Segmentation and Recognition of Fingers Using Microsoft Kinect

Smit Desai
Sarvajanik College of Engineering and Technology
Surat, Gujarat (India)
thesmitdesai5@gmail.com

Abstract. Hand gesture identification is a very important part of HCI. In this paper, I have presented a very efficient algorithm for finger segmentation. Using fingers as an input medium, our interaction with the computer can become easier. Microsoft Kinect, which is a depth sensor is used to capture the image which is used for finger segmentation. Background is removed from the captured image by accepting pixels, which fall in a fixed range of depth. The image is further pre-processed and then palm area is identified and removed to obtain separate fingers. Further, to identify open fingers as gesture-kNN classifier is used. This proposed algorithm has achieved more than 90% accuracy.

Keywords: Microsoft Kinect, Gesture recognition, Finger Segmentation, centroid, Feature Extraction, kNN Classifier

1. Introduction

Since evolution of computers, wired input methods and devices are used for Human Computer Interaction (HCI). With development of new and innovative computer based applications and with ever evolving computing technology, researchers have developed new HCI methods. After successful use of wired and wireless HCI methods and devices, scientists are working on more complex but useful HCI methods. Voice, touch and gestures are the next three HCI mediums that the scientists are working on. Body posture, body language, facial expression, hand gestures are very powerful mediums of interaction. Normally it is seen that these medium of interaction are more powerful and convenient for human beings in general and specially for physically challenged people. However, providing understanding of such intelligence to a computer is a very challenging and complex task. Such real time gesture identification task is even more difficult. Up until now many different hand gesture recognition techniques and methods are used for interaction with machine. Some of them have made use of colored gloves or color code on fingers such as [1] and some researchers have attempted to recognize gestures without any color code. In this paper I have presented an algorithm wich recognizes open finger based gestures without any color code. I have used Microsoft Kinect for capturing hand image from 3D space. This paper is divided in seven section. This introduction section is followed by literature review. Section three is devoted to introduction to Kinect sensor which is followed by a discussion on hand capturing and its segmentation. In section six processing of segmented hand images is discussed. In the following section feature extraction and classification is described which is followed by concluding remaraks.

2. Literature Review

Lately in modern computing technology hand gestures are used frequently for interaction with computers. For example, Pavlovic et.al. [1] have categorized human gestures in three categories viz. one, hand and arm gestures, two, head and face gestures and three, body gestures. The first category of gestures, that is, hand and arm gestures including the palm and fingers are mainly used for applications like game playing and understanding and interpreting sign languages etc. The second category of gestures is manly utilized for applications like face recognition or emotions or actions like blowing whistle, smiling, and anger recognition. The third and last category of gestures include gestures and postures of entire human body and therefore they are more useful for teaching and training of athletes, dancers etc. Hand gestures are very useful and more and more researchers have concentrated on it. Initially hand gestures were recognized by some color coding. This technique used color gloves, bands or some indicative gears on their hand, palm, wrist or fingers for identifying the gestures. C. Keskin et.el [2] presented their work based on colored gloves for communicating with a healthcare system. I. Agrawal. Et.el. [3] developed a gesture recognition system based on colored gloves. This system is designed as a tutor for hearing impaired people. In this paper authors have used red color gloves for localization of hands, where as in Zhou Ren et.el. [4] use black colored wrist band for localization of palm. Though Zhou Ren et.el. [4] use Kinect sensor, they have use used color code for hand segmentation. Further they made use of Finger – Earth Mover's Distance for measuring dissimilarities between gestures. In this work they demonstrated applications like arithmatic computation and Rock-Paper-Scissors Game playing using hand gestures. Giulio Marin [5] has presented an algorithm, which finds are of the palm using the centroid of the palm. Even though authors used a depth sensor, they used a colored band for isolation of hand like Zhou et.el. [4]. There are many other work which are based on such color codes, but since this technique is based on vision, it requires proper lighting and illumination condition Erol [6] and Suwarez et.el. [7] . After introduction of Microsoft Kinect researchers started using it for gesture recognition. Since it has a built in depth sensor and also gives skeletal details of human body it has become more popular. Many interesting applications of Kinect are found in literature. It consists of game playing, assistive tools and many more. Van BangLe et.el. [8] presented a work where they proposed an algorithm which finds the center point of the palm and also identifies fingers. Here they segmented the palm area and then image contour is used to find moments. J.L. Raheja et.el. [9] presented a work to find the finger tips and center of the palm using finger tracking algorithm. This algorithm can track both the hands. Here hand is segmented by taking a calculated threshhold and NITE library [10] . J Shukla & A. Dwivedi [11] used Kinect to capture hand image with depth information. They also used countour of the segmented hand image. For identifying hand gestures, they used convex hull and convexity defects of the hand countour as the featureset. They suggested Naiive Bayes classifier for gesture recognition. K.K. Biswas et.el [12] propsoed an algorithm which recognizes eight hand gestures clap, call, greet etc. Here they capture hand gesture image through Kinect and then this image is prossed for back ground removal. They used depth histogram equilizer for identify foreground. They further divided the depth image in ten bins to find features and then they identified the features using support vector machine (SVM). Pan Jing [13] proposed an algorithm where they developed virtual touch screen tool using Kinect. This is done by identfying and tracking the fingers.

