I.J. Information Technology and Computer Science, **2019, 12, 50-56**
Published Online December 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2019.12.05

An Efficient Framework for Hand Gesture Recognition based on Histogram of Oriented Gradients and Support Vector Machine

Ahmed Abdal Shafi Rasel
Department of Computer Science and Engineering, Stamford University Bangladesh
E-mail: shafi.cse@stamforduniversity.edu.bd

Mohammad Abu Yousuf
Institute of Information Technology, Jahangirnagar University, Bangladesh
E-mail: yousuf@juniv.edu

Received: 12 September 2019; Accepted: 15 October 2019; Published: 08 December 2019

***Abstract*—**This paper focuses on an empirical hand gesture recognition system in the domain of image processing and machine learning. The hand gesture is probably the most intuitive and frequently used mode of nonverbal communication in human society. The paper analyzes the efficiency of the Histogram of Oriented Gradients (HOG) as the feature descriptor and Support Vector Machine (SVM) as the classification model in case of gesture recognition. There are three stages of the recognition procedure namely image binarization, feature extraction, and classification. The findings of the paper show that the model classifies hand gestures for the given dataset with satisfactory efficiency. The outcome of this work can be further utilized in practical fields of real-world applications dealing with non-verbal communication.

***Index Terms*—**Hand Gesture Recognition, Support Vector Machine, Histogram of Oriented Gradients, Human-Computer Interaction.

I. Introduction

By providing a natural and intuitive way of interaction and communication, hand gesture-based methods stand out from other approaches to Human-Computer Interaction. Efficient hand gesture recognition methods have therefore become much more important.

The motivation of the work comes from the need to understand nonverbal communications across different social contexts. It has many practical purposes like sign language interpretation, body language interpretation, etc. Non-linguistic communication involves various modes like gestures and eye-contacts. The most naturalistic and convenient way of communicating simple expressions are done through hands. That's why understanding gestures are so important.

We will concentrate on a specific technique of gesture recognition in this paper and analyze its effectiveness. Following the preprocessing procedure, we extracted features from the training images using Histogram of Oriented Gradients (HOG). Different methods based on SVM are then used in the classification of binary and multiclass datasets. The system model proposed in this paper can be applied as part of their system in many practical applications requiring gesture recognition.

The rest of the paper is organized as follows. Section 2 describes the literature review, section 3 demonstrates the methodology based on the configuration of the proposed hand gesture recognition system. Section 4 shows the experimental results & discussion and conclusions of our work are drawn in Section 5.

II. Literature Review

Many researchers have worked together on the topic of service robots accepting requests through nonverbal communication [1] and hand gesture recognition of requests [2]. Before this interaction between caregivers and elderly people was analyzed from a sociological point of view to serve as a model to be emulated in service robots[3].

In the classification task, the extraction of features is a primary task. In the field of computer vision, several feature extraction algorithms are used such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and Independent Component Analysis (ICA). In several works, Scale-invariant Feature Transform (SIFT)[4], Speeded Up Robust Features (SURF)[5] and Histogram Oriented Gradient (HOG)[6] were also used. Both PCA and SVD reduces the dimensionality of the feature space. It projects the data into a smaller space making it easier to focus on the important parts. SURF and HOG are known to be susceptible to the rotation of objects which SIFT can deal with very swiftly. SIFT is also good at illumination changes but can be slow. This drawback can be remedied by adding variety to the images in the database[7]. Different features extraction methods can be further added up to so what the strengths of both can be combined. However, that can also add some complexity concerns.

Support Vector Machine (SVM)[8], Neural Network[9], HMM, fuzzy c-means[10] clustering were used as computer vision classification methods. The technique based on the Neural Network needs big sets of training data[9]. Although it is nice to recognize dynamic gestures[11], HMM is also computationally expensive[12]. Some studies also suggest using fuzzy logic[13] and Kalman filtering to recognize gestures[14].

Neural Networks use biologically inspired models to map input data to target classes. Different types of structures and models can be employed based on the problem type. Deep Neural Networks like Convolutional Neural Networks (CNNs) are used in case of computationally intensive tasks. Hidden Markov Model is very useful in case of finding a data sequence that isn't easily visible. HMM is basically a statistical model that is capable of predicting the class labels on a scale of probability. Fuzzy c means clustering provides a mechanism that permits one piece of data to belong to more than one class. Support Vector Machines (SVMs) uses a separating hyperplane to classify between classes. However, when the data is not so linearly separable, we may need to use kernel tricks.

