

Machine Control Using Hand Gesture Recognition

¹QURRAT-UL-AIN, ¹SHAH NAWAZ TALPUR, ¹NOOR-UZ-ZAMAN LAGHARI,
AND ¹MUHAMMAD SAEED RATTAR

¹Institute of Information & Communication Technologies, Mehran University of
Engineering & Technology, Jamshoro, Pakistan.

qurratulain.rattar@gmail.com, shahnawaz.talpur@faculty.muets.edu.pk,
nooruzaman@gmail.com, saeedrattar@hotmail.com

Abstract

Human to machine collaboration is an important game changer in industries as well as in daily lives. Hand gesture recognition assists this interaction by automating the process. This collaboration is supported by proposing system which can recognize human hand gestures and perform tasks based on those gestures. This paper presents the system that uses hand gesture recognition to control machine. The machine executes programmed task according to gestures recognized. The hand gestures are recognized using vision-based approach. The defined gestures are static and carried out by bare hands. The recognition of the system involves hand segmentation, feature extraction, and classification. The hand is segmented using bounding box in webcam preview, and features are extracted by determining signature of hand and histogram of signature. The classification process uses supervised feed forward neural network system to train some samples of different gestures. Once gesture is recognized, the programmed tasks start to execute. The proposed system has found to have an average accuracy of 92.6% for the classification of gestures.

Key Words: Gesture Recognition, Human Computer Interaction, Hand Detection, Feature Extraction, Neural Networks.

I. INTRODUCTION

Gesture recognition and interpretation of those gestures to control the machines is gaining significant attention due to its application in numerous areas such as human computer interaction, gaming, robot control, automobiles, visual environments, television control, sign language interpretation, etc. A gesture can be defined as expressive form of non-verbal communication with different body parts such as hands, face, limbs, head or body with the intention to interact with the environment [1]. Gesture recognition is the ability of machines to recognize the gestures and execute commands based on those gestures. There are two approaches to gesture recognition; hardware based approach and vision based approach. Hardware based approach requires user to wear some bulky device such as data glove or computing device. Although it provides accuracy, it is not practical to wear cumbersome device

as it hinders the natural approach for human computer interaction. Vision based approach uses camera and computer vision techniques to interpret gestures which is more natural to humans. [2]. In this research, proposed system adopts vision based approach to recognize static hand gestures and perform meaningful commands based on the interpretation of those gestures.

This paper is organized into five sections. Section II describes previous work that has been carried out for machine control using gesture recognition. Section III explains the methodology. The results are presented and discussed in Section IV. Finally, the conclusions drawn based on the results and discussed in Section V.

II. RELATEDWORK

There has been an active research in the area of gesture recognition that allow gesture enabled command implementation in recent years.

The researches in [3-6] emphasize on efficient hand gesture recognition algorithm but do not utilize the recognized gestures into functional inputs. Some research has been done on fuzzy logic for the recognition using a learning and feedback procedures. Gesture is recognized by capturing the image, extracting its features that are learned by merely adding it to the database and then matching with the new gestures. The recognition rate is reliant on the quality of images and the type of hardware used [7].

Ren and Zhang [8] classify static hand gestures by learning the Fourier descriptor of contour line. The hand is segmented by mean shift algorithm. Nonetheless, this identification is computationally rigorous. Joseph and LaViola [9] classify gestures that are sensitive to illumination conditions which yields erroneous segmentation of the hand region. Shin et. al., [10] have carried hand segmentation by selecting high entropy regions with near-skin color distribution and recognition is performed through entropy analysis.

Kim and Fellner [11] have developed hand gesture recognition system using back-projection wall environment which is vision-based approach but they have used florescent white paper to mark finger tips in the captured image, and it is not practical to wear white florescent strips.

Jing et. al. [12] proposed algorithm that extracts morphological features and classifies different gestures using AdaBoost algorithm. AdaBoost is a cumbersome method and can be sensitive to noisy data and outliers

Acharya [13] proposed an algorithm in which features of gestures extracted using PCA (Principle Component Analysis) method and feature matching is carried out with the help of template matching. PCA does not have good track record when it comes to feature extraction and it is highly dependent upon linearity. Similarly, template

atching does not work well on scaling. The authors in [14] proposed a way in which features are extracted using discrete wavelet transformation decomposition, and SVM (Support Vector Machine) classifier is used to classify various gestures.