3. Suggested Algorithm

The proposed algorithm captures images from Microsoft kinect. The captured images are in RGB color space, therefore it is converted into binary image and then smoothing of the image is done. This smoothened image may have some discontinuity which are removed by performing morphological operations. These three processes are done under preprocessing stage. Now using moments, centroid of the image is found and an ellipse which fits the hand image is considered and its minor axis is computed. This minor axis is used as the diameter of the ellipse which is then removed from the image. This circular area is calculated using the centroid as the origin. Removal of this circular area will remove the palm from the image leaving us with fingers and some noise. Here I have made one observation that most of the hand images are connected with the boundary of image and threfore when circular area is removed the forearm will remain with the boundary. I remove all the parts which are connected with the boundary of the image and we will be left with fingers. Now region of interest(ROI) is cropped for finding the featureset. After finding featureset, kNN classifier is used for identifying open fingers as a gesture. The block diagram of algorithm is shown in Fig 1.

4. Segmentation Of Hand

As mentioned earlier, fingers are very reliable for HCI. When user wishes to use hand gestures as a medium for interaction, he generally keeps his hand in front of the imaging device. Here I have used Kinect as an imaging device which has an inbuilt depth sensor. This depth sensor measures the depth of an object from the device. I use this depth for segmenting hand from the body of the user. Here I suggest depth range of [250mm,650mm] from the device, as region of interest (ROI) separating background from the hand as it is suggested in [4], [10], [14], [9]. That means when users' hand comes in the depth range of [250mm,650mm] it will be cropped and used for further processes. This is done using NITE library as it is suggested in cliff et.el. [10]. The cropped hand images are shown in Fig. 2.

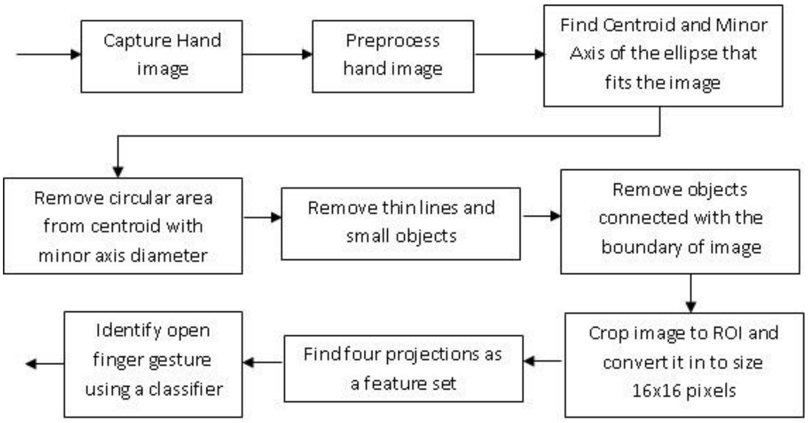


Fig. 1: Block Diagram of proposed algorithm

As discussed, the image captured from Kinect is in RGB color space. For processing this image we need to convert this image into binary. For converting RGB image into binary, I propose to use Otsu's global threshold [15].