Among many feature extraction methods and classification techniques used in computer vision, some of the methods over-perform other methods with substantial efficiency. We have studies different classes of problems and how are they have been solved in many different works. In our case, we have decided to use the Histogram Oriented Gradient (HOG) as our feature extraction method and Support Vector Machine (SVM) as the classifier for hand gesture identification. The SVM is tuned using Error-Correcting Output Codes (ECOC) which is an ensemble method. We will discuss in more detail the proposed model and how it works in the next chapter.

III. Methodology

The suggested technique of recognition comprises of picture binarization (pre-processing), extraction of features using Histogram of Oriented Gradients (HOG), and classification of gestures using a Support Vector Machine (SVM). The feature extraction using HOG takes into consideration different parameters and how we can modulate the extraction method for better optimization. The SVM can be used for both the binary classification of gestures as well as multiclass classification we will discuss in detail. Fig. 1 illustrates the general flow of the suggested scheme for recognizing hand gestures.

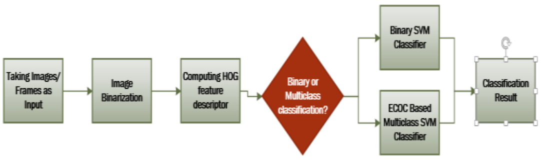
Taking Images/ Frames as Input
Image Binarization
Computing HOG feature descriptor
Binary or Multiclass classification?
Binary SVM Classifier
ECOC Based Multiclass SVM Classifier
Classification Result

Fig.1. Overall flow of the proposed hand gestures recognition system

A. Image Binarization

We need pre-processing before extracting characteristics from training dataset pictures. Pre-processing significantly improves picture analysis reliability. We regarded the picture binarization to be the first step forward. The picture is transformed into a binary image from RGB to grayscale. Conversion of a color picture to a grayscale, however, is not unique; the different weighting of color channels effectively reflects the effect of shooting a black and white film on cameras with different color photographic filters.

In our work, the values of the grayscale are calculated by using the equation (1) to form a weighted sum of the components R, G, and B.

$$Y = 0.2989 * R + 0.5870 * G + 0.1140 * B \qquad (1)$$

Lastly, the grayscale values are transformed to form a binary image. Fig. 2 demonstrates the binarization method. The purpose of this step is to enhance image information that suppresses unwanted distortions or enhances some significant image characteristics for further processing.

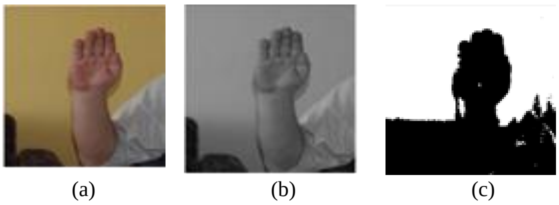
(a) (b) (c)

Fig.2. Binarization process (a) Original RGB image (b) Grayscale image (c) Binary image

B. Feature Extraction Using Histogram of Oriented Gradients

After preprocessing, we used HOG descriptors to extract features from binary images. HOG is a feature descriptor introduced in 2005 by Navneet Dalal and Bill Triggs[6]. HOG is similar to edge-oriented histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is calculated on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization to improve accuracy. HOG works on cells that make geometric and photometric transformations invariant. This provides a benefit to the HOG descriptor over the other descriptors except for the rotation of objects. However, by enhancing the training dataset by including information from different orientations, the orientation issue can be overcome. Fig. 3 shows the process of extraction of the HOG function.

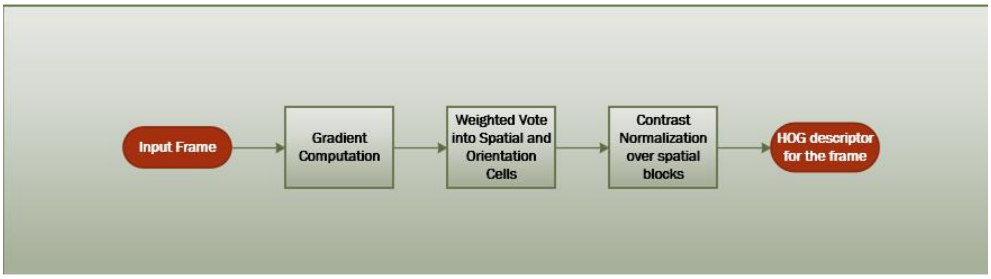
Input Frame
Gradient Computation
Weighted Vote into Spatial and Orientation Cells
Contrast Normalization over spatial blocks
HOG descriptor for the frame

Fig.3. HOG feature extraction procedure

The steps in calculating HOG feature representation for an image are as follows:

1. *Gradient computation of the input image*

The calculation's first stage is the calculation of gradient values. Applying the 1-D focused, point discrete derivative mask in one or both horizontal and vertical directions is the most prevalent technique. This method specifically requires filtering the image's color or intensity information using the following filter kernels: [−1,0,1] and [−1,0,1] T.