SVM is a non-linear classifier in which technique is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification [12].

However, restriction of the SVM is the large number of support vectors used from the training set to perform classification.

It is, therefore, necessary to develop methods that may perform recognition more accurately and are less computationally intensive. This paper proposes a system that minimizes the limitations of previous work discussed above.

III. METHODOLOGY

The proposed system comprises of various steps discussed below. The flowchart of the system is represented in Fig. 1.

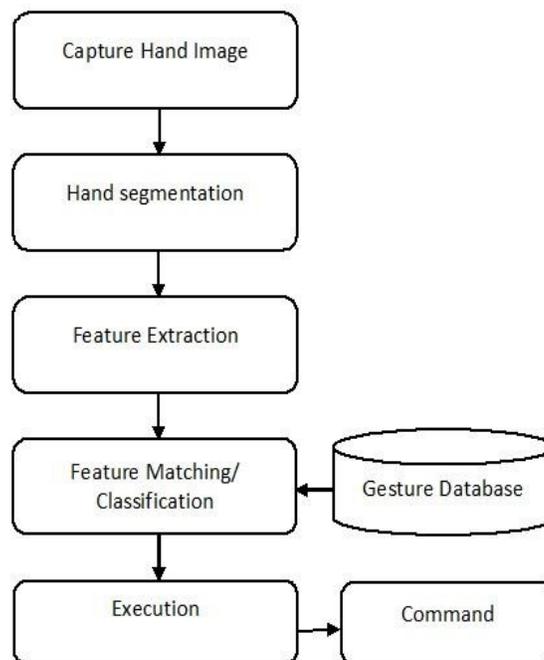


FIG. 1. METHODOLOGY

Hand Segmentation: After capturing images of hand, hand segmentation is the first step. It is carried out to handle the challenges of vision-based system such as proper illumination conditions, complex background removal and variable scaling. In the

proposed system, segmentation is performed using bounding box in webcam preview. The bounding box help to get ROI (Region of Interest) which, in our case, is hand. The hand is placed in a bounding box and the portion of image inside the bounding box is cropped. The cropped image is of hand gesture whose features need to be extracted.

Feature Extraction: Feature extraction is the process by which new data set or image is generated from the existing image by retaining their significant features [15]. Usually, features are found from the segmented image, which could be anything such as mean, centroid, area, edge, histogram, etc. of the image. Feature extraction is performed to reduce dimensionality and hence computational complexity. The steps involved in the feature extraction of image are shown in Fig. 2.

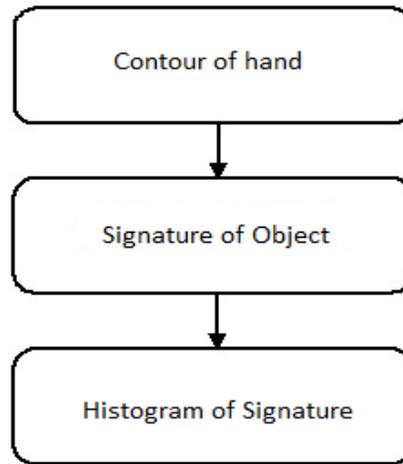


FIG. 2. FEATURE EXTRACTION

First, the contour of the segmented image is determined. YCbCr skin color model is used to detect skin for the initial mask of contour. YCbCr skin thresholds are mentioned in form of Equations (1-3):

$$80 < Y \tag{1}$$

$$85 < Cb < 135 \tag{2}$$

$$135 < Cr < 180 \tag{3}$$

Transformation from RGB (Red, Green Blue) color space to YCbCr is done through the following expression in Equation (4-6):

$$Y = C1 R + C2 G + C3 B \tag{4}$$

$$Cb = \frac{(B - Y)}{(2 - 2 \times C3)} \tag{5}$$

$$Cr = \frac{(R - Y)}{(2 - 2 \times C1)} \tag{6}$$

Where C1 is 0.2989, C2 is 0.5866, C3 is 0.1145 for standard images.

Then the images were converted into grayscale images using the equation stated in Equation (7).

$$0.2989R + 0.5870G + 0.1140B \quad (7)$$

After getting grayscale images, the threshold is determined using Otsu's method.