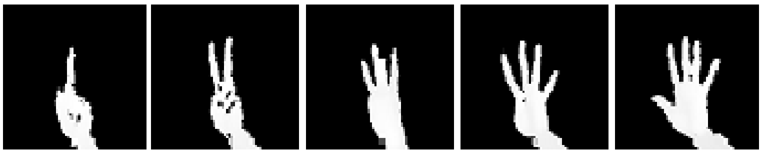


Fig. 2. Gesture images captured through Kinect

The next pre-process step is dilation of the image. I propose dilation because some times while converting the image from RGB to binary, some portion of the image gets separated. To connect these separated parts, dilation is used. Even after dilation, there is a possibility to have some aberrations in the form of holes in the image and therefore in the next step we fill all such holes to make a complete image. Major stages of this process have been illustrated in Fig. 3.

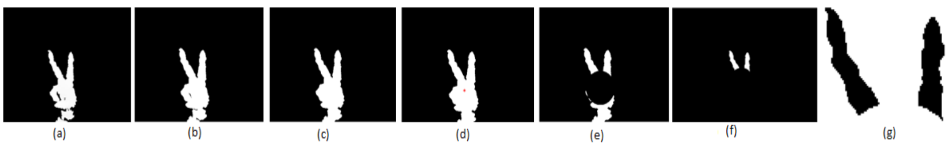


Fig. 3. (a) original gesture image (b) boundary of hand smoothen (c) holes are removed from image (d) image with centroid of the hand (e) image with circle drawn from centroid with minor axis diameter (f) non-finger components are removed from image (g) fingers cropped from the image

5. Processing

The next task is to determine the position of the hand in the image. Zhou Ren et.el. [4] and Guilo et.el. [5] have used centroid for palm area identification. For estimating position of the hand, we find centroid of the hand image, which is nothing but a point where all the mass of an object is concentrated without changing its first moment. It is a statistical measure which can be derived from the raw moments of an object image. Let f(x,y) be a binary image, then its centroid is;

$$\left(\frac{\mu_{10}}{\mu_{00}}, \frac{\mu_{01}}{\mu_{00}}\right)$$

Where

$$\mu_{00} = \sum_{x=0}^{width} \sum_{y=0}^{height} x^0 y^0 f(x,y)$$

$$\mu_{10} = \frac{\sum_{x=0}^{width} \sum_{y=0}^{height} x \; f(x,y)}{\mu_{00}}$$

$$\mu_{01} = \frac{\sum_{x=0}^{width} \sum_{y=0}^{height} y \; f(x,y)}{\mu_{00}}$$

Here μ_{00} is zeroth raw moment and it gives the area of object image. The other variables that we calculate for the hand image is the minor axis of an ellipse which fits the object of the hand image. I noticed that when centroid of hand image is calculated it generally falls in the area of palm. So if the palm area is separated and removed from the hand image, we can segment fingers of the hand. Considering this observation, we remove a circle taking centroid as the origin and minor axis as the diameter. It is shown in Fig. 4.

Fig. 4. Ellipse over the object to determine minor radius of the object

In this process, some of the thin portions and tiny objects of the palm may still be present in the image. To remove these thin portions we perform morphological operation erosion. This process completely separates out forearm from the other parts of the hand image. I noticed one more fact about the hand images that I had captured with Kinect: Forearm of hand along with the wrist is usually also captured while capturing the image and that portion of the image is usually connected with the boundary of the image. Taking advantage of this fact I removed all the objects which are connected with the boundary of the image.

