

HAND GESTURE RECOGNITION USING KINECT SENSOR

Dr.M.AntoBennet¹, R.Himaya², G.V.K.JothiSree³, K.Deepa⁴, P.Preethi⁵,

V.Aarthi⁶, M.Sowmiya⁷

¹ Professor, ^{2,3,4,5,6,7} UG students, **Department of Electronics and Communication Engineering**
VEL TECH, Chennai-600062

Email: bennetmab@gmail.com¹, himayaramesh1997@gmail.com²

ABSTRACT

The recently developed Kinect sensor have provided new opportunities for Human Computer Interaction (HCI). Although great progress has been made by leveraging the Kinect sensor in human body tracking, face recognition and human action recognition, robust hand gesture recognition remains an open problem. Compared to the entire human body, the hand is a smaller object with more complex articulations and more easily affected by segmentation errors. It is thus a very challenging problem to recognize hand gestures. The project focuses on building a robust part-based hand gesture recognition system using Kinect sensor. To handle the noisy hand shapes obtained from the Kinect sensor, a novel distance metric, Finger Earth Mover's Distance (FEMD) is used to measure the dissimilarity between hand shapes. As it only matches the finger parts while not the whole hand, it can better distinguish the hand gestures of slight differences. The extensive experiments demonstrate that our hand gesture recognition system is accurate and robust to hand articulations, distortions and orientation or scale changes, and can work in uncontrolled environments. The superiority of our system is further demonstrated in two real-life HCI applications.

Keywords: Finger Earth Mover's Distance (FEMD), PIC Digital Signal Controller (DSCs), Minimum Near-Convex Decomposition (MNCD),

INTRODUCTION

Embedded system contain processing cores that are typically either microcontroller or digital signal processor (DSP). The embedded system is a computer system designed for specific control function within a larger system, often with real-time computing constraints. It is embedded as part of a complete device often including hardware and mechanical parts. By contrast, a general-purpose computer, such as a personal computer (PC), is designed to be flexible and to meet a wide range of end-user needs [1,2]. Embedded system control many devices in common use today. Physically, embedded system range from portable devices such as digital watches and MP3 player, to large stationary installations like traffic light, factory controllers. Complexity varies from low, with a single microcontroller chip, to very high with multiple units, peripherals and networks mounted inside a large chassis or enclosure [3,4]. Many embedded system have substantially different design constraints than desktop computing application. Embedded system is typically a design making use of the power of a small microcontroller, like the Microchip

PICmicro MCU or PIC Digital Signal Controller (DSCs). These microcontrollers combine a microprocessor unit (like the CPU in a desktop PC) with some additional circuits called "peripherals", plus some additional circuits on the same chip to make a small control module requiring few other external devices. This single device can then be embedded into other electronic and mechanical devices for low-cost digital control. This paper will design and build a man-machine interface using a video camera to interpret the American one-handed sign language alphabet and number gestures. The keyboard and mouse are currently the main interfaces between man and computer. In other areas where 3D information is required, such as computer games, robotics and design, other mechanical devices such as roller balls, joysticks and data gloves are used. Humans communicate mainly by vision and sound using Human Computer Interaction (HCI), therefore, a man machine interface would be more intuitive if it made greater use of vision and audio recognition. Another advantage is that the user not only can communicate from a distance, but need have no physical contact with the computer. However, unlike audio commands, a visual system would be preferable in noisy environments or in situations where sound would cause a disturbance[6-12]. The visual system chosen was the recognition of hand gestures Finger Earth Mover's Distance (FEMD). The amount of computation required to process hand gestures is much greater than that of the mechanical devices, however standard desktop computers are now quick enough to make this paper hand gesture recognition using computer vision a viable proposition. It describes the various design options, rationale and conclusions[4,5]. The structure of the write-up and completed project architecture is outlined below. The system will use a single, color camera mounted above a neutral colored desk surface next to the computer. The output of the camera will be displayed on the monitor. The user will be required to wear a white wrist band and will interact with the system by gesturing in the view of the camera. Shape and position information about the hand will be gathered using detection of skin and wrist band color shown in fig 1.