2. *Weighted Vote into spatial and orientation cells*

Creating cell histograms is the second step in the calculation. Each pixel in the cell casts a weighted vote based on the values discovered in the gradient computation for an orientation-based histogram channel. The cells can be either rectangular or radial in form and the histogram channels are distributed uniformly over 0 to 180 degrees or0 to 360 degrees depending on whether the gradient is "unsigned" or "signed." As for the weight of the ballot, the contribution of pixels can either be the magnitude of the gradient itself, or some magnitude function. We merely used the magnitude of the gradient.

It is therefore essential to ensure the proper quantity of data about the item is encoded by the HOG feature vector. By varying the parameter of the HOG cell size and visualizing the outcome, we can see the impact of the cell size parameter on the quantity of shape data encoded in the feature vector as shown in Fig. 4.

The visualization in Figure 4.4 demonstrates that a cell size of [1 1] does not encode much data about the form, whereas a cell size of [4 4] encodes a lot of data about the form but considerably improves the dimensionality of the HOG feature vector. A good compromise is the size of a cell 2-by-2. This size setting encodes enough spatial data to visually define a digit shape while restricting the number of measurements in the vector of the HOG function, which helps accelerate instruction.

3. *Block formation*

The gradient strengths must be locally normalized to account for modifications in illumination and contrast, which involves grouping the cells together into bigger, spatially linked blocks. The HOG descriptor is the concatenated vector from all the block areas of the parts of the normalized cell histograms. Each cell contributes to the final feature vector.

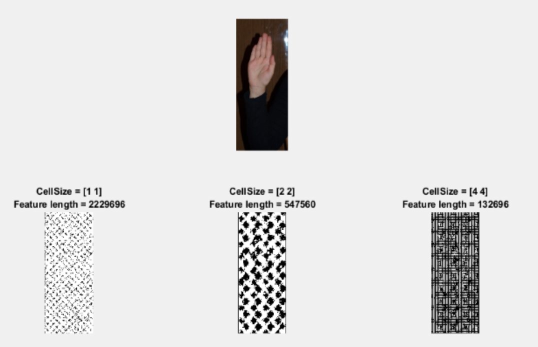

Fig.4. HOG feature visualization along with cell size for a training image in dataset 1

There are rectangular R-HOG blocks and circular C-HOG blocks in two primary block geometries. The C-HOG and R-HOG perform almost identically. In our case, R-HOG has been used.

4. *Contrast Normalization*

Different normalization techniques are there. However, the basic technique remains the same. Let the non-normalized vector comprising all the histograms in a specified block be k −norm for k = 1,2 and e a little constant. Then it is possible to calculate the normalization factor by the equation (2):

$$L1 - sqrt : f = \sqrt{\frac{v}{\|v\|_k + e}} \qquad (2)$$

Contrast Normalization makes it more amenable for the features extracted to be processed by the classifier. After, contrast normalization the feature vector is prepared to be fed into the classifier.

This is how descriptors of the HOG function are extracted from a particular picture. To train the Support Vector Machine (SVM), we extracted HOG features from all the images from the training dataset.

C. *Classifying Gestures using Support Vector Machine*

The Support Vector Machine (SVM) is developed using a training data function vector information. Then pictures from the test dataset will be placed in the forecast classifier. Two different datasets were used in this job to analyze our system model's efficiency. In the first situation, only two feasible classes need to be identified. So we need only binary classification using SVM in this situation. In the second scenario, multi-class SVM based on ECOC was used to classify various kinds of pictures.

The binary classification algorithm for SVM is searching for an ideal hyperplane separating the information into two groups. The optimal SVM hyperplane is the one with the greatest margin between the two classes. Margin implies the maximum slab width parallel to the hyperplane that does not have any internal information points. The support vectors are the nearest information points to the separating hyperplane; these points are on the slab's border.

