

Hand Gesture Recognition using TOF camera

Jyotirmayee Routray, Sarthak Rout, Jiban Jyoti Panda,
Bhabani Shankar Mohapatra, Hitendrita Panda
College of Engineering and Technology, Bhubaneswar, India.

Abstract

Gesture recognition is an important topic in computer vision because of its wide range of applications, such as Human computer Interaction, sign language interpretation, and visual surveillance. 3-D gesture recognition has attracted a lot of research interests in computer vision, pattern recognition, and human-computer interaction. The rising use of depth sensors greatly increases various hand gesture recognition approaches and applications, which were severely limited in the 2D domain with conventional cameras. Conventional approach gesture control in automobile industry etc possess certain challenges such as high computational power ,environmental lighting constraints or integrating voice control commands in noisy environments has also certain challenges,which can be solved and can be overcome by the use of Gesture Recognition Control using Time of flight camera which is fast and efficient. This paper aims at the use PicoZense ToF camera sensor and OpenCV libraries, we can easily overcome the challenges of recognising the gestures in low lighting conditions with noisy environment.

Keywords: Depth sensor, OpenCV, Hand gesture recognition, hand tracking, skeleton detection and tracking

I. INTRODUCTION

Recently, computer-assisted communication technology(HCI) has attracted attention as a promising means of human-computer communication. The development of HCI was guided by the co-ordinated development of computer power, various sensors, and display techniques. The interest in human gesture communication with HCI has also increased. This includes the movement of human hands and arms. Human hand gestures are verbal communication from simple pointing to complex interactions between people. The great advantage of body hands is the ability to communicate remotely. The use of HCI hand gestures requires that the human hand configuration be measured by a computer. [1] Performance is highly dependent on the accuracy of the acquisition and tracking of handicrafts. Current methods of finding and tracking use a variety of sensors including direct hand attachments, special feature gloves, and color or depth colors. Hand detection and image sensor tracking can be done with 2D or 3D information. However, since 3D data acquisition requires high computer power and expensive equipment, 2D methods are much more advanced than 3D. In the 2D visual and auditory components, the most common method is the visual-based approach, which uses details

such as color, shape, and edge. Visual-based methods can be divided into color-based and template-based methods. A color-based approach starts by finding the location of the hand using color data (RGB, HSV, YCbCr). After that, a color histogram was created from the acquired hand. Depending on the color of the histogram a region similar to hand color can be traced. A template-based approach creates a separate image in color or gray image. The image on the edge is aligned with the image of a trained hand, and the hand is followed. However, hand movements usually occur in 3D space. After that, the 2D method can only use 2D data, which removes travel details near the axis. This creates limitations for 2D methods naturally. Recently, 3D information acquisition equipment has become faster, more accurate, and more costly. These devices include depth sensors such as ToF cameras and PrimeSensor. After the advent of these devices, real-time 3D hand tracking methods grew rapidly. For example, Breuer et al. use the infra red ToF camera to create a real-time touch detection system. Grest et al. suggest a method of tracking human movement using a combination of in-depth and additional information. [2] In this article, we propose a real-time novel tracking method in deep space using the Picozense TOF sensor and OpenCV libraries. We took an input image and defined the ROI, then a gray measurement and transformation was performed and later convexity impairment points were calculated and the action separation was performed. The layout of this article is as follows. In Section 2, the related functions are briefly reviewed. In Section 3, an in-depth analysis of the proposed method of accessing and tracking is described. In Section 4, a number of tests are performed on our tracking system. Finally, we conclude the article in section 5.

