

Recognition of Static Hand Gestures of Alphabet in Bangla Sign Language

Md. Atiqur Rahman¹, Dr. Ahsan-Ul-Ambia², Md. Ibrahim Abdullah³ & Sujit Kumar Mondal⁴

¹Department of Computer Science & Engineering, University of Information Technology & Sciences, Dhaka, Bangladesh.

^{2,3,4}Department of Computer Science & Engineering, Islamic University, Kushtia, Bangladesh.

Abstract: This paper presents a system for recognizing static hand gestures of alphabet in Bangla Sign Language (BSL). A BSL finger spelling and an alphabet gesture recognition system was designed with Artificial Neural Network (ANN) and constructed in order to translate the BSL alphabet into the corresponding printed Bangla letters. The proposed ANN is trained with features of sign alphabet using feed-forward back-propagation learning algorithm. Logarithmic sigmoid (logsig) function is chosen as transfer function. This ANN model demonstrated a good recognition performance with the mean square error values in this training function. This recognition system does not use any gloves or visual marking systems. This system only requires the images of the bare hand for the recognition. The Simulation results show that this system is able to recognize 36 selected letters of BSL alphabet with an average accuracy of 80.902%.

Keywords: Artificial Neural Network, Back-propagation Neural Network, Bangla Sign Language, Gestures Recognition, Sign language Recognition.

I. Introduction

A sign language means using gestures instead of sound or spoken words to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. This sign language is normally considered as the first language and main means of communication for deaf individuals. The signers, however, still face some difficult problems to communicate with the people who are not well acquainted with the signs. The communication complexity severely influences on the life and interpersonal relationships in the deaf community. Deaf individuals communicate with speaking people usually via interpreters or text writing. Although interpreters can help the communication between deaf and hearing persons, they are often expensive and have negative effect on independency and privacy.

There is no universal sign language. Different countries use different sign languages. To help finding a communication aid for deaf people, many researchers have been working on recognition of various sign languages, e.g. Australian, Japanese, Chinese, German, and American sign languages, etc. Research in sign language recognition started to appear in literature at the beginning of 1990s. Stephan Liwicki and Mark Everingham [1] develop a method in which recognition based on hand shape alone is used where motion cues is not required. It also includes scalability to large lexicon recognition with no re-training. But there exists similar discrimination problem like previous work. The letter 'V' and 'I' has very similar representation. Moreover, the dataset contains only a single inexperienced signer and that the imaging conditions are only moderately challenging. Takahashi and Kishino [2] used a range classifier to recognize 46 Japanese Kana manual alphabet with a VPL Data Glove™. The hand gestures were simply encoded with data ranges for joint angles and hand orientations based on experiments. This system could recognize 30 out of 46 hand gestures correctly, but the remaining 16 signs could not be reliably recognized.

Artificial neural networks have been widely used in sign language recognition research. Murakami and Taguchi [3] investigated the use of recurrent neural nets for Japanese Sign Language recognition. Although it achieved a high accuracy of 96%, their system was limited only to 10 distinct signs. Kramer and Leifer [4, 5] developed an ASL finger-spelling system using a Cyberglove, with the use of neural networks for data segmentation, feature classifier, and sign recognition. Using a tree-structured neural classifying vector quantizer, a large neural network with 51 nodes was developed for the recognition of ASL alphabets. They claimed a recognition accuracy of 98.9% for the system.

Begum, S. and Hasanuzzaman, M.[6] have developed a computer vision-based Bangladeshi sign language recognition system. In this system, separate PCA (principal component analysis) is used for Bengali Vowels and Bengali numbers recognition. The system is tested for 6 Bengali vowels and 10 Bengali numbers. In recognition of the linguistic of the Deaf community, the government of Bangladesh has pronounced to promote Bangla Sign language as of 1 February, 2009. Such an initiative by the government will allow the 2.4 million Individuals and Sign Language users to excel far. This work presents a BSL recognizer that has been

developed to recognize 36 letters of alphabet of BSL using Back-propagation Neural Network technique. In this recognition system, we have used a new set of features for training and testing of ANN. This system allows fast training and intelligent learning of new gestures. It can recognize 36 letters with average accuracy of 80.902%. The paper is organized as follows. Section 2 illustrates alphabetic finger signs. Section 3 briefly explains Neural Network Training Process. Proposed recognition method is described in section 4. Section 5 presents recognition results. Finally Conclusion is given in section 6.