1. The detection will be illustrated by a color change on the display. The design of the detection process will be covered in it. The shape information will then be refined using spatial knowledge of the hand and wrist band.
2. The refined shape information will then be compared with a set of predefined training data to recognize which gesture is being signed. In particular, the contribution of this project is a novel way of speeding up the comparison process. A label corresponding to the recognized gesture will be displayed on the monitor screen shows the successful recognition of a series of gestures.
3. This describes an application of the system a gesture driven windows interface.
4. Finally, This describes how the work has achieved the goals set and further work that could be carried out.



Fig 1. Wrist band and neutral colored background

In order to detect hand gestures, data about the hand will have to be collected. A decision has to be made as to the nature and source of the data. Two possible technologies to provide this information are: 1. A glove with sensors attached that measure the position of the finger joints. 2. An optical method has been chosen, since this is more practical (many modern computers come with a camera attached), Minimum Near-Convex Decomposition (MNCD), cost effective and has no moving parts, so is less likely to be damaged through use. The first step in any recognition system is collection of relevant data. In this case the raw image information will have to be processed to differentiate the skin of the hand (and various markers) from the background. Chapter 3 deals with this step. Once the data has been collected it is then possible to use prior information about the hand (for example, the fingers are always separated from the wrist by the palm) to refine the data and remove as much noise as possible. This step is important because as the number of gestures to be distinguished increases the data collected has to be more and more accurate and noise free in order to permit recognition. The next step will be to take the refined data and determine what gesture it represents. Any recognition system will have to simplify the data to allow calculation in a reasonable amount of time (the target recognition time for a set of 36 gestures is 25 frames per second). Obvious ways to simplify the data include translating, rotating and scaling the hand so that it is always presented with the same position, orientation and effective hand-camera distance to the recognition system.

PROPOSED WORK

GESTURE RECOGNITION

Hand gesture recognition is of great importance for human-computer interaction (HCI), because of its extensive applications in virtual reality, sign language recognition, and computer games. Despite lots of previous work, traditional vision-based hand gesture recognition methods are still far from satisfactory for real-life applications. Because of the nature of optical sensing, the quality of the captured images is sensitive to lighting conditions and cluttered backgrounds, thus optical sensor based methods are usually unable to detect and track the hands robustly, which

largely affects the performance of hand gesture recognition. To enable a more robust hand gesture recognition, one effective way is to use other sensors to capture the hand gesture through the data glove. Unlike optical sensors, such sensors are usually more reliable and are not affected by lighting conditions or cluttered backgrounds.

