

Dynamic Hand Gesture Recognition using Multi-Color Modules Segmentation Method and Artificial Neural Network

Faiza Mahmood Shuker
College of Medicine of Hamorabi, University of Babylon, Hillah, Iraq
faiza.mahmood1@yahoo.com

Abstract: Although, voice is main way used for communication, facial expressions and body language are also used while interacting with others. In our regular life, hand gestures are used to communicates with deaf and mute people by sign language, to express the feelings like ‘stop’ or ‘bye’. With the advancement in machine learning, artificial intelligence and applications like Human Computer Interaction (HCI), hand gestures are getting widely used to interact with machines and computers. This study presents the effective approach to identify dynamic hand gesture. Hand region detection is a vital task in dynamic hand gesture recognition module. For that purpose an improved method is introduced by combining HSV, YCgCr and YCbCr color spaces. After detection of hand portion from the video sequence their color texture and edge features are extracted. Then for gesture recognition an artificial neural network is used. Thus, this proposed system can be viewed as a complete dynamic hand gesture recognition system that can be used for various purposes.

Key words: Communication, facial expressions, hand gestures, Human Computer Interaction (HCI), HSV, artificial intelligence

INTRODUCTION

Many industries have started to implement human computer interaction techniques to make machines more intelligent. There are some entertaining applications like video games where touch-less consoles are made which operate by various hand gestures. Also in daily life, in offices, homes, transport vehicles, gesture recognition system reduce the appliances like remote controls, buttons and increase the usability. This hand gesture recognition technology has reduced a lot of manual labour and makes living a lot easier.

Hand gestures are the movements done by hand to express the meaningful information that can be a human feelings or thoughts. Gesture recognition is the process of identifying hand movements performed by human, so that, machine can carry out the equivalent action. It is a process of understanding, analysing and interpreting meaningful movements of hand. This is an atural way of human-machine interaction. To recognise any hand movement it is necessary to train the system by providing gestures with their meaning, so, after that it will automatically identify the correct gesture meaning. The objective of any hand gesture recognition is to make a system that has capacity to identify hand movement provided by a person and then use it to control the machine. Hand gesture are categorised into.

Static gesture: A static hand gesture also called as hand posture or hand poses are the hand configuration whose position does not methods are used for extracting features of image. First method used is color structure descriptor with HMMD color space and second feature extraction is edge map extraction using edge his to gram descriptor. Change with respect to time period and that can be shown by single image.

Dynamic gesture: A dynamic gesture is the hand configuration whose position changes continuously with respect to time and they are shown by sequence of various image frames. This research develops an algorithm for effective hand gesture recognition with good accuracy. This involves:

- Acceptance of video frames sequence
- Hand segmentation
- Feature extraction from segmented region
- Then identify the correct label for that hand gesture

Segmentation is a process of capturing only interested region from the image. The aim of segmentation process is to change the orientation of an image to make it something more informative and easier to analyze by simplifying the image. Some existing methods for segmentation are TSL skin color model for segmentation

(Mo *et al.*, 2011) and Kalman filter, adaptively background subtraction with help of threshold based on skin color (Elsayed *et al.*, 2015) histogram based segmentation (Kshirsagar and Shinde, 2017) and object plane generation segmentation and there are other methods such as Thresholding methods like color-based segmentation, Otsu's method like k-means clustering, texture methods using texture filters. In hand gesture recognition hand portion will be interested region, so, try to capture that region. Skin color is the most important feature of hand, so, we can extract hand portion by using skin color detection algorithm. It is considered that skin colour identification is the popular process for the detection of skin region as it is simple and faster (Mohanty and Raghunadh, 2016). Therefore, in this study for skin color detection based segmentation a new algorithm is proposed by combining three color spaces YCbCr, HSV and YCgCr.

Feature extraction is of the crucial task in the process as it helps to detect accurate hand gesture. Feature extraction is like a data reduction process which accepts the 2D image array as an input and gives feature vector output. Identifying which features has to select from image such that it will give the most descriptive information about the image such as edges, corners, ridges, blobs, etc. is the important task. The mostly used feature extraction techniques are template matching, unitary image transforms, gabor features, deformable templates, graph description, contour profiles, zoning, projection histograms, geometric moment invariants, spline curve approximation, fourier descriptors, zernike moments and gradient feature (Kumar and Bhatia, 2014). Here, two last part in hand gesture detection process is gesture identification. After extracting appropriate features from the image the gesture can be identified using different classifiers or using machine learning techniques. The common methods used for gesture recognition are template matching, statistical matching, dictionary look-up, neural network, linguistic matching and ad hoc method (Pranjali and Ubale, 2015). Gesture identification needed classification algorithm to classify features according to their common attributes. In this project, the feed forward neural network is used for identification of gesture.