In the next stage, these images are processed for counting fingers visible in the image. This is what we are interested in. In most of the cases we get separate fingers in the images but merely counting them does not serve our purpose because in many cases we get joint fingers too. Fig 4 shows step wise process of algorithm up to getting cropped fingers and Fig 5 is a case where cropped image has joint fingers.

(a) (b) (c) (d)

Fig. 5. (a) image with centroid of the hand (b) image with circle drawn from centroid with minor axis diameter (c) non-finger components are removed from image (d) fingers cropped from the image

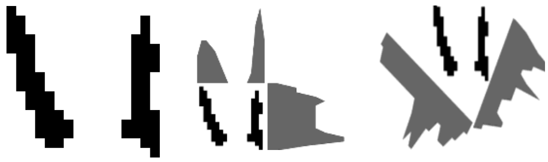
Fig. 6 : an example of joint separated fingers

Since there are possibilities of having images where joint fingers are there and under segmentation we need to identify open finger gesture applying classification technique. In the following section feature extraction and classification are described.

6. Feature Extraction And Classification

Many feature extraction methods can be traced in the literature. Feature are usually divided in three categories, structural features [5], [11], [16], [17], statistical features [4], [10], [17], [18] and hybrid feature. Since extracted fingers may be in any direction, here we have used four projection profiles as it used in [19] for Gujarati handwritten numeral identification. It is a structural feature of an object. To find four projection viz vertical, horizontal and two diagonal profiles, first extracted finger image is reduced to an image of size 16 x 16. For reduction of image size, I propose to use nearest neighbor interpolation method. Now projection profile is found for the extracted finger image. By projection profile we mean total black pixels in each of the vertical, horizontal and two diagonals of the 16x16 image. That comes to 94 features,16 each for horizontal and vertical profiles and 31 each for two diagonal profiles, of an image. This projection feature set is illustrated in Fig 6.

Many classification methods [3], [4], [11], [12], [19], [17], [18] are used for classification of hand gestures. For classification purpose I propose to use kNN algorithm with Euclidian distance. I tried out classification with different values of k, but after all experiments I found that the best result is obtained with k=1. Here I have used a data set of five hand gestures collected from fifty people. Thus we have a total of two hundred and fifty hand gesture images. I have trained our algorithm with fifty percentage of dataset and tested it for the rest of the fifty percent of the images. Table [1] and table [2] show the accuracy of success of proposed algorithm for k=1 and n=25.



(a) (b) (c)

Fig. 7. (a) Finger figure converted in size of 16x16 (b) horizontal and vertical profile of 16x16 image (c) two diagonal profiles of 16x16 image

The results show that the proposed algorithm has achieved 91.20% of success for unseen data and 95.60% of accuracy for complete dataset. The algorithm gives 100% accuracy for the training data set. We have tried out Support Vector Machine (SVM) as a classifier to identify open finger gestures, but it gave us a relatively lower accuracy of 65%.

Table 1 :Confusion matrix of test data set of size 25

	1	2	3	4	5
1	23	1	0	0	1
2	0	25	0	0	0
3	0	1	22	0	2
4	0	0	0	19	6
5	0	0	0	0	25

Table 2 :Confusion matrix of test data set of size 50

	1	2	3	4	5
1	23	1	0	0	1
2	0	25	0	0	0
3	0	1	22	0	2
4	0	0	0	19	6
5	0	0	0	0	25

7. Conclusion And Future Work

The proposed morphological operation based algorithm gives high accuracy of identification of open fingers based hand gestures. In this proposed algorithm I have made a couple of assumptions. First, the hand must be in the range of [250mm,650mm] for segmentation, and two the segmented hand must be connected with one of the boundary of the image. Here if the segmented hand image is not connected to the boundary of the image, we have cropped it to the lower boundary of the image. The main reason of non-identification of gestures is the violation of the second assumption made for the algorithm. There is ample scope of correction of the results without any assumption.

Fingers create important and expressive gestures for human communication and they are very widely used in game playing. This algorithm has open up various avenues to work in the direction of developing an interaction media for communicating with any electronic devices including computers. Touch screen computers and other devices are currently market, but by extending this work, an effective interaction will be made possible just by using fingers and without the need of any touch.