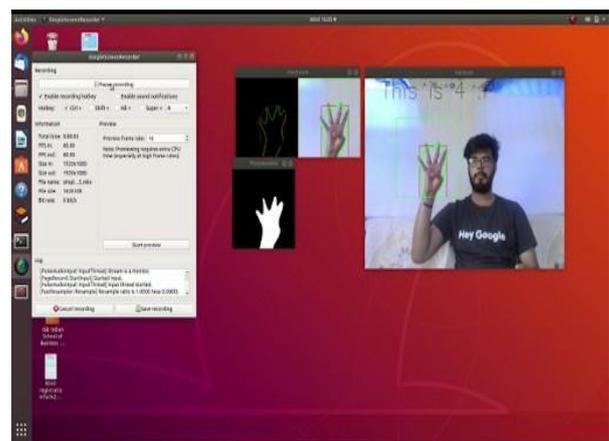


Fig. 1: Simple Gesture Classification

The rest of the paper is organized in following sections. Section II describes the Novel Contribution. Literature Reading is presented in Section III. Section IV includes Proposed Framework. System Structure and Design is presented in Section V. The working methodology is presented in Section VI. Section VII includes the Simulations and Results. Section VIII includes Conclusion.

II. NOVEL CONTRIBUTION

Post-Covid there has been an insurgency in the world of technology and our product which is a touchless technology using depth sensors i.e gesture recognition control from now on can be the next-gen technology for the world because of its infinite applications and easy-to-use features. We have implemented the algorithm which is fast and efficient than existing system for hand gesture recognition.

III. LITERATURE READING

Eshed Ohn-Bar has developed a vision-based system which employs a combined RGB and depth descriptor to classify human hand gestures. The technology stack used here is Computer Vision for Human Computer Interaction.

Asish Sharma was able to recognize and classify the Standard American Sign Language (ASL) images under several different environmental conditions of a person's hand to their correct meaning with the maximum accuracy possible using Image Processing, Feature Extraction Technique (Principal Component Analysis (PCA), Histogram of Gradients (HOG), Local Binary Patterns (LBP))[3]

Jay Prakash explained how gesture is a non-vocal way of communication that uses hand motion, facial expressions, and different postures of the body. They are enabling the machine to take command by recognizing different hand gestures using Computer Vision-Based Approach, and Instrumented Glove to make the machine smart.

Deaf Mute Communication Interpreter- A Review:[4] This paper aims to cover the various prevailing methods of deaf-mute communication interpreter system. The two broad classification of the communication methodologies used by the deaf –mute people are - Wearable Communication Device and Online Learning System. Under Wearable communication method, there are Glove based system, Keypad method and Handicom Touch-screen. All the above mentioned three subdivided methods make use of various sensors, accelerometer, a suitable micro-controller, a text to speech conversion module, a keypad and a touch-screen. The need for an external device to interpret the message between a deaf –mute and non-deaf-mute people can be overcome by the second method i.e online learning system. The Online Learning System has different methods. The five subdivided methods are- SLIM module, TESSA, Wi-See Technology, SWIPELE System and Web-Sign Technology.

An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform:[5] The proposed ISLR system is considered as a pattern recognition technique that has two

important modules: feature extraction and classification. The joint use of Discrete Wavelet Transform (DWT) based feature extraction and nearest neighbour classifier is used to recognize the sign language. The experimental results show that the proposed hand gesture recognition system achieves maximum 99.23 distance classifier.

Hand Gesture Recognition Using PCA:[6] In this paper authors presented a scheme using a database driven hand gesture recognition based upon skin color model approach and thresholding approach along with an effective template matching with can be effectively used for human robotics applications and similar other applications. Initially, hand region is segmented by applying skin color model in YCbCr color space. In the next stage thresholding is applied to separate foreground and background. Finally, template based matching technique is developed using Principal Component Analysis (PCA) for recognition.

Hand Gesture Recognition System For Dumb People:[7] Authors presented the static hand gesture recognition system using digital image processing. For hand gesture feature vector SIFT algorithm is used. The SIFT features have been computed at the edges which are invariant to scaling, rotation, addition of noise.

Hand Gesture Recognition for Sign Language Recognition:[8] A Review: Authors presented various method of hand gesture and sign language recognition proposed in the past by various researchers. For deaf and dumb people, Sign language is the only way of communication. With the help of sign language, these physical impaired people express their emotions and thoughts to other person.

Design Issue and Proposed Implementation of Communication Aid for Deaf Dumb People:[9] In this paper author suggested a system to aid communication of deaf and dumb people communication using Indian sign language (ISL) with normal people where hand gestures will be converted into appropriate text message. Main objective is to design an algorithm to convert dynamic gesture to text at real time. Finally after testing is done the system will be implemented on android platform and will be available as an application for smart phone and tablet pc.