Literature review: Considerable efforts have been made by many researchers to develop efficient techniques for each step of hand gesture recognition. Various video features are extracted to classify the data (Salahat and Qasaimeh, 2017). These features are used by different classification frameworks to identify the gesture of that

image. By Rosalina *et al.* (2017), researcher uses ANN (Artificial Neural Network) for classification whereas by Wu *et al.* (2016) segmentation and recognition of gesture done simultaneously using hierarchical dynamic framework which is a semi-supervised and based on a HMM (Hidden Markov Model) is proposed then DDNN (Deep Dynamic Neural Networks) is used for multi modal gesture identification.

Various feature extraction methods can also be used for the same image, for example, Ahmed (2012) uses rotation, translation, scaling and orientation in variant feature extraction method to extract the feature of the input image based on moment feature extraction method. That extracted features are then classified by neural network.

Hand region detection from the image is a key aspect for hand gesture detection. Detecting skin color from the image can provide better results for hand detection. By Muhammad and Abu-Bakar (2015), hybrid color spaces are used which include Hue Saturation Value (HSV) and luminance, chrominance in green, chrominance in red (YCgCr) color spaces. And the proposed algorithm provides well response to different skin color tones. Also Kolkur *et al.* (2017) uses RGB (Red, Green, Blue), HSV and YcbCr (Luminance, Chrominance) color models for skin detection for better accuracy in recognizing the skin area in a given image.

Reshna and Jayaraju (2017) proposed system for Indian sign language recognition system which uses skin segmentation and HoG features are extracted then they are classified using supervised machine learning based support vector machine. Researchers Salah NASR, Muhammad Shoaib, Kais Bouallegue and Hassen Mekki (NASR *et al.*, 2017) has implemented the system for face recognition in which bag features are extracted and multi class support vector machine is used to classify the face images.

By Wang *et al.* (2015) proposed a new method for HGRS which is based on the super-pixel using Kinect sensor. In this project, all images are acquired using Kinect sensor and for segmentation purpose they used depth information and skeleton features. As there is benefit of super-pixel that it retain depth, color and shape information of hand in super-pixels. A new identification distance measure known as earth mover's distance is planned which is based on the super-pixel demonstration. This method is robust to translation, scaling and to local distortions. For analyzing the results they experiment this method on the database which includes 20 posture and 10 different gestures from 5 different persons and they get 99.6% of accuracy. This method is also applied on public database and tested on 2 HCI applications.

MATERIALS AND METHODS

Proposed system: In order to recognise hand gesture a method is proposed for segmentation of hand region using threshold values of skin color from that segmented images we extract their shape and color texture features which will further use for classification. Figure 1 shows the system architecture of proposed system. The proposed model is divided into following steps:

Input: This system accepts dynamic hand gesture which represents some specific hand movement. Here, the cambridge hand gesture dataset is used where there are total 9 gestures and each gesture is represented by 16 image frames captured from video of showing hand movement. Both for training and testing these frames have to give as an input. The sample sequences of input gestures are shown in Fig. 2.

Pre-processing: While capturing the images by digital camera, light source may introduce color cast in the images which degrade their quality due to unwanted lightning conditions which make them less reproductive for further processing. It is not always possible to control lightning illumination condition, it is good idea to pre-process the images for color correction before

actual starting of image processing stages. The aim of pre-processing is to minimize the unwanted parameters of images or enhance features of images that made them more important for further processing.

Here, gray world algorithm is used for pre-processing. It is a color consistency algorithm to detect the true color of image which is independent of light source. According to gray world algorithm, the average value of the image should converge to the mean value of gray. Hence, this algorithm calculate the color of image considering unknown illumination of light by assuming that images have similar average gray values of all red, green and blue components. As a result, we get color cast free images which have color conditions closer to the original scene. This can be show mathematically by below Eq. 1:

$$r_i = \frac{\text{Max}(\text{avg}_R, \text{avg}_G, \text{avg}_B)}{\text{avg}_i} \tag{1}$$

Here:

- r_i = A scaled value along either red
- green or = Component and avg_i is channel mean along
- blue respective red, green or blue component