Since most practical applications involve multi-class classification, scientists have suggested a number of techniques for generating multi-class SVMs from binary SVMs. One vs. one, one vs. remainder, Directed Acyclic Graph (DAG), and multi-class methods based on Error Corrected Output Coding (ECOC) generate many binary classifiers and combine their outcomes to determine a sample pixel's class label.

A classification problem of the K-class is converted into two-class N problems in ECOC. First, a matrix of K*N binary coding is necessary.

Table 1. An ECOC Matrix

Class	Code Word						
	f_1	f_2	f_3	f_4	f_5	f_6	f_7
Co	0	0	0	0	0	0	0
C1	1	1	1	1	0	0	0
C2	1	1	0	0	1	1	0
C3	1	0	1	0	1	0	1

By training N binary classifiers, ECOC combined with the SVM approach works to distinguish between different classes of K. According to a binary matrix M, each class is given a codeword length N. Each M row corresponds to a specific class. For K= 4 classes and N= 7-bit code words, Table 1 demonstrates an instance. Each class has a matrix row. To train a separate binary classifier, each column is used. The output codeword of the N classifiers is compared with the given K code words when testing an unseen example, and the one with the minimum hamming distance is considered the test data class label.

Binary SVM classifiers are used in this work to recognize gestures and non-gestures, and ECOC SVM classifiers are used to classify among multiple classes.

IV. Experimental Results and Discussion

In this section, we discuss the result and the efficiency of our system model in classifying hand gestures correctly. For hand gesture recognition, we have used the dataset created by M. Kawulok, J. Kawulok, and J. Nalepa [15,16,17]. We have simulated the efficiency of this system model for two different problems namely the recognition of hand gestures as a form of request and recognition of the number of fingers from the gestures.

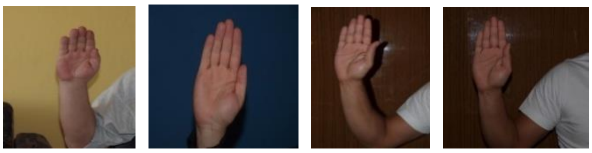

Fig.5. Sample Images from dataset 1

For the first problem, we have created a database containing two sets of the image. One is for training purposes and the other for testing purposes. Fig. 5 shows some random images from the data set 1.

Dataset 1 contains two classes of data namely images of gestures and images of non-gestures. The gesture in this dataset is defined as a signal of request using hands whether left or right. Non-gestures are images that have in them the position of palm or, wrist in a manner that doesn't necessarily mean request. For classification purposes, the dataset is divided into training and testing part. We used 80% of the data for training and the rest for testing.

In the second problem, we considered the recognition of a number of fingers using hand gestures. Fig. 6 illustrates the numbers as indicated by the fingers. We have created five classes of data for extracting features and creating the classifier. The five classes are namely gestures with one to five-pointed fingers. And, the dataset is split into training and testing set.

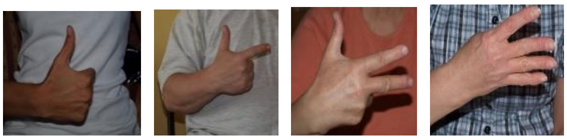

Fig.6. Sample Images from dataset 2

A. *Performance Evaluation on Data Set 1*

In this step, we assessed the performance of our model on data set 1. First, the images are pre-processed in the training data set. Then we used HOG to extract feature vector data from the training image data which was then fed into the SVM binary classifier which we have described in the previous chapter.

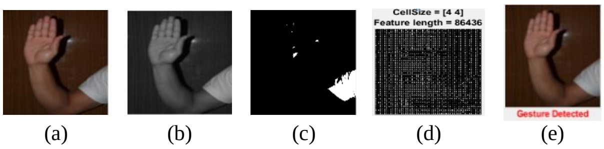

Fig.7. Gesture Recognition on a test image from dataset 1 (a) Original RGB image (b) Grayscale image (c) Binary image (d) HOG feature descriptor on the image (e) Gesture Detection

Table 2. Confusion Matrix for Data Set 1

	Gesture	Non-Gesture
Gesture	95%	5%
Non-Gesture	0%	100%

Fig. 7 demonstrates the step by step procedure in recognizing hand gestures using the system model. The confusion matrix in Table 2. shows the accuracy of the classifier when tested with the test dataset images.

