Hand Gesture Recognition using Multiclass Support Vector Machine

Md. Hafizur Rahman
Lecturer, Department of EEE
International University of Business Agriculture and Technology (IUBAT)
Dhaka-1230, Bangladesh

Jinia Afrin
B.Sc. in Computer Science & Engineering
Khulna University (KU)
Khulna, Bangladesh

ABSTRACT
Vision-based recognition system has developed rapidly over the past few years. This paper presents hand gesture recognition system that can be used for interfacing between computer and human using hand gesture. In natural Human Computer Interactions (HCI), visual interpretation of gestures can be very useful. In this paper we propose a method for recognizing hand gestures using Support Vector Machine (SVM). We propose a system which can identify specific hand gestures and use them to convey information. In this system we select the feature vectors by Biorthogonal Wavelet Transform. These extracted features are used as input to the classifier. Multi Class SVM is used for classifying hand gestures into ten categories: A, B, C, D, G, H, I, L, V, Y. This system gives us good performance for recognizing the gestures. We can get up to 92% correct results on a particular gesture set.

Keywords
Gesture Recognition, Canny Edge Detection, Radon Transform, Biorthogonal Wavelet, Multiclass Support Vector Machine.

1. INTRODUCTION
With the development of technology and computing, computer is becoming more and more important in our day to day life. Various input and output devices have been designed and used over the years for the purpose of easing the communication between computers and humans. In fact, more and more unrestricted interaction between humans and computers are required in computer applications. The idea behind it is to shape computers in a way so that computers can understand human language and develop a user friendly human computer Interactions (HCI). Making a computer understand human gesture is a step towards it. A gesture is spatiotemporal pattern, which may be static or dynamic or both. In [1], Bobick and Wilson defined gestures as the motion of the body to communicate with other agents. In recent years, gesture recognition system has become very popular in the field of research, especially facial and hand gesture recognition system. Hand gestures can be classified in two categories: static and dynamic. Whereas static gesture is a particular hand configuration and pose, represented by a single image, a dynamic gesture is a moving gesture, represented by a sequence of images. Sign language is one of the most structured set of gestures. In sign language, each gesture has a specific meaning (or meanings). Hand gestures provide a more human-computer interface, allowing us to point, or rotate a 3D model by rotating hands. For instance, transformation of human hand motion for tele-manipulation is especially important in hazardous environment. One of the most important applications of hand gesture recognition is the easy way communication with the deaf or non-vocal persons through a hand-gesture to speech system to improve the quality of life. Deaf or non-vocal individuals are unable to communicate with others through speech due to congenital malfunction, disease, head injuries. Deaf or non-vocal persons use sign language or hand gestures to express themselves. However, most of the people who have hearing power do not have the special sign language expertise. This is a major barrier between these two groups in daily communication. To overcome this barrier and to help those people to integrate into society is a very challenging research area.

The first gestures that were applied to computer interactions date back to the PhD work of Ivan Sutherland [2], who demonstrated Sketchpad, an early form of stroke-based gestures using a light pen to manipulate graphical objects on a tablet display. This form of gesturing has received widespread acceptance in the human-computer interaction (HCI) community. In the last decade, several methods of potential applications [3-4] in the advanced gesture interfaces have been suggested but these differ from one to another in their models. Some of these models are Neural Network [3], Fuzzy Systems [4] and HMM [5-8]. In rapid object detection [9], Viola uses integral images as Haar wavelet features. Integral images allow for the fast implementation of box type convolution filters, which makes very fast feature extraction. A real-time hand gesture recognition system using skin color segmentation and multiple-feature based template matching techniques was introduced by Hasanuzzaman et al. in [10]. This method shows that, the three largest skin-like regions are segmented from the input images by skin color segmentation technique from YIQ color space then they are compared for feature based template matching through the use of a combination of two features correlation coefficient and minimum (Manhattan distance) distance qualifier. Ho-Sub et al. [11] has used the combined features of location, angle and velocity to determine the discrete vector which is used as input to HMMs. Wu [12] developed a hand gesture recognition system for media player control. This system firstly separates the left arm by background subtraction and detects the straight line by both Hough transform and Radon transform.