II. Sign Alphabets

The Sign Language Alphabet is a set of alphabetic finger signs. The Bangla Sign Language alphabet is shown in below:



Figure 1: BSL Alphabet

III. Artificial Neural Network

Artificial Neural Network (ANN) or commonly neural network (NN) is an interconnected group of artificial neurons that use a mathematical or computational model for information processing based on a connectionist approach to computation [7,8]. According to Principle et al. [9], one of the most significant strength of ANN is its ability to learn from a limited set of examples. ANN has been successfully used in solving complicated problems in different domain such as pattern recognition, identification, classification, speech, vision, and control systems [10]. An ANN, which imitates the human brain in problem solving, is capable in modeling the complex relationship between input and output to find patterns in data. Typically, an ANN consists of a set of interconnected processing elements or nodes called perceptrons. The nodes are organized in different ways to form a network structure where each ANN is composed of a collection of perceptron grouped in layers. Each perceptron is designed to mimic its biological counterpart, the neuron and to accept a weighted set of input and respond with an output [11]. A sophisticated ANN may have several hidden layers, feedback loops and time delay elements, which designed to make the networks as effective as possible in discriminating relevant features or patterns [11]. The most well known ANN is feed-forward ANN. A feed-forward ANN consists of a set of nonlinear neurons connected together, in which the information flows in the forward direction [7]. Among the feed-forward network, multilayer Perceptron (MLP) is the most widely and commonly used model free estimators. A MLP consists of at least three layers, which are the input, hidden and output layer. The input and output layer contains a collection of neurons representing input and output variables. There are three learning types of ANN models, which are supervised, unsupervised and reinforcement learning. The network modifies the weight based on a sequence of training vector with an associated target output node, known as supervised training. On the other hand, unsupervised training refers to a network that modifies weight by assigning the most similar input vectors to an output unit. And the third learning type is reinforcement training, which lies between supervised and unsupervised learning. Among the various neural network models, back propagation is the best general purpose model and is probably the best at generalization [12, 13]. The back propagation is the classical algorithm used for learning. It is an iterative gradient descent algorithm which is designed to minimize the mean squared error between the desired output and the generated output for each input pattern [14]. In this research, focus is given to the feed-forward and back- propagation model with multi layer perceptron.

IV. Recognition Method

We have used MATLAB tools to implement the algorithm [15]. The block diagram of our recognition system is given in the following figure 2. In this recognition system, we have used multilayer feed forward neural network. For learning the network, Back-propagation algorithm has been used. The activation function which has been used is “Sigmoid” [16].

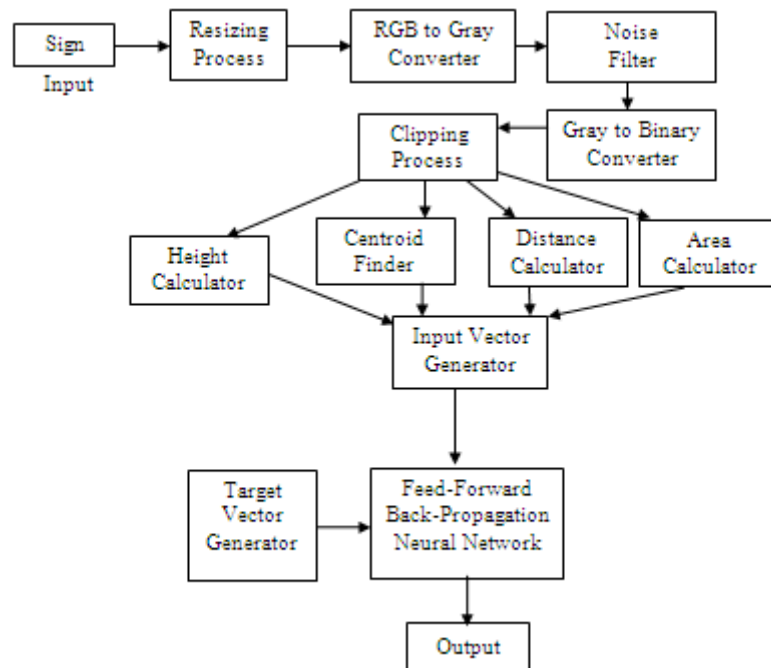


Figure 2: Block diagram of recognition system

4.1 Average Height of a Sign

To find the height of a sign from binary image where 0's represent black pixels and 1's represent white pixels, we have used pixel scanning method [17]. Starting from the left most column of top row, we scan each row of each column. When a 0 (zero) is found, the corresponding row is taken as the first row (Y_{first}). Continuing in this way, we find the last row (Y_{last}) of 0 (zero) of the same column. The height of each column of image is then calculated as

$$Height_{column(i)} = Y_{last}(i) - Y_{first}(i) + 1 \quad (1)$$

i , is the column number that contains at least one black pixel (0).