Real Time Detection And Recognition Of Indian And American Sign Language Using Sift In:[10] Author suggested a real time vision based system for hand gesture recognition for human computer interaction in many applications. The system can recognize 35 different hand gestures given by Indian and American Sign Language or ISL and ASL at faster rate with virtuous accuracy. RGB-to-GRAY segmentation technique was used to minimize the chances of false detection. Authors proposed a method of improvised Scale Invariant Feature Transform (SIFT) and same was used to extract features. The system is model using MATLAB. To design and efficient user friendly hand gesture recognition system, a GUI model has been implemented.

Neelam K. Gilorkar et al.[11] emphasised on the static signs of Indian Sign Language and American Sign Language from images or video sequences which have been recorded under controlled conditions using special hardware like data gloves,

coloured gloves.

Shahrouz Yousefi introduced a novel solution for real-time 3D gesture-based interaction by finding the best match from an extremely large gesture database. This database includes images of various articulated hand gestures with the annotated 3D position/orientation parameters of the hand joints. A unique matching algorithm is based on the hierarchical scoring of the low-level edge-orientation features between the query frames and database and retrieving the best match. Once the best match is found from the database in each moment, the pre-recorded 3D motion parameters can instantly be used for natural interaction. The proposed bare-hand interaction technology performs in real-time with high accuracy using an ordinary camera.

Yikai Fang paper said vision-based hand interaction is more natural and efficient compared to the traditional interaction approaches, such as a keyboard, mouse, pen, etc.,. In this paper, they proposed a robust real-time hand gesture recognition method in which a specific gesture is required to trigger the hand detection followed by tracking; then they use motion and color cues to segment the hand; finally, the scale-space feature detection is integrated into gesture recognition in order to break the limitation of aspect ratio encountered in most learning-based hand gesture methods. Their experimental results gave them a very satisfactory performance by applying the proposed method to navigation of image browsing.

G. R. S. Murthy paper explained that with the ever-increasing diffusion of computers into the society, it is widely believed that the mouse and keyboard will become a bottleneck in the effective utilization of information flow between the computers and the human. Vision-based Gesture recognition can be a natural and powerful tool supporting efficient and intuitive interaction between the computer and the human. Visual interpretation of hand gestures can be helpful in achieving the naturalness desired for Human Computer Interaction (HCI). They surveyed the literature on visual interpretation of hand gestures in the context of its role in HCI. The purpose of the review was to introduce the field of gesture recognition as a mechanism to interact with computers.

IV. PROPOSED FRAMEWORK

Figure 2 Picozense camera is used to capture frames containing different gestures. After that we preprocess the library where we change the scale and then applied thresholding over it in order to separate the hand from the background. We then apply contour mapping over the hand and after that we apply convexity defect, which will indicate different gesture depending upon the convexity defects. It is integrated with pyautogui which helped us in doing tasks by giving commands by gesture.

V. SYSTEM STRUCTURE AND DESIGN

The whole system is divided into two major portion. Picozense depth camera is used in order to capture depth frames and further processing is done using OpenCV. The depth camera

and processing of depth frames are done realtime. The image is preprocessed using grayscale and thresholding in order to separate the hand from the background. We then apply contour mapping over the hand and after that we apply convexity defect, which will indicate different gesture depending upon the convexity defects. It is integrated with pyautogui which helped us in doing tasks by giving commands by gesture.