Segmentation: After the color balancing, we convert that image to YCbCr color space using in-build MATLAB function. YCbCr is the most popular color space for representation for digital video where Y represents the luminance, Cb shows the difference between blue component and reference value and Cr shows difference between red component and reference value. In this color space we use below threshold values for skin pixel detection:

$$77 \leq Cb \leq 127 \text{ and } 133 \leq Cr \leq 173 \tag{2}$$

With YCbCr another color space ‘HCgCr’ that is combination of HSV and YCgCr is proposed for skin pixels detection. For this purpose we conver RGB color image to HSV using MATLAB in-build function ‘rgb2hsv’. And red and green component from YCgCr components are calculated using following Eq. 3 and 4:

$$\text{Red component} = [(0.148 * R) - (0.291 * G) + (0.439 * B)] + 128 \tag{3}$$

$$\text{Green component} = [(0.439 * R) - (0.368 * G) - (0.071 * B)] + 128 \tag{4}$$

In this new color module skin pixels are masked using threshold values shown:

$$140 \leq Cg \leq 165 \text{ and } 140 \leq Cr \leq 195 \text{ and } 0.01 \leq H \tag{5}$$

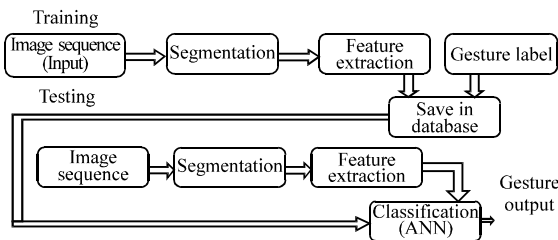


Fig. 1: Architecture diagram of proposed system

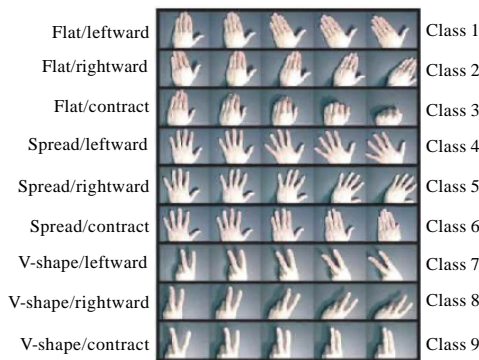


Fig. 2: Sample sequences of the 9 gesture classes

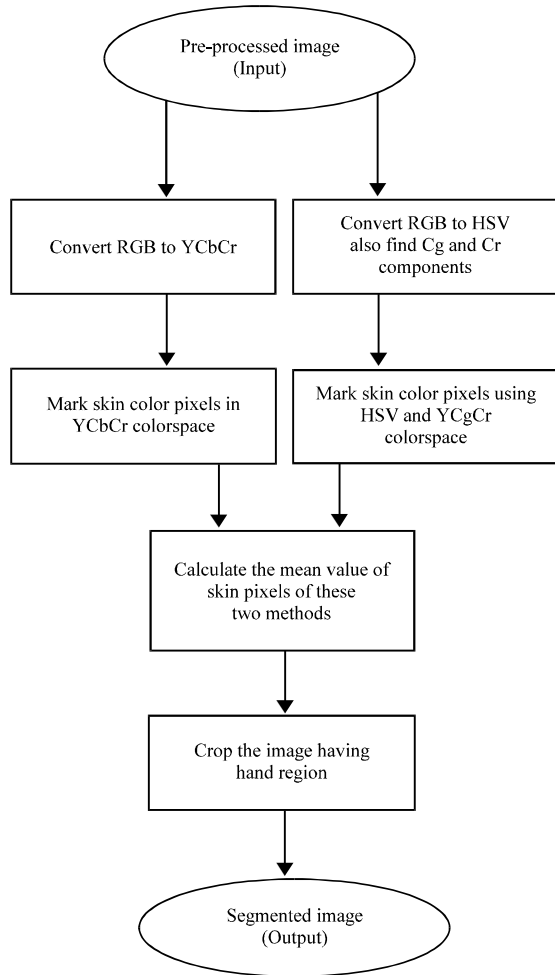


Fig. 3: Flow chart of proposed segmentation algorithm

After masking the skin pixel in these two different color spaces we find their mean values for better results. Then we extract only skin pixels, so that, we get only interested hand portion by subtracting non skin pixels portion. Figure 3 shows flowchart of proposed segmentation method.

Feature extraction: Here, two methods are used for extracting features of image. First method used is color structure descriptor with HMMD color space and second feature extraction is edge map extraction using edge histogram descriptor.

Color features: Structure descriptor is used to identify the local color pattern of the image using 8×8 structuring element. This method gives the color histogram which indicates number of count of particular number at each level of structuring element. Color features are calculated

using HMMD colorspace where Hue, Sum and Difference (Diff) are the three components are used. Hue is same as the component from the HSV color modal and has the value between $0-360^\circ$. Diff component indicates presence of gray color in image and how much purity presents to the original color giving the favour of tone. Difference component has value between $0-1$. And sum component also has a value between $0-1$, it shows how much brightness of color is present in the image. These components are obtained from the RGB color modal (Nasr *et al.*, 2017).