In Otsu's method, threshold is considered to be that value which has the minimum variance σ^2_w within the class.

$$\sigma^2_w = W_b\sigma^2_b + W_f\sigma^2_f \quad (8)$$

Where σ^2 represents the variance of two classes, "W" represents weight and subscript "b" and "f" represents foreground and background respectively.

$$W_b = \frac{\sum_{i=1}^n h_i}{l} \quad (9)$$

Where h is histogram and l is maximum level.

$$\mu_b = \frac{\sum_{i=1}^n (ind_i h_i)}{\sum_{i=1}^n h_i} \quad (10)$$

$$\sigma^2_b = \frac{\sum_{i=1}^n ((ind_i - \mu_b)^2 h_i)}{\sum_{i=1}^n h_i} \quad (11)$$

Foreground variance was calculated in the same manner. If $g(x,y)$ is a threshold version of $f(x,y)$ at some global threshold T, then image pixel values are converted into binary values using threshold T as:

$$g(x,y) = \begin{cases} 1, & \text{if } f(x,y) > T \\ 0, & \text{Otherwise} \end{cases} \quad (12)$$

After determining the contour of hand, signature of hand is obtained by calculating the distance from centroid to the points that form the boundary. The centroid is calculated by finding the indices of columns of non-zero pixel and indices of rows of non-zero pixel of binary image.

If indices of columns are $x_1, x_2, x_3 \dots x_n$ and indices of rows be $y_1, y_2, y_3 \dots y_n$, the centroid of binary image can be calculated by using Equation (13).

$$C(x,y) = \left(\frac{x_1 + x_2 + x_3 + \dots + x_n}{n}, \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} \right) \quad (13)$$

The distances from centroid to the boundary of hand are calculated using Equation (14).

$$d = \sqrt{(x - b_{n,x})^2 + (y - b_{n,y})^2} \quad (14)$$

Where $b_{n,x}$ and $b_{n,y}$ are x and y coordinates of boundary respectively and $n= 1,2,3,4 \dots$

After getting distance value, the histogram of the signature of hand is determined, as shown in Fig.3, which provides 360 values. Those 360 values work as the elements of feature vector.

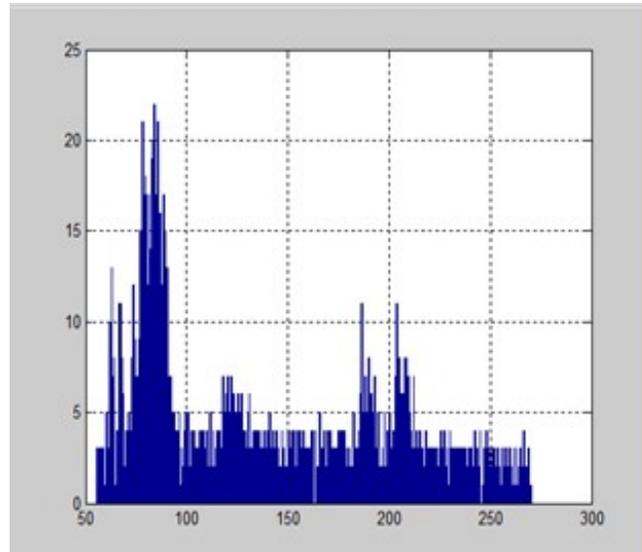


FIG. 3. HISTOGRAM OF SIGNATURE

Classification: The next step after hand segmentation and feature extraction is to perform classification. To perform classification, ANN (Artificial Neural Networks) is used as the classifier. The best network for the NN (Neural Network) is generally obtained by trial and error method. After running several simulations, it is considered suitable to use a two-layer FNN (Feedforward Neural Network) to train the gestures with 10 hidden neurons, 360 input values and 3 output values as shown in Fig. 4.

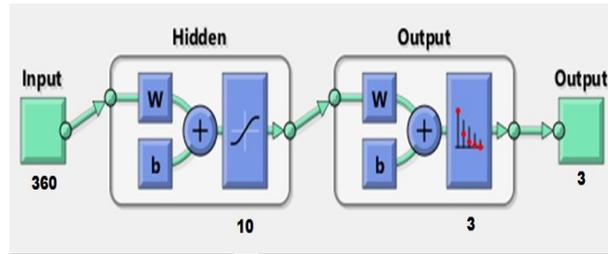


FIG.4. FEED FORWARD NETWORK

Twelve (12) number of samples are used to train single gesture and 3 gestures are classified.