References

1. Pavlovic V., Sharma R., Huang T., Visual Interpretation of Hand Gesture for Human Computer Interaction, IEEE transaction on Pattern Analysis and Machine Intelligence, vol. 19(7), pp. 677 – 695 (1997)
2. Keskin C., Balci K., Aran O., Akrun L., A Multimodal 3d Healthcare Communication System, In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video(2007)
3. Agrawal I, Johar S., Santhosh J., A Tutor for the Hearing Impared (Developed using Automatic Gesture Recognition), International Journal of Computer Science, Engineering and Application, vol. 1(4), pp. 49 – 61(2011)
4. Ren Z., Youn J., Meng J. Zhang Z., Robust Part Based Hand Gesture Recognition Using Kiect Sensor, IEEE Transaction Multimedia, vol. 15(5), pp. 1110 – 1120(2013)
5. Marin G., Fraccaro M., Donadeo M., Dominio F., Zanuttigh P., Palm Area Detection for Reliable Hand Gesture Recognition, In IEEE International Workshop on Multimedia Signal Processing 2013 (MMSP-2013)(2013)
6. Erol A., Bebis G., Nicolescu M., Boyle R., Twombly X., Vision Based Hand Pose Estimation : A Review, Computer Vision and Image Understanding, pp. 52 – 73(2007)
7. Suarez J., Murphy R., Hand Gesture Recognition with Depth Image : A Review, In IEEE International Symposium on Robot and Human Interactive Communication, pp. 9 – 13(2012)
8. Le V.B., Nguyen A.T., Zhu Y., Hand Detecting and Positioning Based on Depth Image of Kinect Sensor, International Journal of Information and Electronic Engineering, vol. 4(3), pp. 176 – 179(2014)
9. Raheja J., Chaudhary A., Singal K., Tracking of Fingertips and Centre of Palm using Kinect, In IEEE International Conference on Computational Intelligence, Modelling and Simulation, Malaysia(2011)
10. Cliff C., Mirfakhroei S.S., Hand Gesture Recognition Using Kinect, Technical Report No ECE-2013-04, Boston University(2013)
11. Shukla J., Dwivedi A., A Method for Hand Gesture Recognition, Fourth International Conference on Communication System and Network Technologies(2014)

12. Biswas K., Basu S., Gesture Recognition using Microsoft Kinect, In 5th International Conference on Automation, Robotics and Application, Wellington, Newzeland(2011)
13. Jing P., Ye-Peng G., Human Computer Interaction using Poining Gesture based on an Adaptive Virtual Touch Screen, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6(4), pp. 81-91(2013)
14. Du H., To T.H., Hand Gesture REcognition using Kinect, Technical Report No. ECE-2011-04, Boston University(2011)
15. Otsu N., A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man, and Cybernetics, vol. 9(1), pp. 62 – 66(1979)
16. Verma H.V., Eshan A., Chandra S., Gesture Recognition Using Kinect for Sign Language Translation, In IEEE Second International Conference on Image Information Processing (ICIIP-2013)(2013)
17. Arun K., Chris Z., Joseph, J. L. Jr., Poster : Real Time Markless Kinect Based Finger Tracking and Hand Gesture Recognition, In IEEE Symposium on 3D User Interface, Orlando, USA(2013)
18. Hamissi M., Foez K., Real Time Hand Gesture Recognition Based on Depth Map for Human Robot Interaction, International Journal of Electrical and Computer Engineering, vol. 3(6), pp. 770 – 778(2013)
19. Desai A.A., Gujarati Handwritten numeral optical character recognition through neural network, Pattern Recognition, vol. 40(1), pp. 2581 – 2589(2010)
20. Verma H., Aggarwal E., Chandra S., Gesture Recognition using Kinect for Sign Language Transalation, In IEEE International Conference on Image Processing (ICIIP-2013)(2013)