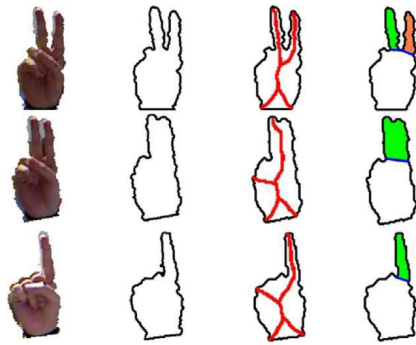


Fig.2 Hand Gesture Recognition

However, as it requires the user to wear a data glove and sometimes requires calibration, it is inconvenient to use and may hinder the natural articulation of hand gesture. Fig 2 shows such data gloves are usually more expensive than optical sensors, e.g., cameras. As a result, it is not a very popular way for hand gesture recognition. The recent development of inexpensive depth cameras, e.g., the Kinect sensor, new opportunities for hand gesture recognition emerge. Instead of wearing a data glove, using the Kinect sensor can also detect and segment the hands robustly, thus it provides a valid base for gesture recognition. In spite of many recent successes in applying the Kinect sensor to articulated face recognition, human body tracking and human action recognition, it is still an open problem to use Kinect for hand gesture recognition. Due to the low-resolution of the Kinect depth map, typically, of only 640 480, although it works well to track a large object, e.g., the human body, it is difficult to detect and segment a small object from an image with this resolution, e.g., a human hand which occupies a very portion of the image with more complex articulations. The segmentation of the hand is usually inaccurate, thus may significantly affect the recognition step. To illustrate the above problem, the first column shows three examples. It can be seen that the contours (in the second column) have significant local distortions in addition to pose variations. Kinect sensor, the two fingers of the second hand are indistinguishable as they are close to each other. Unfortunately, classic shape recognition methods, such as correspondence-based shape matching algorithms and skeleton matching methods cannot robustly recognize shape contour with severe distortions. The red skeletons of the last two hands are very similar. Hence skeleton matching algorithms classify them as the same gesture.

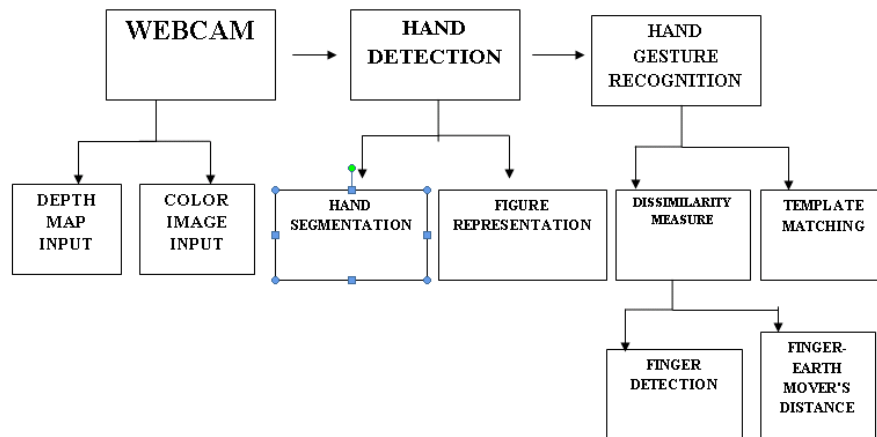


Fig 3 FEMD

based hand gesture recognition system

Due to the low resolution and inaccuracy of the Kinect sensor, the two fingers of the second hand are indistinguishable as they are close to each other. Unfortunately, classic shape recognition methods, such as correspondence-based shape matching algorithms and skeleton matching methods cannot robustly recognize shape contour with severe distortions. For example, as shown in the third column of Fig.3 shows the red skeletons of the last two hands are very similar. Skeleton matching algorithms classify the same gesture. Clearly, recognizing such noisy hand contours is challenging, especially if there are many hand gestures to recognize. In order to address this problem, a novel shape distance metric called Finger-Earth Mover's Distance (FEMD). FEMD is specifically designed for hand shapes. As it only matches the fingers while not the whole hand, it can better handle the noisy hand shapes obtained by Kinect sensor. The framework of our hand gesture recognition system. Kinect sensor used as the input device which captures both the color image and its corresponding depth map. With the help of depth cue, it can detect the user's hand robustly to the cluttered backgrounds and lighting conditions. Then, it represents the hand shape by its finger parts, which is detected by shape decomposition. Finally, the dissimilarity between the obtained hand shape and each gesture template is measured by the proposed distance metric, FEMD, for gesture recognition. This method build a new challenging dataset (containing 1000 cases collected in uncontrolled environments). Tests on this dataset shows that our hand gesture recognition system not only operates accurately and efficiently (a mean accuracy of 93.2% in 0.0750 s per frame), but also is robust to uncontrolled environments and hand gesture variations in orientation, scale, articulation, and shape distortions. To compare our algorithm with shape contexts, and skeleton path similarity and show our superiority in hand gesture recognition. Furthermore, on top of our gesture recognition algorithm, build two real-life HCI demos to illustrate the effectiveness of the method.

The main contributions of this method are as follows:

- A part-based hand gesture recognition system, based on a novel distance metric Finger Earth Mover Distance (FEMD). It is robust to orientation, scale, articulation changes as well as local distortions of hand shapes.