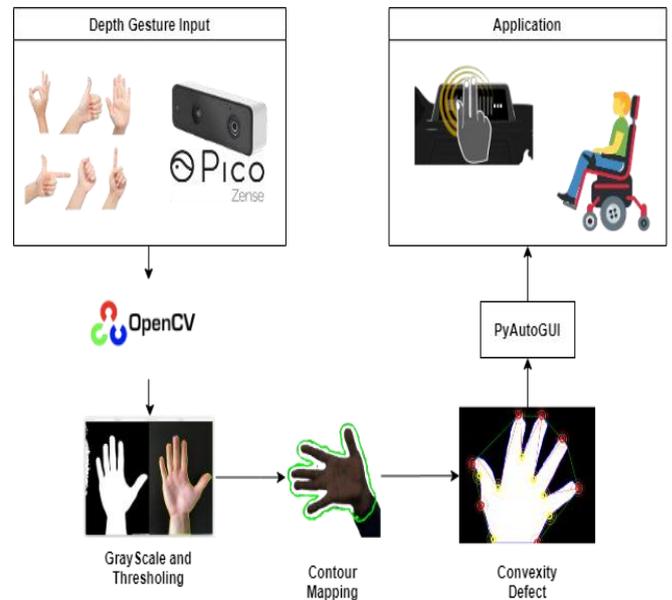


Fig. 2: The Proposed Framework of the System

PicoZense SDK: In the proposed architecture the computer is connected to the Picozense depth camera module. Picozense have their own SDK kit to research and build projects or products. This kit help us to work with the picozense depth camera. Using this we could avail the functions like thresholding the video frames, adjusting the depth values and much more. The depth frames are captured by the depth camera and using the SDK and its functions such as thresholding, grayscaleing, contour mapping and convexity defect points are applied using the script running behind.

A. Hardware System Structure

1) *Depth Camera Module* : For vision-based applications, selecting a proper camera plays a major role because noiseless images increase the efficiency of the computer vision algorithm. In this system, we have used a Picozense DCAM700 camera to detect the hand gestures. As it is a depth camera with multi functional capability, it finds a suitable application in our system.

B. Software System Design

1) *Picozense SDK*: Picozense SDK is the brain of our proposed model. It is suitably chosen for our design because it has lots of advantages as it helps to capture both RGB and depth frames. The camera is connected to it's USB port and

our computer vision algorithm is running on the SDK itself. This kit help us to work with the picozense depth camera. Using this we could avail the functions like thresholding the video frames,adjusting the depth values,etc.

2) *Image Processing Tool*: Basic steps followed to perform the task:

a) *Converting Color Image to GrayScale*: Transformations among RGB area like adding/removing the alpha channel, reversing the channel order, conversion to/from 16-bit RGB color (R5:G6:B5 or R5:G5:B5), furthermore as conversion to/from grayscale using:

$$RGB[A] \text{ to Gray} : Y, 0.299R + 0.587G + 0.114B \quad (1)$$

$$\text{Gray to RGB}[A] : R, Y, G, Y, B, Y, A, \max(\text{ChannelRange}) \quad (2)$$

b) *Image Filtering*: In one-dimensional signals, pictures can also be filtered with numerous filters like low-pass filters(LPF), high-pass filters(HPF) etc. LPF removes noises and helps in blurring the photographs. HPF filters facilitate to find edges within the pictures. OpenCV provides a perform `cv2.filter2D()` to flex a kernel with a picture. for instance, during a 5x5 averaging, this perform keeps this kernel on top of a component, adds all the twenty five pixels below this kernel, takes its average and replaces the central component with the new average price. It continues this operation for all the pixels within the image.

c) *Gaussian blurring method for blurring images*: In this, rather than box filter, the Gaussian kernel is employed. it's through with the perform, `cv2.GaussianBlur()`. The breadth and height of the kernel ought to be per order to work out that ought to be positive and odd. the quality deviation in X and Y direction, `sigmaX` and `sigmaY` ought to be given severally. If the sole `sigmaX` is given, letter Y is taken as same as letter X. If each area unit given as zeros, they're calculated from kernel size. Gaussian blurring is extremely effective in removing Gaussian noise from the image.

d) *Haar Cascade Classifier*: Object Detection exploitation Haar feature-based cascade classifiers is an efficient object detection technique projected by Paul Viola and archangel Jones in their paper, "Rapid Object Detection employing a Boosted Cascade of straightforward Features" in 2001. it's a machine learning based mostly approach wherever a cascade perform is trained from tons of positive and negative pictures. it's then accustomed notice objects in different pictures. OpenCV comes with a trainer furthermore as a detector. If we wish to coach our own classifier for any object like automotive, planes etc. we will use OpenCV to make one. The project deals with the detection. OpenCV already contains several pre-trained classifiers for face, eyes, smiles, etc. Those XML files area unit hold on within the `OpenCV/data/haarcascade/` folder. At first, a face and eye detector with OpenCV is made. currently the hand within the image is known. If hand is known, it returns the positions of the detected hand as `Rect(x,y,w,h)`. Once these locations area unit found, ROI for the hand in real time is created.