Since the training and testing dataset is properly balanced, we are only showing the percentage rather than instances in the Confusion Matrix in Table 2. As shown in Table 2. the system accurately classifies an image to be a gesture in 95% of the cases when it is actually a gesture and with perfection, in case it is not a gesture. The accuracy here is 97.5%.

B. Performance Evaluation on Data Set 2

Then, we evaluated the efficiency of our model on data set 2 which identifies the number of pointed fingers from a given image. The dataset contains images showing numbers from one to five using a finger in different orientations. The steps of execution are the same as before. However, we used an ECOC based multiclass classifier in this case. Fig. 8 shows the step by step procedure of recognizing the number of fingers on a hand through hand gesture. The simulation results are shown in Fig. 9.

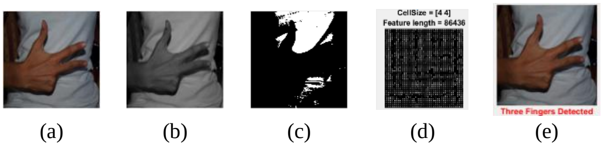

Fig.8. Finger Number Recognition on a test image from dataset 2 (a) Original RGB image (b) Grayscale image (c) Binary image (d) HOG feature descriptor on the image (e) Number of fingers detected.

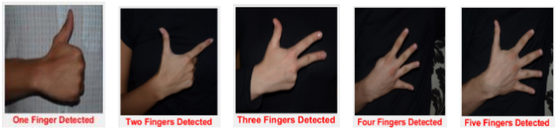

Fig.9. Detection of Fingers in Dataset 2

Table 3. Confusion Matrix for Data Set 2

	One Finger	Two Finger	Three Finger	Four Finger	Five Finger
One Finger	97%	0%	0%	0%	3%
Two Finger	0%	99%	1%	0%	0%
Three Finger	0%	4%	96%	0%	0%
Four Finger	0%	0%	1%	99%	0%
Five Finger	4%	0%	0%	0%	96%

The confusion matrix in Table 3. shows the efficiency of the multi-class classifier on test data from dataset 2. The overall accuracy of the system model using the multiclass classifier is 97.4%.

For, overall performance measurement we will evaluate four measures of efficiency for the model - accuracy, precision, recall, and F1 score. The equation (3),(4),(5),(6) shows how they are calculated.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - score = \frac{2 * (recall * precision)}{recall + precision} \tag{6}$$

Here, TP stands for True Positive, FP for False Positive, TN for True Negative and, FN for False Negative.

Accuracy shows the total correct classification rate. The precision shows the ratio between correct positive class prediction and, the total positive class prediction. The recall shows how many are correctly classified among all the positively classified cases. The F1 score is a measure of the weighted average of precision and recall. Table 4. shows the precision, recall, and F1 values.

Table 4. Accuracy, Precision, Recall, F1 Value

	Precision	Recall	F1 Value
Binary Classification on Dataset No. 1	**95%**	**100%**	**97.43%**
For Class 1 in Multiclass Classification	**97%**	**88.18%**	**92.37%**
For Class 2 in Multiclass Classification	**99%**	**96.11%**	**97.53%**
For Class 3 in Multiclass Classification	**96%**	**97.95%**	**96.96%**
For Class 4 in Multiclass Classification	**99%**	**100%**	**99.49%**
For Class 5 in Multiclass Classification	**96%**	**96.96%**	**96.47%**

From the confusion matrix, it is evident that the proposed method is highly efficient and can meet the real-time application requirements.

V. Conclusion

This paper outlines an empirical approach to hand gesture recognition. The system model presents a simple yet powerful and suitable method that can detect and recognize hand gestures by combining HOG based feature extraction method and SVM based classification. The work presented here provides a groundwork which can be utilized in real-life applications. The simplicity of the model provides support for lite-applications. We can further extend the work to include more nonverbal communication modes, eye-contacts, understanding the meaning of gesture signals in a cross-culture context, etc.

Acknowledgment

I would like to express my gratitude to the Ministry of ICT, Government of Bangladesh for providing me with a fellowship during my research period.

References

[1] Cho, Mi-Young, and Young-Sook Jeong. "Human gesture recognition performance evaluation for service robots." 2017 19th International Conference on Advanced Communication Technology (ICACT). IEEE, 2017.

[2] Rahman, Md Abdur, and M. Shamim Hossain. "A gesture-based smart home-oriented health monitoring service for people with physical impairments." International Conference on Smart Homes and Health Telematics. Springer, Cham, 2016.