2. SYSTEM OVERVIEW
We propose an automatic system that recognizes static hand gesture for alphabets (A,B,C,D,G,H,I,L,V,Y) using Biorthogonal Wavelet Transform. In particular, the proposed system consists of several steps. In the first step images are at first read and then pre-processed. Then noise is removed from the image by using filters. In the next step the proposed method detect the edge from the images and then compute *projections* of an image along specified directions by RT (Radon Transformation). After that, the Biorthogonal Wavelet Transformation is performed on the projections of an image which we get from the RT (Radon Transformation). Then the

gestures are trained by Multiclass SVM and then testing is performed. After the completion of testing period the gestures are recognized. The overall system is described in Fig. 1.

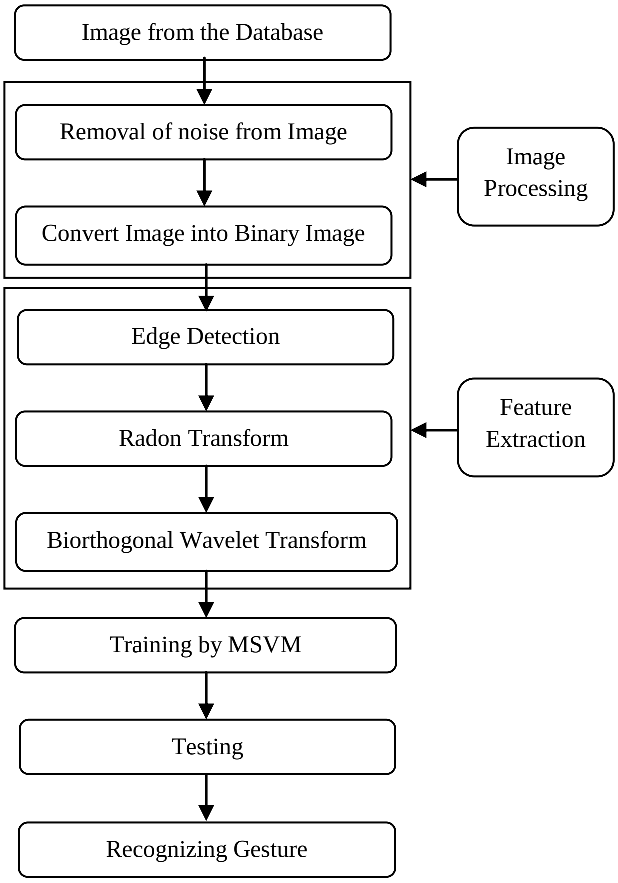

Fig 1: Block diagram of Hand Gesture Recognition System.

3. OVERALL METHODOLOGY

3.1 Canny Edge Detection Algorithm

In this system we detect the edges of an image by Canny Edge Detection Algorithm [14], because this edge detector performs better than reference algorithms. In this algorithm optimal edge is detected based on some criteria which include finding edges by minimizing error rate, making edges closely to the actual edges to maximize location and marking edges only once when a single edge exists for minimal response. The optimal filter that meets all three criteria above can be approximated using the first derivative of a Gaussian function.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

$$\frac{\partial G(x,y)}{\partial x} \;\alpha\; xe^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

$$\frac{\partial G(x,y)}{\partial y} \;\alpha\; ye^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

In the first stage the image is convoluted with a Gaussian filter. Then the gradient of the image is measured by feeding the convoluted image. The 2-D convolution operation is described in the following equation.

$$I'(x, y) = g(k, l) \times I(x, y) \quad (4)$$

$$I'(x,y) = g(k,l) \times I(x,y)$$

$$I'(x,y) = \sum_{K=-N}^{N} \sum_{l=-N}^{N} g(k,l) I(x-k, y-l) \quad (5)$$

Where: $g(k,l)$ = convolution kernel

$I(x,y)$ = original image

$I'(x,y)$ = filtered image

$2N + 1$ = size of convolution kernel

In Fig. 2 the Canny Edge Detection process is demonstrated. Firstly, noise is removed from the input image, then image is converted to binary image, then Canny Edge Detection algorithm is applied to detect the boundary of the hand.