This color space is partitioned into 5 sub spaces (0-4) along the diff component axes and then it quantized using 128 bin along hue and sum axes. For each of five sub spaces the quantization levels for hue and sum components are: Hue (14888) and Sum (16444).

Shape features: EHD (Edge Histogram Descriptor) is used to calculate shape features of image sequence. In EHD local edge distribution in the image is calculated. Horizontal, vertical, diagonal and anti-diagonal edges are identified by dividing image into 4×4 grid (Wu *et al.*, 2001). Then for each such sub-image, edge histogram is generated to represent the edge distribution in each sub-space. To identify different edge types, each sub-space is further divided into small squares named as image-blocks.

As neural network cannot take the input in the image format, hence, the binary form of image is given as an input to this neural network. After the feature extraction process, the image is converted into 16×16 resolution means image contains total 256 pixels. Therefore, 256 inputs are accepted by the input layer of neural network. And the output layer contains 9 neurons representing 9 different classes of hand gestures. Hidden layer is made up of number of neurons by taking mean value of neurons from input and output layer. While testing the image, whichever value finds largest among 9 classes is predicted as output gesture for that image.

RESULTS AND DISCUSSION

The proposed system uses combination of YCbCr, YCgCr and HSV colors for segmentation of images. These segmented images are further used in feature extraction algorithm and then it is classified using artificial neural network. After the proposed segmentation method performance of the system is measured on the basis of results given by ANN.

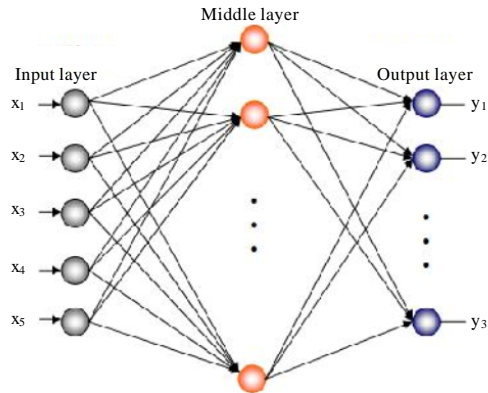


Fig. 4: Three layers feed forward neural network

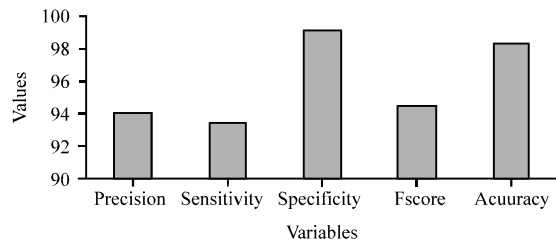


Fig. 5: Average performance measure of the proposed system; Classification results with proposed multiple color space algorithm for segmentation

Classification: Classification is done using un-supervised feed forward Artificial Neural Network (ANN). It is one of the popular machine learning techniques. It is inspired from the human brain system which copy the way that human learns. It is a mathematical model consists of layers that are input layer, hidden layer and output layers as shown in Fig. 4. Each layer is consisting of neurons that accept the input and transfer it into something that next layer can understand.

To analyze the accuracy of proposed system, this system is tested on the Cambridge hand gesture database. This database consists of image sequences for single gesture representation which are captured from the video of a person. This database consists of total 9 different gestures. Each such gesture class has 22 folders with 15 frames of images. Out of these 22 folders, 13 folders of each gesture are used for training purpose. That means out of total 198 folders, 117 folders are used for training. And all 198 folders are tested for evaluating the results. Table 1 shows the confusion matrix and Fig. 5 shows the bar chart of performance of system.

Table 1: Confusion matrix for classification with proposed segmentation method using multiple color spaces

Class	1	2	3	4	5	6	7	8	9
1	22	0	0	0	0	0	0	0	0
2	0	22	0	0	0	0	0	0	0
3	0	0	22	0	0	0	0	0	0
4	0	0	0	19	0	3	0	0	0
5	0	0	0	1	20	1	0	0	0
6	0	0	0	3	0	19	0	0	0
7	0	0	0	0	0	0	22	0	0
8	0	0	0	0	0	0	1	21	0
9	0	0	0	0	0	0	4	0	18

CONCLUSION

Thus proposed approach addresses the challenge of recognition of hand gesture. Here, the effective segmentation method is used by detecting skin portions for the images which is extracting more useful information that proceeded for further feature extraction process from which edge map and color structure descriptor features are extracted, finally, artificial neural network is used for classification and identification of gestures. This proposed method is suitable for vision-based dynamic hand gesture recognition. This system is providing overall good results.