- Demonstration of hand gesture recognition algorithm in two HCI applications. The proposed system operates accurately and efficiently in uncontrolled environments. It is applicable to other HCI applications.

HIDDEN MARKOV MODEL:

HMMs have been widely and successfully used in speech recognition and handwriting recognition. Consequently, they seem to be effective for visual recognition of complex, structured hand gestures such as sign language recognition. A HMM can be employed to represent the statistical behaviour of an observable symbol sequence in terms of a network of states. For each observable symbol, it can be modelled as one of the states of the HMM, and then the HMM either stays in the same state or moves to another state based on a set of state transition probability associated with the state.

The variety of the observable symbols for which the HMM uses a particular state is described in terms of the distribution of probability that each observable symbol will occur from that state. Thus, an HMM is a doubly (observable and hidden) stochastic model where the observable symbol probability distribution for each state captures the intra-state variability of the observable symbols, and the state transition probability describe the underlying dynamic structure of the observable symbols. HMMs used to recognize different gestures because of their simplicity and reliability.

The HMM uses only three parameters: the initial state probability vector, the state transition probability matrix, and the observable symbol probability matrix. Analysis of dynamic images naturally will yield more accurate recognition than that of a single static image. Gestures are recognized in the context of entire image sequences of non-constant lengths. Using an HMM for gesture recognition is advantageous because it is analogous to human performance which is a doubly stochastic process, involving a hidden immeasurable human mental state and a measurable, observable human action.

ROBUSTNESS TO CLUTTERED BACKGROUNDS:

Our hand gesture recognition system is robust to cluttered backgrounds because the hand shape is detected using the depth information thus the backgrounds can be easily removed. When the hand is cluttered by the background, which is hard for other hand gesture recognition methods that use colored markers to detect the hand. A difficult case for the skin color-based hand gesture recognition approaches, where the hand is cluttered by the user's face. However, our hand segmentation is very accurate using Kinect sensor.

ROBUSTNESS TO DISTORTIONS AND HAND VARIATIONS:

In real-life environment, a hand can have variations on orientation, scale and articulation. Besides, because of the limited resolution of the depth map, the hand shapes are always distorted or ambiguous. However, it can demonstrate that the proposed dissimilarity distance metric, Finger-Earth Mover's Distance (FEMD), is not only robust to the orientation and scale changes of the hand, but also insensitive to distortions and articulations and 3 hands with different orientations. The initial point (the red point on the figure) and the center point (the blue point) are relatively fixed in the shapes. Thus the time-series curves of these hands are similar and their

distances are very small. There are 3 hands of different size. Because the time series curve and the FEMD distance are normalized, they are correctly recognized as the same gesture. Hence it can conclude that FEMD is robust to orientation and scale changes. Furthermore, our hand gesture recognition method is robust to the articulations and distortions brought by imperfect hand segmentation. Since the proposed FEMD distance metric uses global features (fingers) to measure the dissimilarity, local distortions are tolerable. As for articulations, the leftmost column shows 4 hand images of the same gesture the middle column shows the corresponding hand shapes and the rightmost column shows their time-series curves. The rightmost column of detecting the finger parts (the yellow regions), represent each shape as a signature whose clusters are the finger parts. Particularly the signatures have 2 clusters and therefore FEMD metric is insensitive to distortions and articulations.

ACCURACY AND EFFICIENCY:

In order to evaluate the accuracy and efficiency of our system, two experiments are conducted on the new dataset. Experiment I uses thresholding decomposition to detect the finger parts for FEMD measurement uses near-convex decomposition for FEMD finger detection.