The confusion matrix plot for the classification is shown in Fig. 5, in which the rows correspond to the predicted class (Output Class) and the columns correspond to the true class (Target Class). Further, the diagonal cells show the number and percentage of gestures correctly classified by the trained network. In the first column, 12 out of 12 samples are classified correctly which corresponds to 33.3% of all gestures. In the second column 11 out of 12 samples are classified correctly which corresponds to 30.6% of all gestures. Similarly, in the third column, 12 out of 12 samples are classified appropriately which corresponds to 33.3% of all gestures. Overall, 97.2% of the predictions are correct and 2.8% are incorrect.

	1	2	3	
1	12 33.3%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	11 30.6%	0 0.0%	100% 0.0%
3	0 0.0%	1 2.8%	12 33.3%	92.3% 7.7%
	100% 0.0%	91.7% 8.3%	100% 0.0%	97.2% 2.8%
	1	2	3	

FIG. 5. CONFUSION MATRIX

IV. RESULTS AND DISCUSSION

The proposed system uses 12 samples of each gesture for the training. The total number of tested samples are 250. In case of Gesture 1, 230 samples are correctly classified (92%). While in case of Gesture 2, 240 samples are correctly classified (96%). Similarly, in Gesture 3, 225 samples are correctly classified (90%). Thus, the

average accuracy is 92.6%. The total number of trained and tested samples are shown in Table 1.

TABLE 1. TOTAL TRAINED AND TESTED NUMBER OF SAMPLES			
Gesture	Total Tested Samples	Trained Number of Samples	Correctly Classified Samples
1	250	12	46
2			48
3			45

Similarly, the percentages of correct and incorrect classifications are shown in Table 2.

TABLE 2. PERCENTAGES OF CORRECT AND INCORRECT CLASSIFICATIONS			
Classes	G1 (%)	G2 (%)	G3 (%)
Correctly Classified Gestures	92	96	90
Incorrectly Classified Gestures ECGs	9	5	10

Fig.6 demonstrates the GUI (Graphical User Interface) developed for the proposed system in which an input gesture is captured by the camera and its features are extracted. Contour and histogram are obtained in feature extraction process. Classification is carried out after feature extraction and programmed task is executed on recognition of gesture.

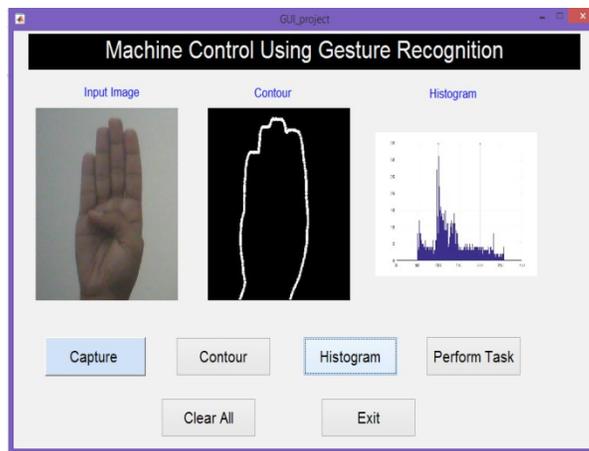


FIG. 6. GRAPHICAL USER INTERFACE

The programmed tasks for the different gestures are shown in Fig. 7.