REFERENCES

Ahmed, T., 2012. A neural network based real time hand gesture recognition system. *Intl. J. Comput. Appl.*, 59: 17-22.

Elsayed, R.A., M.S. Sayed and M.I. Abdalla, 2015. Skin-based adaptive background subtraction for hand gesture segmentation. *Proceedings of the 2015 IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, December 6-9, 2015, IEEE, Cairo, Egypt, ISBN:978-1-5090-0246-7, pp: 33-36.

Kolkur, S., D. Kalbande, P. Shimpi, C. Bapat and J. Jatakia, 2017. Human skin detection using RGB, HSV and YcbCr color models. *Adv. Intell. Syst. Res.*, 137: 324-332.

Kshirsagar, K.P. and R.A. Shinde, 2017. Comparing techniques of segmenting hand region. *Proceedings of the 2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, October 19-20, 2017, IEEE, Coimbatore, India, ISBN:978-1-5090-5014-7, pp: 721-724.

Kumar, G. and P.K. Bhatia, 2014. A detailed review of feature extraction in image processing systems. *Proceedings of the 2014 4th International Conference on Advanced Computing and Communication Technologies*, February 8-9, 2014, IEEE, Rohtak, India, ISBN:978-1-4799-4910-6, pp: 5-12.

- Mo, S., S. Cheng and X. Xing, 2011. Hand gesture segmentation based on improved kalman filter and TSL skin color model. Proceedings of the 2011 International Conference on Multimedia Technology, July 26-28, 2011, IEEE, Hangzhou, China, ISBN:978-1-61284-771-9, pp: 3543-3546.
- Mohanty, R. and M.V. Raghunadh, 2016. Skin color segmentation based face detection using multi-color space. Intl. J. Adv. Res. Comput. Commun. Eng., 5: 470-475.
- Muhammad, B. and S.A.R. Abu-Bakar, 2015. A hybrid skin color detection using HSV and YCgCr color space for face detection. Proceedings of the 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), October 19-21, 2015, Kuala Lumpur, Malaysia, ISBN:978-1-4799-8996-6, pp: 95-98.
- Nasr, S., K. Bouallegue, M. Shoaib and H. Mekki, 2017b. Face recognition system using bag of features and multi-class SVM for robot applications. Proceedings of the 2017 International Conference on Control, Automation and Diagnosis (ICCAD), January 19-21, 2017, IEEE, Hammamet, Tunisia, ISBN:978-1-5090-5988-1, pp: 263-268.
- Pranjali, S. and V.S. Ubale, 2015. Hand gesture recognition systems: A survey. Intl. J. Inventive Eng. Sci., 3: 1-5.
- Reshna, S. and M. Jayaraju, 2017. Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM. Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), March 22-24, 2017, IEEE, Chennai, India, ISBN:978-1-5090-4443-6, pp: 386-390.
- Rosalina, L. Yusnita, N. Hadisukmana, R.B. Wahyu and R. Roestam *et al.*, 2017. Implementation of real-time static hand gesture recognition using artificial neural network. Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), August 8-10, 2017, IEEE, Kuta Bali, Indonesia, ISBN:978-1-5386-0601-8, pp: 1-6.
- Salahat, E. and M. Qasaimeh, 2017. Recent advances in features extraction and description algorithms: A comprehensive survey. Proceedings of the 2017 IEEE International Conference on Industrial Technology (ICIT), March 22-25, 2017, IEEE, Toronto, Canada, ISBN:978-1-5090-5321-6, pp: 1059-1063.
- Wang, C., Z. Liu and S.C. Chan, 2015. Superpixel-based hand gesture recognition with kinect depth camera. IEEE. Trans. Multimedia, 17: 29-39.
- Wu, D., L. Pigou, P.J. Kindermans, N.D.H. Le and L. Shao *et al.*, 2016. Deep dynamic neural networks for multimodal gesture segmentation and recognition. IEEE. Trans. Pattern Anal. Mach. Intell., 38: 1583-1597.
- Wu, P., Y.M. Ro, C.S. Won and Y. Choi, 2001. Texture descriptors in MPEG-7. Proceedings of the International Conference on Computer Analysis of Images and Patterns, September 5-7, 2001, Springer, Berlin, Heidelberg, Germany, ISBN:978-3-540-42513-7, pp: 21-28