

# Smart Photography using Kinect

Aishwarya C  
Department of ECE  
GSSSIETW  
Mysuru, India.

Manasa H K  
Department of ECE  
GSSSIETW  
Mysuru, India.

Spoorthi Y  
Asst. Professor  
Dept. of ECE  
GSSSIETW  
Mysuru

Manasa B K  
Department of ECE  
GSSSIETW  
Mysuru, India.

Nandini J  
Department of ECE  
GSSSIETW  
Mysuru, India.

**Abstract** - Kinect sensor is a peripheral device designed for Xbox and Windows PCs. It has key features such as full body 3D motion capture, gesture recognition, facial recognition and voice recognition capabilities. The timer is set using voice recognition feature. This is used to capture the photo on the camera by giving commands. We propose a new technology or capture of digital images. Facial recognition capability is employed so that the photo is captured only when people are smiling and their eyes are open. Also we make use of gesture recognition capability to set the timer.

**Keywords** - Face recognition, Face tracking, Image capture, Image resolution, Kinect sensor, Microsoft Visual Studio, Software Development Kit (SDK).

helping children with autism to assisting doctors in operating rooms. Microsoft calls this the Kinect Effect. On 1 February 2012, Microsoft released the Kinect Software Development Kit (SDK) for Windows, which will undoubtedly amplify the Kinect Effect. The SDK will potentially transform human-computer interaction in multiple industries-education, healthcare, retail, transportation and beyond.



Figure I.1 Microsoft Kinect sensor.

Recent advances in 3D depth cameras such as Microsoft Kinect sensors have created many opportunities for multimedia computing. Kinect was built to revolutionize the way people play games and how they experience entertainment. With Kinect, people are able to interact with the games with their body in a natural way. The key enabling technology is human body language understanding; the computer must first understand what a user is doing before it can respond. This has always been an active research field in computer vision, but it has proven formidably difficult with video cameras. The Kinect sensor lets the computer directly sense the third dimension (depth) of the players and the environment, making the task much easier. It also understands when users talk, knows who they are when they walk up to it and can interpret their movements and translate them into a format that developers can use to build new experiences.

Kinect's impact has extended far beyond the gaming industry. With its wide availability and low cost, many researchers and practitioners in computer science, electronic engineering and robotics are leveraging the sensing technology to develop creative new ways to interact with machines and to perform other tasks, from

Kinect was launched on 4 November 2010. A month later there were already nine pages containing brief descriptions of approximately 90 projects. The Kinect sensor incorporates several advanced sensing hardware. Most notably, it contains a depth sensor, a colour camera and a four-microphone array that provides full-body 3D motion capture, facial recognition and voice recognition capabilities.

In this project, the Kinect2 sensor is used to recognize facial attributes such as happy expression and activities such as eyes closed, looking away. Kinect acts as an eye to the robot. It consists of 1080p HD colour camera using which the photo can be captured. It has ability to track six people and twenty six skeletal joints per person. It has ability to track the voice of a person.

## II. EXISTING SYSTEM

Facial expression tracking has been an active research area in computer vision for decades. It has many applications including human computer interaction, performance-driven facial animation and facial recognition. A Kinect sensor produces both 2D color video and depth images. A linear deformable head model with a linear combination of a neutral face, a set of shape basis units with co-efficients that represent a particular person and a static over time, and a set of action basis units with co-efficients that represent a person's facial expression and are dynamic over time. A face cannot perform all facial expressions simultaneously, so it is believed that in general the set of co-efficients for the action basis units should be sparse.

## III. PROPOSED SYSTEM

### A. Kinect Sensor

It can detect and track faces of up to 6 persons. The detected face is represented by a rectangle. The position of the eyes, nose and mouth are also recognised.