[3] Yang, Geng, et al. "A Novel Gesture Recognition System for Intelligent Interaction with a Nursing-Care Assistant Robot." Applied Sciences 8.12 (2018): 2349.

[4] Azeem, A., et al. "Hexagonal scale invariant feature transform (H-SIFT) for facial feature extraction." Journal of applied research and technology 13.3 (2015): 402-408.

[5] Kashif, Muhammad, et al. "Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment." Computers in biology and medicine 68 (2016): 67-75.

[6] Dalal, N.; Triggs, B., "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol.1, no., pp.886,893 vol. 1, 25-25 June 2005

[7] Prasuhn, L.; Oyamada, Y.; Mochizuki, Y.; Ishikawa, H., "A HOG-based hand gesture recognition system on a mobile device," Image Processing (ICIP), 2014 IEEE International Conference on, vol., no., pp.3973,3977, 27-30 Oct. 2014

[8] Tavakoli, Mahmoud, et al. "Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier." Biomedical Signal Processing and Control 46 (2018): 121-130.

[9] Tsironi, Eleni, et al. "An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition." Neurocomputing 268 (2017): 76-86.

[10] Maharani, Devira Anggi, Hanif Fakhrurroja, and Carmadi Machbub. "Hand gesture recognition using k-means clustering and support vector machine." 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, 2018.

[11] Kumar, Pradeep, et al. "Coupled HMM-based multi-sensor data fusion for sign language recognition." Pattern Recognition Letters 86 (2017): 1-8.

[12] Raheja, J. L., et al. "Robust gesture recognition using Kinect: A comparison between DTW and HMM." Optik 126.11-12 (2015): 1098-1104.

[13] Beke, Aykut, Ahmet Arda Yuceler, and Tufan Kumbasar. "A rule based fuzzy gesture recognition system to interact with Sphero 2.0 using a smart phone." 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE, 2017.

[14] Hongyi Liu, Lihui Wang, “Gesture recognition for human-robot collaboration: A review”,International Journal of Industrial Ergonomics,Volume 68,2018,, Pages 355-367, ISSN 0169-8141

[15] Michal Kawulok, Jolanta Kawulok, and Jakub Nalepa. 2014, “Spatial-based skin detection using discriminative skin-presence features”, Pattern Recogn. Lett. 41, C (May 2014), 3-13.

[16] Michal Kawulok, Jolanta Kawulok, and Jakub Nalepa. 2014, “Self-adaptive algorithm for segmenting skin regions”, EURASIP Journal on Advances in Signal Processing2014.

[17] Michal Kawulok, Jolanta Kawulok, and Jakub Nalepa. 2014, “Wrist Localization in Color Images for Hand Gesture Recognition”, Volume 242 of the series Advances in Intelligent Systems and Computing pp 79-86

Authors’ Profiles

Ahmed Abdal Shafi Rasel was born on the 23rd of October, 1993 in Bangladesh. He has earned his B.Sc. in Information Technology from the Institute of Information Technology from Jahangirnagar University Bangladesh in 2014. He completed his M.Sc. from the same institute in 2016.

He is currently employed as a faculty at the Department of Computer Science and Engineering, Stamford University Bangladesh.

Mr. Rasel has got Fellowship from the Ministry of ICT, Bangladesh for his research work during his M.Sc. degree completion.

His primary field of research is Machine Learning and Image Processing.

Mohammad Abu Yousuf was born on 1st January 1978 in Bangladesh. He completed his B.Sc. from Shahajalal University of Science and Technology, Bangladesh in 1999. Then he earned his M.Sc. degree Kyung Hee University, South Korea specializing in Biomedical Engineering in 2009. Then he acquired his Ph.D. from Saitama University, Japan in 2013.

He is currently working as an Associate Professor at the Institute of Information Technology, Jahangirnagar University.

Dr. Yousuf is a member of the IEEE Computer Society. He has got more than thirty scientific publications in the field of Image Processing, Computer Vision and Pattern Recognition. He has received the Japanese Government Fellowship for the

Ph.D. program.

How to cite this paper: Ahmed Abdal Shafi Rasel, Mohammad Abu Yousuf, "An Efficient Framework for Hand Gesture Recognition based on Histogram of Oriented Gradients and Support Vector Machine", International Journal of Information Technology and Computer Science(IJITCS), Vol.11, No.12, pp.50-56, 2019. DOI: 10.5815/ijitcs.2019.12.05