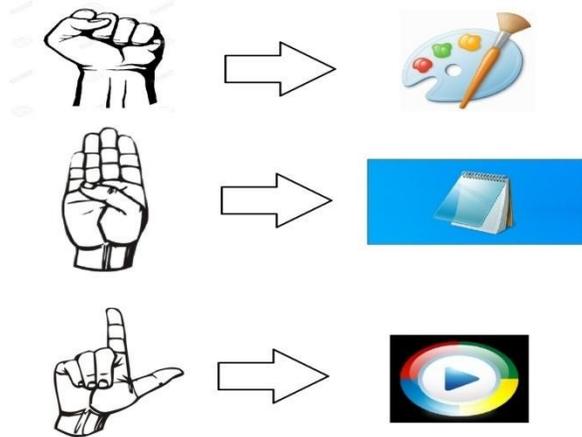


FIG. 7. PROGRAMMED TASKS

V. CONCLUSION

In the proposed system, the gestures are recognized to perform programmed tasks using vision-based recognition approach. The system is cost effective as it avoids the use of sensor-based recognition which require expensive sensitive sensors. The steps of recognition involve hand segmentation, feature extraction, and classification. On recognition of gestures, programmed tasks are executed. Such system can be embedded in many areas of gesture-controlled applications such as robotics, automobiles, mobile applications, television control, gaming, etc. and is scalable enough to add new gestures and commands. Integrating such systems in machines will allow humans to command them via hand gestures.

VI. ACKNOWLEDGMENT

This research was supported by Mehran University of Engineering & Technology, Jamshoro, Pakistan.

VII. REFERENCES

- [1] Mitra, S., and Acharya, T., "Gesture Recognition: A Survey", IEEE Transactions on Systems, Man, and Cybernetics Part-C, Applied Review, Volume 37, No. 3, pp. 311-324, 2007.
- [2] Lee, H.B., and Lim, H., "Hand Tracking and Gesture Recognition System for Human-Computer Interaction Using Low-Cost Hardware", Multimedia Tools and Applications, Volume 74, No. 8, pp.2687-2715, 2013.
- [3] Burns, A.M., and Mazzarino, B., "Finger Tracking Methods Using Eyesweb", Gesture in Human-Computer Interaction and Simulation, pp. 156-167, 2006.
- [4] Dias, J.M.S., Nande, P., Barata, N., and Correia, A., "OGRE: Open Gestures Recognition Engine", IEEE Proceedings on 17th Brazilian Symposium on Computer Graphics and Image Processing, pp. 33-40, 2004.

- [5] Han, S.I., Mi, J.Y., Kwon, J.H., Yang, H.K., and Lee, B.G., “Hand Tracking and Gesture Recognition System for Human-Computer Interaction Using Low-Cost Hardware”, Vision Based Hand Tracking for Interaction, 2008.
- [6] Hasan M.M., and Mishra, P.K., “Real Time Fingers and Palm Locating Using Dynamic Circle Templates”, International Journal of Computer Applied, Volume 41, No. 6, pp. 33-43, 2012.
- [7] Bimber, O., “Continuous 6DOF Gesture Recognition: A Fuzzy Logic Approach”, Proceedings of WSCG’99, Volume 1, pp. 24-30, 1999.
- [8] Ren, Y., and Zhang, F., “Hand Gesture Recognition Based on Meb-SVM”, IEE 2nd International Conference on Embedded Software and Systems, Computer Society, Los Alamitos, pp. 344–349, 2009.
- [9] Joseph, J., and LaViola Jr., “A Survey of Hand Posture and Gesture Recognition Techniques and Technology”, Master Thesis, Science and Technology, Center for Computer Graphics and Scientific Visualization, USA, 1999.
- [10] Shin J.-H., Lee, J.-S., Kil, S.-K., Sehn, D.-F., Ryn, J.-G., Lee, E.-H., Min, H.-K., and Hon, S.-H., “Hand Region Extraction and Gesture Recognition Using Entropy Analysis”, International Journal of Computer Science and Network Security, Volume 6, Issue 2A, February, 2006.
- [11] Kim, H., and Fellner D.W., “Interaction with Hand Gesture for a Back-Projection Wall”, Proceedings of Computer Graphics International, pp. 395-402, 19 June, 2004,
- [12] Sun, J., Ji, T., Zhang, S., Yang, J., and Ji, G., “Research on the Hand Gesture Recognition Based on Deep Learning. 12th International Symposium on Antennas, Propagation and EM Theory, 2018.
- [13] Acharya, R.U., “Advances in Cardiac Signal processing”, Berlin, Springer, 2007.
- [14] Agarwal, R., Raman, B., and Mittal, A., “Hand Gesture Recognition Using Discrete Wavelet Transform and Support Vector Machine”, 2nd International Conference on Signal Processing and Integrated Networks, 2015.
- [15] Kumar, G., and Bhatia, P., “A Detailed Review of Feature Extraction in Image Processing Systems”, 4th International Conference on Advanced Computing & Communication Technologies, [DOI: 10.1109/ACCT.2014.74], 2014.