Additionally, the following facial features/ attributes are determined:

- Happy : yes/ no/ unknown
- Engaged : yes/ no/ unknown
- Left eye closed : yes/ no
- Right eye closed : yes/ no

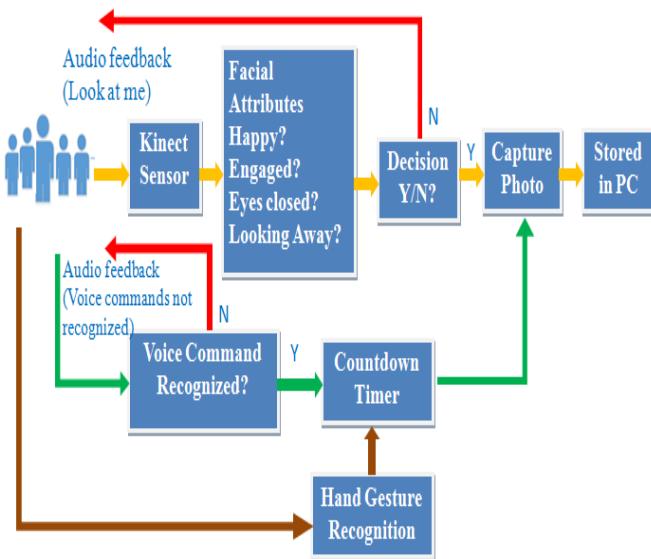


Figure A.1 Block diagram of the proposed Smart Photography System.

### B. Decision logic

Weights or scale factors are assigned to each of the facial attributes such as:

- Happy
- Engaged
- Left eye closed
- Right eye closed

$$\text{Decision logic : } \frac{\sum_{i=1}^4 \alpha_i A_i}{\sum_{i=1}^4 \alpha_i} \times 100 > \text{Threshold}$$

Where  $A_i$  represents the  $i^{th}$  facial attribute and  $\alpha_i$  are the corresponding scale factors

### C. Audio feedback

Play a pre-recorded audio file which will say “SMILE PLEASE/LOOK AT ME” whenever the decision logic fails.

### D. Timer

#### D.1 Voice command

The Kinect sensor is equipped with speech recognition. We can set the timer by using voice commands such as:

- START CAPTURE
- CAPTURE

#### D.2 Audio Feedback

Play a pre-recorded audio file which will say “VOICE COMMAND NOT RECOGNIZED” whenever the Kinect fails to recognize voice commands.

#### D.3 Hand gesture

The Kinect can also detect basic hand gestures such as:

- OPEN HAND
- CLOSED HAND
- LASSO (POINTER POSITION)

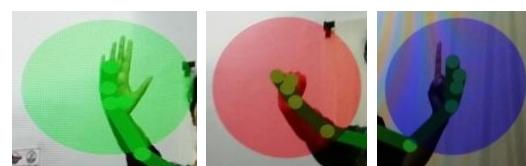


Figure D.3.1 The basic hand gestures.

Photo is captured by the 1080p HD wide-angle camera present in the Kinect. These pictures will be stored in the local directories of PC to which the Kinect is connected.

#### IV. EXPERIMENTAL RESULTS

- The photo is captured when people are smiling and their eyes are open.
- Timer is set using voice commands. It can also be set using hand gesture.
- Face detection
- Facial attributes or features recognition.
- C# code to detect voice commands is executed and timer is set using command “ START TIMER”.
- C# code to detect basic hand gesture is also executed and timer is set using the gesture “ LEFT HAND CLOSED AND RIGHT HAND OPEN ”.

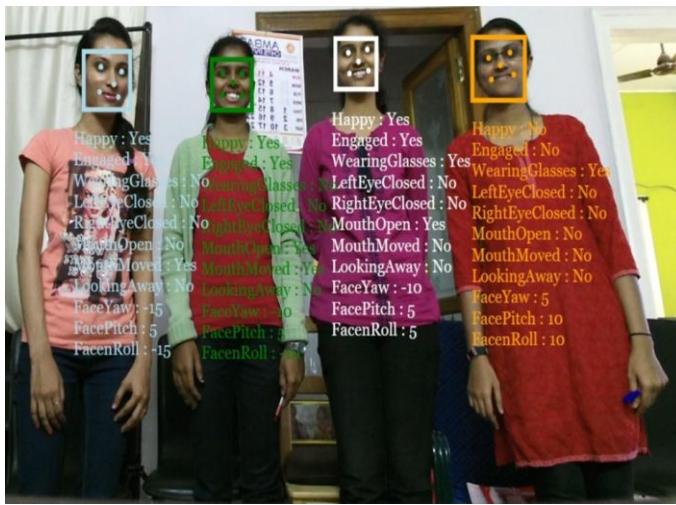


Fig IV.1: Kinect detecting all the facial attributes

In this figure the Kinect is detecting all the facial attributes and captures the photo. We modify this code accordingly as per our project needs.