VI. METHODOLOGY AND SOFTWARE IMPLEMENTATION

A web camera attached to the system or a simple camera is required to capture the video feed of the gestures. Here we have used a Picozense DCAM700 camera to detect the hand gestures. The region of interest will be the area of the image containing the hand. After capturing the image, it must be converted to grayscale. It is done by the `cvtColor` function, which allows converting the image from one color space to another. As the first input, this function receives the original image. As the second input, it receives the color space conversion code. Since we want to convert our original image from the BGR color space to gray, we use the code `COLOR_BGR2GRAY`. Using one global value as a threshold might not be good in all cases, e.g. if the image has different lighting conditions in different areas. Therefore we have used `cv.adaptiveThreshold` method for better results.

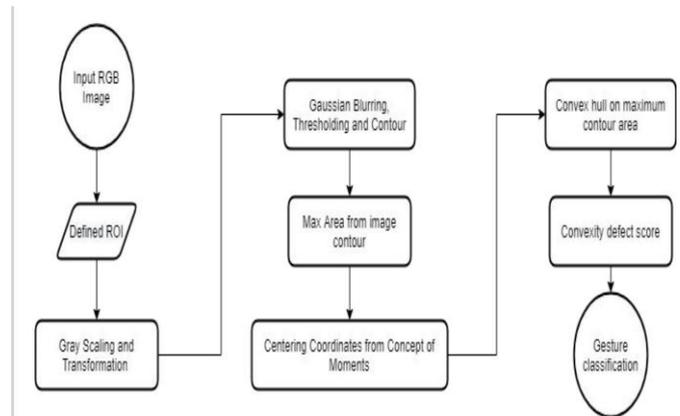


Fig. 3: Workflow

Our suggested system uses an extensively convex methodology to locate the hand in the area of interest which includes the following steps: sorting the depth of the object using eroded and dilated methods, removing the contours of the hand into a rough polygon, finding convex and concave points in the approximated polygon, and categorizing them into convex/concave points. But our main focus is on the convex points that correspond to the fingers in the hand. The convex points of one image at a time is determined by the convexity approach. For better efficiency and reliable results, the convex points are minimised as much as possible.

The palm is considered with the fingers as the convexity defects to calculate the convex hull of the given contour that helps in simplifying the complex structure of the image. The convex hull of the contour of the hand gesture is the convex polygon surrounded in gesture contour by all the convex vertices.

The convexity defect can be depicted as the difference between the convex hull and the contour. The convexity defect is defined as the points farthest from the convex points. So, if the finger tips are considered as the convex points, the trough between the fingers is considered as the convexity defects. So, it is possible to detect the fingertips by using contour

and convex defects. Convex defect character points of the hand contour are defined, and the finger angles and fingertip positions are measured to recognize the hand gestures. The OpenCV is used to perform our analysis. And we have built in functions for all the gestures provided by the PyAutoGUI module.

VII. SIMULATION AND RESULT

In our system we tried to assign functions to the gestures using a simple GUI. We were able to control the ball movements, zoom in and zoom out and rotate the ball on the screen using different hand gestures as shown in Gesture I, Gesture II and Gesture III.