PERFORMANCE OF ANALYSIS

The block diagram for hand gesture recognition shown in fig 4. A hand gesture has every position in the frame, every hand yaw angle and for various distances from the camera then the comparison task becomes impossibly large. However, the inclusion of a wrist band in detection helps simplify the process by removing these degrees of freedom. The angle between the centroids of the wrist band and the hand designates the yaw of the hand. so this degree of freedom can be removed. The distance between the centroids allows the hand to be scaled to a constant size so the hand-to-camera distance degree of freedom can be removed. Finally, since the centre of the hand is indicated by the hand centroid the hand position degree of freedom can also be removed by centring detection about this point shown in fig 5&6. The gesture is recognised by the webcam and the respected operation will be processed according to the gesture which are already programmed with necessary functions.

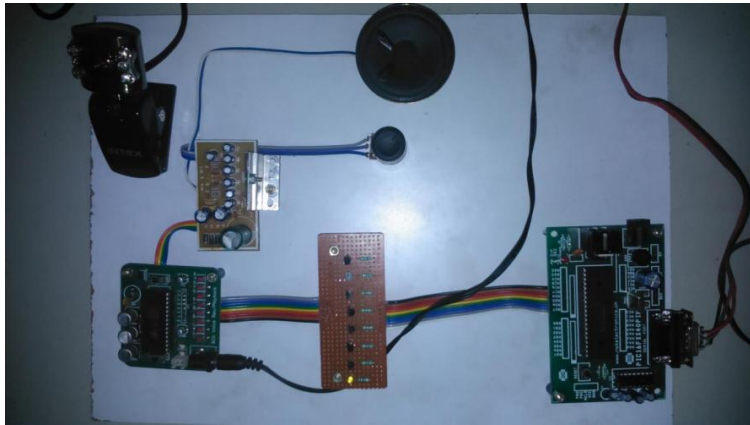


FIG:4 BLOCK DIAGRAM OF HAND GESTURE RECOGNITION

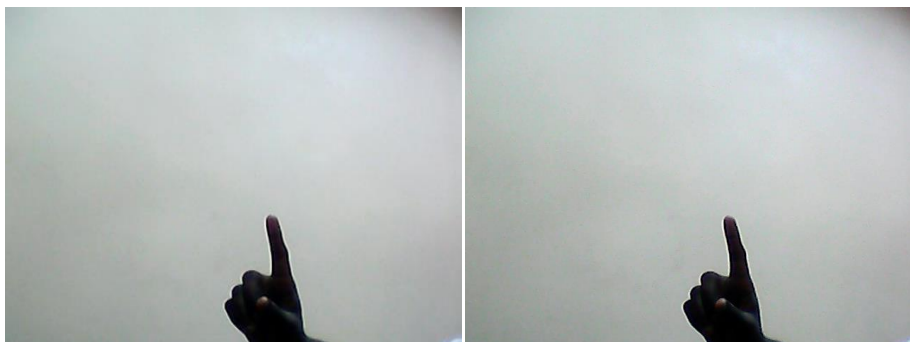


FIG :5. The hand that is clutteredby background can be detected accurately

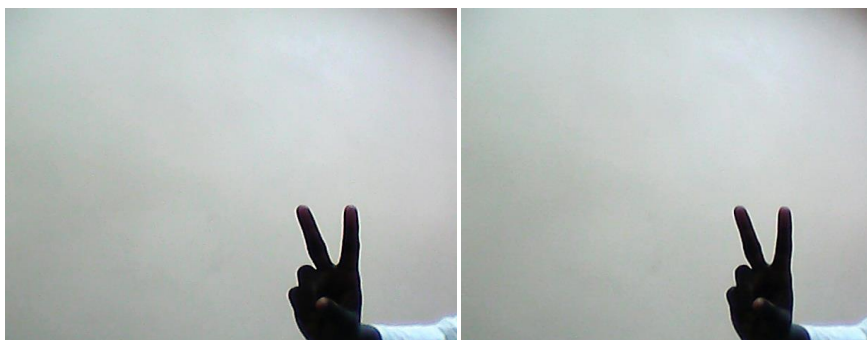


FIG :6.The hand that is cluttered by face can be detected accurately.