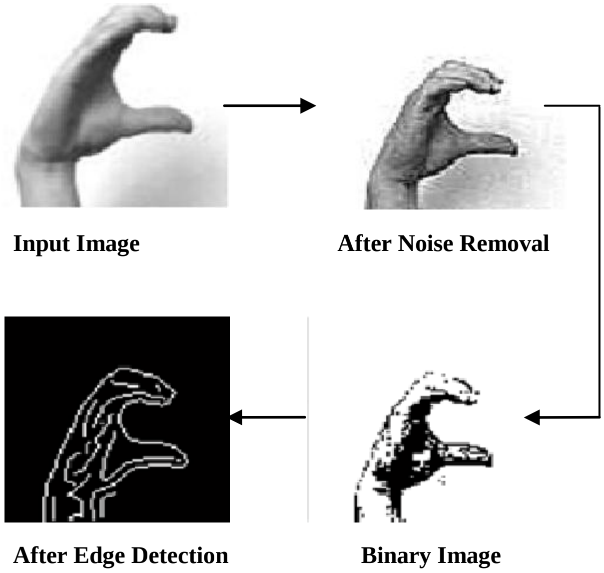

Fig 2: Canny Edge Detection process from input image.

3.2 Radon Transformation

By applying the Radon transform [13] on an image $f(x,y)$ for a given set of angles, the resulting projection is the sum of the intensities of the pixels in each direction and the resultant image is $R(r,q)$. This can be written mathematically by defining

$$r = x \cos q + y \sin q \quad (6)$$

After which the Radon Transformation can be written as

$$R(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(\rho - x\cos\theta - y\sin\theta)dxdy \quad (7)$$

Here δ is dirac delta function.

A projection of a two-dimensional function $f(x,y)$ is a set of line integrals. The *radon* function computes the line integrals from multiple sources along parallel paths, or *beams*, in a certain direction. The beams are spaced 1 pixel unit apart. To represent an image, the *radon* function takes multiple, parallel-beam projections of the image from different angles by rotating the source around the center of the image.

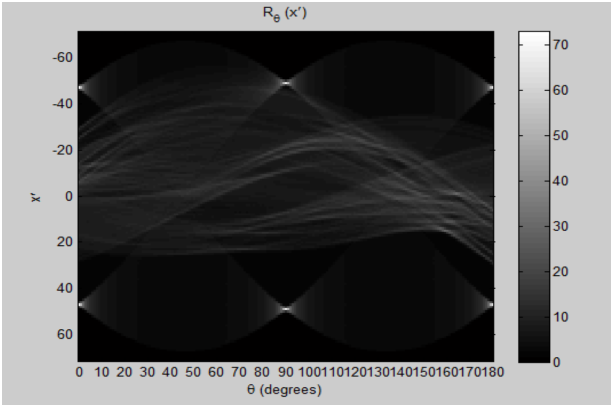

Fig 3: Plot of Radon Transformation of an image at angles from 0° to 180°.

After detecting the edges of the image, computation of the multiple parallel-beam projections of the image from different angles by rotating the source from 0° to 180° was performed around the center of the image. In Fig. 3 we can see that some sine waves appear, it is because the Radon transform of a Dirac delta function (**δ**) is a distribution supported on the graph of a sine wave. Consequently the Radon transform of a number of small objects appears graphically as a number of blurred sine waves with different amplitudes and phases. For that reason the Radon transform data is often called a *sinogram.*

3.3 Biorthogonal Wavelet

In biorthogonal wavelet, the associated wavelet transform is invertible but not necessarily orthogonal which can be defined similarly to orthogonal wavelets, except that the starting point is biorthogonal multi resolution approximations. The following decompositions are performed according to [17]:

$$V_{j-1} = V_j \oplus W_j \quad with \quad W_j \subset \left(V_j^*\right)^{\perp} \tag{8}$$

$$V_{j-1}{}^* = V_j{}^* \oplus W_j{}^* \quad with \quad W_j{}^* \subset \left(V_j\right)^{\perp} \tag{9}$$

Like in the orthogonal case $f(t)$ can be written as

$$f(t) = \sum_{j,n\in\mathbb{Z}} \langle f, \varphi_{j,n}^* \rangle\, \varphi_{j,n}(t) \tag{10}$$

$$= \sum_{n\in\mathbb{Z}} \langle j, \emptyset_{j,n}^* \rangle\, \emptyset_{j,n}(t) + \sum_{k\leq j,n\in\mathbb{Z}} \langle f, \varphi_{k,n}^* \rangle\, \varphi_{k,n}(t) \tag{11}$$

$$= \sum_{j,n\in\mathbb{Z}} \langle f, \varphi_{j,n} \rangle\, \varphi_{j,n}{}^*(t)$$
$$= \sum_{n\in\mathbb{Z}} \langle j, \emptyset_{j,n} \rangle\, \emptyset_{j,n}{}^*(t) + \sum_{k\leq j,n\in\mathbb{Z}} \langle f, \varphi_{k,n} \rangle\, \varphi_{k,n}{}^*(t) \tag{12}$$

By performing biorthogonal wavelet transformation on the computed projection which we obtained from the Radon Transformation, the final feature for our recognition system is selected. We used biorthogonal wavelet 3.7 in this system.