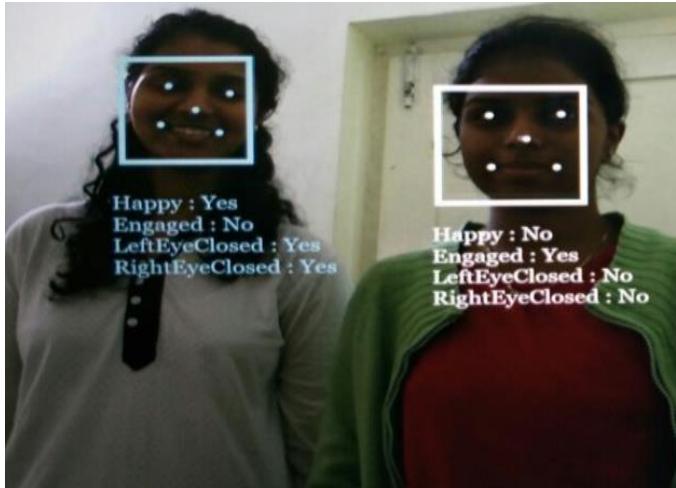


Fig IV.2: Kinect detecting only four attributes

In this figure the kinect detects only four facial attributes such as happy, engaged, right eye closed and left eye closed. Only if happy is yes, engaged is yes, right eye closed and left eye closed is no then the photo is captured.

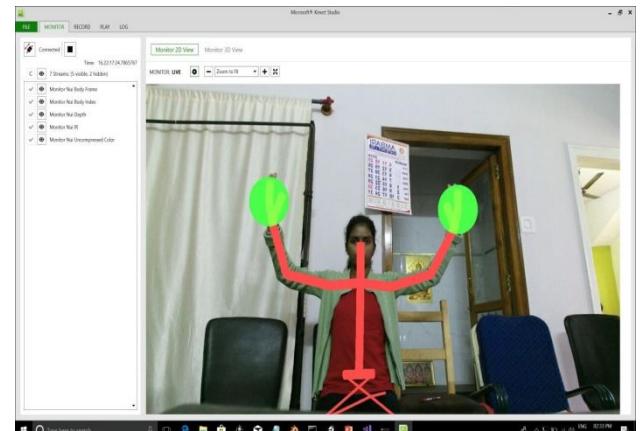


Fig IV.3 : Open hand gesture recognized by Kinect

In this figure, the open hand gesture is recognized by the kinect. This feature is used to set the timer using hand gesture. Combination of gestures are used in our project where RIGHT HAND is open.

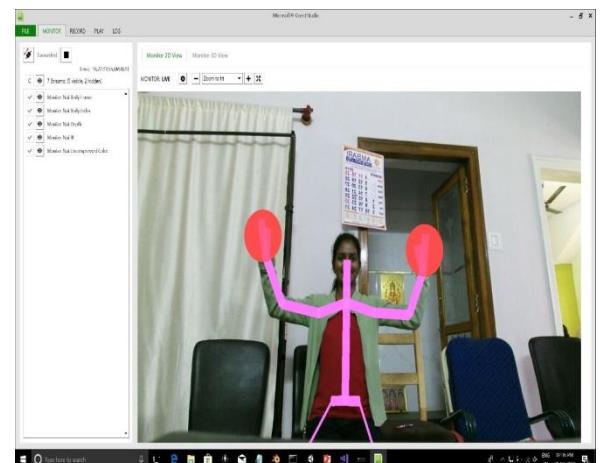


Fig IV.4 : Closed hand gesture recognized by Kinect

In this figure, closed hand gesture is recognized by the kinect. Combination of hand gestures is used to set the timer wherein LEFT HAND is closed.



Fig IV.5 : Hardware setup

In this figure, hardware setup of our project is seen. Kinect is connected to the PC. Kinect captures the photo based on different conditions and stores the captured photo in local directories of PC(different folder). The photo can be captured by using facial attributes, voice commands and hand gestures.