CONCLUSION

A robust part-based hand gesture recognition system using the Kinect sensor. A novel distance metric, Finger-Earth Mover's Distance (FEMD), is used for dissimilarity measure, which represents the hand shape as a signature with each finger part as a cluster and penalizes the empty finger-holes. FEMD based hand gesture recognition system achieves 93.2% mean accuracy and runs in 0.0750s per frame, when using the thresholding decomposition finger detection method. And it achieves better accuracy of 93.9% when using a more accurate finger detection method, however, at the cost of efficiency, thresholding decomposition for finger detection in our real time demo system. Furthermore it exploring a more efficient part-based representation to handle the problem and the efficiency drawback of near-convex decomposition based finger detection method. And we will further develop interesting HCI applications of our hand gesture recognition system. Another contribution is the real-life HCI applications on the top of our hand gesture recognition system. The hand gesture recognition technique can mimic the communications between human and involve hand gesture as a natural and intuitive way to interact with machines.

REFERENCES

1. Huang, T.S. (2011) 'CONDENSATION—Conditional density propagation for visual tracking,' Int. J. Comput. Vis. 'IEEE Trans. Inf. Theory, Vol.13, No.2, pp.83-87.
2. Haykin, S. (2012) Neural Networks: A Comprehensive Foundation. New York 'IEEE Trans. Pattern Anal. Mach. Comput. Vis. Image Understanding, Vol.104, No.82, pp.107-204.
3. Isard, M. and Blake, A. (2013) 'Contour tracking by stochastic propagation of conditional density,' in Proc. Eur. Conf. Comput. Vis. Vol.88, No.54, pp.84-96.
4. Latecki, M. and Bai, X. (2010) 'Path similarity skeleton graph matching,' IEEE Trans. Pattern Anal. Mach. Intell, Vol. 89, No. 7, pp. 23–88.
20. Viterbi, A. K. (2010) 'Error bounds for convolution codes and an asymptotically optimum decoding algorithm,' IEEE Trans. Inf. Theory, Vol. 13, No.2, pp. 260–269.
5. Yamato, J. and Ohya, J. and Ishii, K. (2013) 'Recognizing human action in timesequential images using hidden Markov model,' in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, Vol. 73, No.10, pp. 660–2669.
6. Dr. AntoBennet, M, SankarBabu G, Natarajan S, "Reverse Room Techniques for Irreversible Data Hiding", Journal of Chemical and Pharmaceutical Sciences 08(03): 469-475, September 2015.
7. Dr. AntoBennet, M, Sankaranarayanan S, SankarBabu G, "Performance & Analysis of Effective Iris Recognition System Using Independent Component Analysis", Journal of Chemical and Pharmaceutical Sciences 08(03): 571-576, August 2015.
8. Dr. AntoBennet, M, Suresh R, Mohamed Sulaiman S, "Performance & analysis of automated removal of head movement artifacts in EEG using brain computer interface", Journal of Chemical and Pharmaceutical Research 07(08): 291-299, August 2015.
9. Dr. AntoBennet, M "A Novel Effective Refined Histogram For Supervised Texture Classification", International Journal of Computer & Modern Technology, Issue 01, Volume 02, pp 67-73, June 2015.

10. Dr. AntoBennet, M, SrinathR, RaishaBanuA, "Development of Deblocking Architectures for block artifact reduction in videos", International Journal of Applied Engineering Research, Volume 10, Number 09 (2015) pp. 6985-6991, April 2015.
11. AntoBennet, M & JacobRaglend, "Performance Analysis Of Filtering Schedule Using Deblocking Filter For The Reduction Of Block Artifacts From MPEQ Compressed Document Images", Journal of Computer Science, vol. 8, no. 9, pp. 1447-1454, 2012.
12. AntoBennet, M & JacobRaglend, "Performance Analysis of Block Artifact Reduction Scheme Using Pseudo Random Noise Mask Filtering", European Journal of Scientific Research, vol. 66 no.1, pp.120-129, 2011.