3.4 Multi Class SVM

The SVM [15] is a very well known learning algorithm for classification problem. The main idea of SVM is that; it finds the optimal separating hyper plane such that error for unseen patterns is minimized.

Consider the problem of separating the set of training vectors belonging to two separate classes.

$$\{x_1, x_2 \ldots . x_n\} \tag{13}$$

Which are vectors in $\mathbb{R}^D$.

We consider a decision function of the following form:

$$y(x) = w^T \varphi(x) + b \tag{14}$$

Attached to each observation x_i is a class label, $t_i \in \{-1, +1\}$.

Without loss of generality, we must construct a decision function such that, $y(x_i) > 0$ for all i such that $t_i = +1$, and $y(x_i) < 0$ for all i such that $t_i = -1$. We can combine these requirements by stating,

$$t_i y(x_i) > 0 \quad \forall i \tag{15}$$

The idea is to extend it to multi-class problem is to decompose an M-class problem into a series of two-class problems [16].

Let the j^{th} decision function, with the maximum margin that separates class i from the remaining classes, be

$$y_j(x) = w_j^T \emptyset(x) + b_i \tag{16}$$

Here, w_j is n dimensional vector, $\emptyset(x)$is mapping function which maps x into n dimensional space.

The problem can be equivalently understood in terms of projecting the input data into a higher dimensional space where they are separated using parallel hyper planes. Now, if the classification problem is separable, the training data belonging to class k satisfy $y_k(x) = 0$ and data belonging to other classes must satisfy $y_k(x) \leq 1$. In case of inseparable problem unbounded support vectors satisfy the condition $|y_k(x) = 1|$ and bounded support vectors belonging to class k satisfy the condition $y_k(x) \leq 1$ and other data belonging to other classes satisfy $y_k(x) \geq -1$.

4. EXPERIMENTAL RESULTS

The proposed method was tested on five different users showing ten gestures such as A,B,C,D,G,H,I,L,V,Y shown in Fig. 4. Table 1 shows the recognition results of 10 gestures of 5 users. Fig. 5 shows accuracy vs user curve for five different users.

The result shows that the proposed algorithm successfully detects hand gesture with higher accuracy. In order to achieve robustness of the method to varying conditions of operation such as illumination, posture, zooming conditions and skin color, a large dataset of hand gesture is used to train the classifier. The training set of our detection phase comprised of over 800 positive samples and 1500 negative image samples. The algorithm has been implemented in MATLAB.

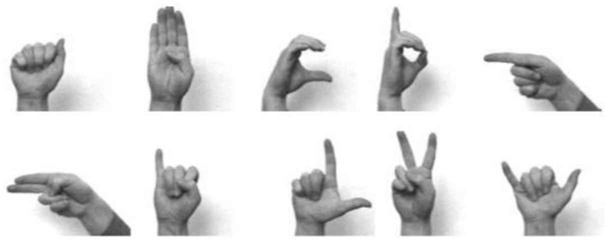

Fig 4: Finger spelled Alphabet (Top row [A, B, C, D, G], Bottom row [H, I, L, V, Y]).

Table 1. Recognition results of 10 gestures of 5 users

User \ Gesture	A	B	C	D	G	H	I	L	V	Y
1	95%	96%	88%	85%	94%	82%	88%	92%	88%	75%
2	92%	90%	85%	88%	92%	90%	85%	90%	82%	82%
3	88%	92%	82%	90%	87%	90%	92%	90%	93%	87%
4	86%	94%	84%	92%	90%	88%	82%	91%	90%	82%
5	75%	88%	88%	93%	90%	82%	80%	88%	84%	80%
Total	87%	92%	85%	90%	91%	86%	85%	90%	87%	81%

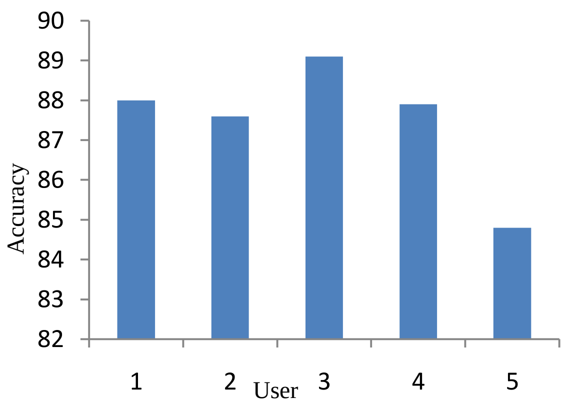

Fig 5: Comparison result of five users.