Fig IV.6 : Photo captured by kinect

In this figure, the photo captured by the kinect can be seen. Its of 1080p resolution captured in a HD camera of Kinect. Thus the photo can be captured using facial attributes, voice commands and hand gestures.

## V. CONCLUSION

We have programmed the Kinect to detect faces, Recognize Voice commands and basic hand gestures. The colour camera is triggered using the voice commands. The photo is captured when people are smiling; their eyes are open and are looking at the camera.

Thus far, additional research areas include body biometrics estimation (such as weight, gender, height etc.,), 3D surface reconstruction and healthcare applications.

## REFERENCES

- [1] Dardan Maraj, Arianit Maraj, Adhurim Hajzeraj, "Application interface for gesture recognition using Kinect sensor", 2017.
- [2] Z. Zhang, "Microsoft Kinect sensor and its effect", IEEE Multimedia Mag., vol. 19, no.2, pp.410, Feb. 2016.
- [3] A. Jalal, S. Lee, J. Kim, T. Kim, "Human activity recognition via the features of labelled depth body parts", Proc. Int. Conf. Smart Homes Health Telematics, pp. 246-249, 2012.
- [4] U. Lee, J. Tanaka, "Hand controller: Image manipulation interface using fingertips and palm tracking with Kinect depth data", Proc. Asia Pacific Conf. Compute. Human Interact, pp. 705-706, 2012.
- [5] H. Liang, J. Yuan, D. Thalmann, "3-D fingertip and palm tracking in depth image sequences", Proc. ACM Int. Conf. Multimedia, pp. 785-788, 2012.
- [6] M. Caputo, K. Denker, B. Dums, G. Umlauf, "3-D hand gesture recognition based on sensor fusion of commodity hardware", Proc. Conf. Mensch Compute., pp. 293-302, 2012.
- [7] Y. Li, "Hand gesture recognition using Kinect", Proc. Int. Conf. Software Eng. Service Sci., pp. 196-199, 2012.
- [8] S. S. Rautaray, A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey", Artificial Intelligence Review, 2012.
- [9] Han, E. Pauwels, P. de Zeeuw, and P. de With, "Employing an RGBD sensor for real-time tracking of humans across multiple re-entries in a smart environment," IEEE Trans. Consumer Electron., vol. 58, no. 2, pp. 255–263, May 2012.
- [10] Z. Ren, J. Meng, J. Yuan, Z. Zhang, "Robust hand gesture recognition with Kinect sensor", Proc. ACM Int. Conf. Multimedia, pp. 759-760, 2011.
- [11] Soumi Paul, Subhadip Basu, Mita Nasipuri, "Microsoft Kinect in Gesture Recognition" I J C T A, 8(5), 2015, pp. 2071-2076 © International Science Press.
- [12] Wongun Choi, Caroline Pantofaru, Silvio Savarese, "Detecting and Tracking People using an RGB-D Camera via Multiple Detector Fusion" Electrical and Computer Engineering, University of Michigan, Ann Arbor, USA, Willow Garage, Menlo Park, CA, USA.
- [13] I. Tashev, "Recent Advances in Human-Machine Interfaces for Gaming and Entertainment," Int'l J. Information Technology and Security, vol. 3, no. 3, 2011, pp. 69-76.
- [14] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 11, 2000, pp. 1330\_1334.
- [15] 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.
- [16] 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.
- [17] A. 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 (ISMAR), IEEE CS Press, 2011, pp. 137\_146.
- [18] Z. Ren, J. Yuan, and Z. Zhang, "Robust Hand Gesture Recognition Based on Finger-Earth Movers Distance with a Commodity Depth Camera," Proc. 19th ACM Int'l Conf. Multimedia (ACM MM), ACM Press, 2011, pp. 1093\_1096.
- [19] W. Li, Z. Zhang, and Z. Liu, "Action Recognition Based on A Bag of 3D Points," Proc. IEEE Int'l Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), IEEE CS Press, 2010, pp. 9\_14.
- [20] C. Velardo and J.-L. Dugelay, "Real Time Extraction of Body Soft Biometric from 3D Videos," Proc. ACM Int'l Conf. Multimedia (ACM MM), ACM Press, 2011, pp. 781\_782.