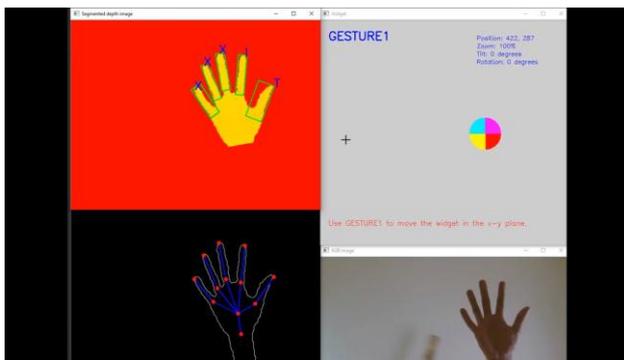


Fig. 4: Gesture I

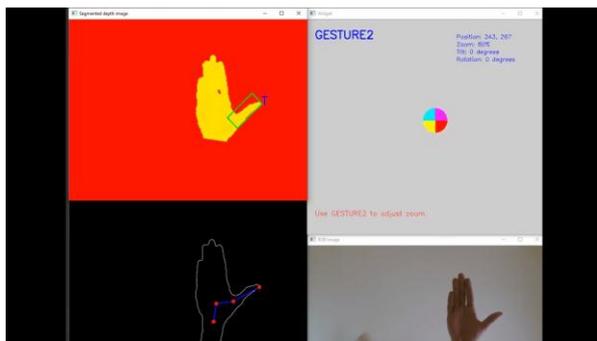


Fig. 5: Gesture II

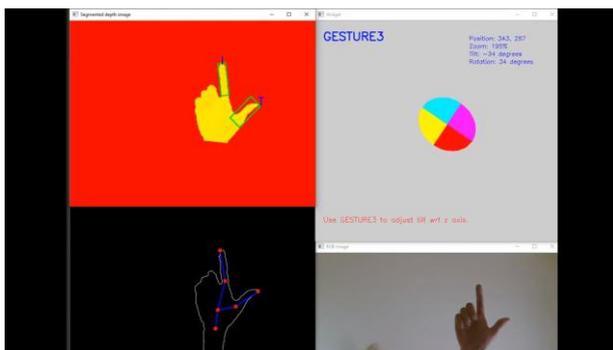


Fig. 6: Gesture III

VIII. CONCLUSION

This paper explains a system that is useful in controlling computer applications with the help of hand gestures. The proposed method here successfully developed a hand gesture recognition system, which is able to recognize what gesture is performed by the person and accordingly perform the functionality associated with it accurately. The main integral part of the system presently are the webcam, microphone and mouse. Our product can completely eliminate the use of mouse that will lead to a new contactless era of Human Computer Interaction(HCI) which is the current need in such pandemic situation.

REFERENCES

- [1] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," in *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 15, no. 6, December 2014.
- [2] R. Okada and B. Stenger, "A single camera motion capture system for human-computer interaction," vol. E91-D, 07 2008.
- [3] A. Sharma, A. Mittal, S.Singh, and V.Awatramani, "Hand gesture recognition using image processing and feature extraction techniques," in *International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020*, 2020, pp. 181–190.
- [4] S. K. A, A. Saraswathi.P, L.Sunny, J. K, and Aarthi.M, "deaf mute communication interpreter- a review," in *International Journal of Applied Engineering Research*, vol. 11, 2016, pp. 290–296.
- [5] M. S. Anand, N. M. Kumar, and A. Kumaresan, "An efficient framework for indian signlanguage recognition using wavelet transform," vol. 7, 2016, pp. 1874–1883.
- [6] M. K. Ahuja and A. Singh, "Hand gesture recognition using pca," in *International Journal of Computer Science Engineering and Technology (IJCSET)*, vol. 5, no. 11, July 2015, pp. 267–274.
- [7] S. P.More and P. A. Sattar, "Hand gesture recognition system for dumb people."
- [8] C. Kaur and N. Gill, "An automated system for indian sign language recognition," in *International Journal of Advanced Research in Computer Science and Software Engineering*.
- [9] P. Pandey and V. Jain, "Hand gesture recognition for sign language recognition: A review," in *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 4, March 2015.
- [10] N. Nagpal, D. A. Mitra, and D. P. Agrawal, "Design issue and proposed implementation of communication aid for deaf dumb people," in *International Journal on*

Recent and Innovation Trends in Computing and Communication, vol. 3, March 2015, pp. 147–149.

- [11] N. K. Gilorkar and M. M. Ingle, “Real time detection and recognition of indian and american sign language using sift,” in *International Journal of Electronics and Communication Engineering Technology (IJECET)*, vol. 5, May 2014, pp. 11–18.