5. CONCLUSION

An efficient hand gesture recognition system requires higher class robustness, accuracy and efficiency. In this paper we propose a method for classifying static hand gestures using Multiclass SVM. There exist many classification algorithms, for example neural networks and classification trees. However Support Vector Machine approach is considered to be a very good candidate for classification problem like hand gesture recognition. This is due to its high generalization performance without the need to add a-prior knowledge, even when the dimension of the input space is very high. Features are selected by Radon transform and Biorthogonal Wavelet. Proposed system works well on wide range of variation in color, position, scale and orientation with image. The experimental results reveal that the proposed method is better than any other method to detect hand gestures with high classification accuracy.

6. REFERENCES

[1] Andrew Wilson and Aaron Bobick, “Learning visual behavior for gesture analysis,” In Proceedings of the IEEE Symposium on Computer Vision, Coral Gables, Florida, pp. 19-21, November 1995.

[2] I. E. Sketchpad: “Aman-machine graphical communication system”. In: Proceedings of the AFIPS Spring Joint Computer Conference 23. pp. 329–346, 1963.

[3] X. Deyou, “A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG”, International Conference ICPR, pp. 519-522, 2006.

[4] E. Holden, R. Owens and G. Roy, “Hand Movement Classification Using Adaptive Fuzzy Expert System”, Journal of Expert Systems, Vol. 9(4), pp. 465-480, 1996.

[5] M. Elmezain, A. Al-Hamadi and B. Michaelis, “Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences”, Journal of WSCG, Vol. 16, pp. 65-72, 2008.

[6] M. Elmezain, A. Al-Hamadi and B. Michaelis, “A Novel System for Automatic Hand Gesture Spotting and Recognition in Stereo Color Image Sequences”. Journal of WSCG, Vol.17, No. 1, pp. 89-96, 2009.

[7] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, “A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory”. International Conference on Pattern Recognition (ICPR), pp. 519-522, 2008.

[8] M. Elmezain, A. Al-Hamadi and B. Michaelis, “Spatio-Temporal Feature Extraction-Based Hand Gesture Recognition for Isolated American Sign Language and Arabic Numbers”. IEEE Symposium on ISPA, pp. 254-259, 2009.

[9] P. Viola and M Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of Computer Vision and Pattern Recognition. Hawaii,U.S., 2001, pp. 511–518.

[10] Md. Hasanuzzaman, V. Ampornaramveth, Tao Zhang, M.A. Bhuiyan ,Y. Shirai and H. Ueno, "Real-time Vision-based Gesture Recognition for Human Robot Interaction". In the Proceedings of the IEEE International Conference on Robotics and Biomimetics, Shenyang China, 2004.

[11] Y. Ho-Sub, S. Jung, J. B. Young and S. Y. Hyun, "Hand Gesture Recognition using Combined Features of Location,Angle and Velocity", Journal of Pattern Recognition, Vol. 34(7), pp. 1491-1501, 2001.

[12] Y. M. Wu, "The implementation of gesture recognition for media player system," Master Thesis of the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, 2009.

[13] Gregory Beylkin, "Discrete Radon Transform", Ieee Transactions On Acoustics, Speech, And Signal Processing, vol. assp-35, no. 2,1987, pp. 162-172

[14] Canny, J., "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986.

[15] Christopher J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition". Kluwer Academic Publishers, Boston, pp. 1-43

[16] Shigeo Abe, "Support Vector Machines for Pattern Classification, second edition", Kobe University,Graduate School of Engineering, 2nd edition, Springer-Verlag London Limited 2005, 2010.

[17] Stéphane G. Mallat, "A Wavelet Tour of Signal Processing". Academic Press, 1999, ISBN 978-0-12-466606-1