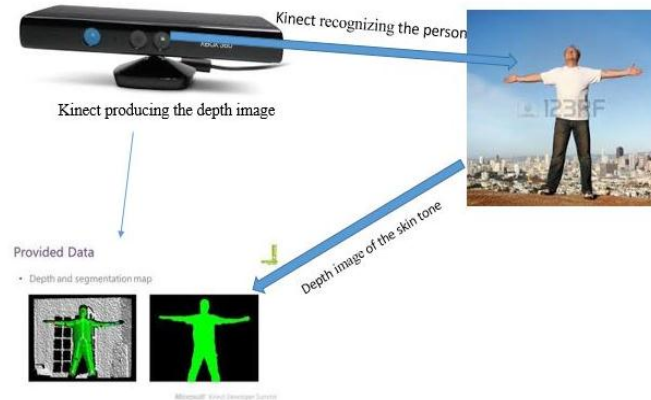


Fig. 11: Proposed System Architecture

The following figure shows an example how the depth images are captured from kinect sensor based on which the kinect can tract different facial expressions and heart rate of person standing before it.

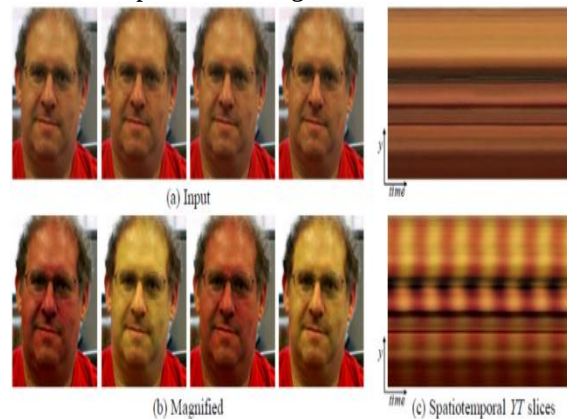


Fig. 12: Example of Depth Images Captured Displaying Skin Tone

### IV. CONCLUSION AND FUTURE IMPLEMENTATIONS

The Kinect sensor offers an unlimited number of opportunities for old and new applications. The dream of building a computer that can recognize and understand scenes like human has already brought many challenges for computer-vision researchers and engineers [3]. The emergence of Microsoft Kinect (both hardware and software) and subsequent research efforts have brought us closer to this goal [7]. We summarized the main methods that were explored for addressing various vision problems. The Kinect also includes topics like object tracking, Facial recognition, human activity analysis, hand gesture analysis, and also indoor 3-D mapping [6].

The future implementations of this paper can be researched in the fields of space where in the Kinect sensor can be used for Space researchers to carry out their experiments using unmanned machine that can be controlled using kinect sensor. Further its applications can also be implemented in Flight Simulators for effective Flight Control and for better navigation system from the depth images captured from the kinect depth Camera [9].

### V. ACKNOWLEDGEMENT

The authors would like to thank the editor and reviewers for their priceless suggestions that appreciably improved the eminence of this paper.

Also we thank to our colleagues who gave valuable inputs.

## REFERENCES

- [1] Roy, A.K. ; Dhirubhai Ambani Inst. of Inf. & Commun. Technol., Gandhinagar, India ; Soni, Y. ; Dubey, S. : “Enhancing effectiveness of motor rehabilitation using kinect motion sensing technology” , Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013 IEEE.
- [2] Kai-Wen Shih ; Dept. of Comput. Sci. & Inf. Eng., Nat. Central Univ., Chungli, Taiwan ; Chia-Jung Wu ; Gwo-Dong Chen: “Developing a Well-Focused Learning through a Kinect-Based Collaborative Setting” , Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference.
- [3] Leyvand, T. ; Meekhof, C. ; Yi-Chen Wei ; Jian Sun : “Kinect Identity: Technology and Experience” , Biometrics Compendium, IEEE 2011.
- [4] Jungong Han ; Civolution Technol., Eindhoven, Netherlands ; Ling Shao ; Dong Xu ; Shotton, J. : “Enhanced Computer Vision With Microsoft Kinect Sensor: A Review” , Cybernetics, IEEE Transactions on (Volume:43 , Issue: 5 ) 2013.
- [5] Islam, M.R. ; Rahaman, S. ; Hasan, R. ; Noel, R.R : “A Novel Approach for Constructing Emulator for Microsoft Kinect XBOX 360 Sensor in the .NET Platform” , Intelligent Systems Modelling & Simulation (ISMS), 2013 4th International Conference.
- [6] Z. Zhang, “A Flexible New Technique for Camera Calibration,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 11, 2000, pp. 1330\_1334.
- [7] J. Shotton et al., “Real-Time Human Pose Recognition in Parts from a Single Depth Image,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS Press, 2011, pp. 1297\_1304.
- [8] Q. Cai et al., “3D Deformable Face Tracking with a Commodity Depth Camera,” Proc. 11th European Conf. Computer Vision (ECCV), vol. III, Springer-Verlag, 2010, pp. 229\_242.
- [9] Maimone and H. Fuchs, “Encumbrance-Free Telepresence System with Real-Time 3D Capture and Display Using Commodity Depth Cameras,” Proc. IEEE Int’l Symp. Mixed and Augmented Reality.
- [10] (ISMAR), IEEE CS Press, 2011, pp. 137\_146.Majdi, A. ; Lab. Riadi, Univ. de Tunis El Manar, Ariana, Tunisia ; Bakkay, C. ; Zagrouba, E.” 3D modeling of indoor environments using Kinect sensor “ , Image Information Processing (ICIIP), 2013 IEEE Second International Conference, 2013.