The average height of the binary image is then calculated as

$$Height_{avg} = N/M \quad (2)$$

Where, N is the total number of black pixels (0's) of the image and M is the total number of columns containing at least one black pixel (0).

4.2 Area of the Sign

The area of a sign means the total number of black pixels in the image. The area is calculated using the following algorithm:

Algorithm for Area Calculation

```

no_of_black_pixels=0;
for row=top-most-row to bottom-most-row
for column=left-most-column to right-most-column
    if image[row,column]=black_pixel
        no_of_black_pixels=no_of_black_pixels+1;
    end if
end for loop
end for loop
area= no_of_black_pixels.
```

4.3 Centroid of a Sign

The centroid of a polyhedron is simply the average of the respective coordinates of all the vertices of the polyhedron [18]. For example, if the coordinates are (x_1, y_1) (x_2, y_2) ... (x_N, y_N) , then the centroid would be,

$$X_c = (x_1 + x_2 + \dots + x_N) / N \quad (3)$$

$$Y_c = (y_1 + y_2 + \dots + y_N) / N \quad (4)$$

In our system, the centroid of an image is calculated using the following relations:

$$X_c = \frac{\sum_{i=1}^n x_i}{area} \quad (5)$$

Where, X_i represents the X-coordinate of each boundary pixel of the image.

$$Y_c = \frac{\sum_{i=1}^n y_i}{area} \quad (6)$$

Where, Y_i represents the Y-coordinate of each boundary pixel of the image and N is the total number of boundary points. The centroid of the image is (X_c, Y_c) .

4.4 Euclidian Distance of the Centroid from the Origin

Given the two points (x_1, y_1) and (x_2, y_2) , the Euclidian distance between these points is given by the formula [19]:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (7)$$

The coordinates of the left-top most pixel of binary image is considered as the origin. Hence the Euclidian distance between centroid and the origin is given by

$$d = \sqrt{x_c^2 + y_c^2} \quad (8)$$

Finally, the features collected from the above sections are combined to form a feature vector in the following order:

Feature vector, $V = [area, average\ height, x\text{-centroid}, y\text{-centroid}, centroid\text{-distance}]$.

V. Results Evaluation And Discussion

In the present work, the signs of BSL are captured by a digital camera of 8 mega pixel resolution. The images are saved as .jpg format. There are 23 samples of each sign of BSL used in the training and recognition. Therefore, a total of 36x23=828 signs are captured. Among 828 images 540 samples are used for training the ANN and the remaining 288 images are used for testing or recognition. The goal of the average root mean square error is specified for training the network and is less than or equal to 0.01. The BSL alphabet recognition rate is presented in table 1. Recognition rate is calculated as

$$\text{Recognition Rate} = \left(\frac{\text{No. of recognized letters}}{\text{No. of total samples of that letters}} \right) \times 100\%$$

For example recognition rate of 'ব' = (8/8)*100=100%.

In the present study, the highest recognition rate is obtained 'এ', 'ক', 'চ', 'ট' and 'ব' for and the lowest recognition rate is found for 'ই', 'দ', 'জ', 'ফ' and 'ং'. Figure 3 illustrates 3 (three) recognized 'ঊ', 'ব' and 'ঘ'.



Figure 3: Some samples of recognized signs of BSL

Table 1: The performance of Recognition of BSL alphabet.

Letters	Recognized Letters	Unrecognized Letters	Wrong Recognized Letters	Recognition Rate (%)
অ(AW)	7	1	0	87.5
আ(A)	7	0	1	87.5
ই(Hrashho E)	5	1	2	62.5
ঊ(Hrashho O)	6	1	1	75
এ(E)	8	0	0	100
ও(O)	7	1	0	87.5
ক(KAW)	8	0	0	100
খ(Khaw)	6	2	0	75
গ(GAW)	7	0	1	87.5
ঘ(GHAW)	5	2	1	62.5
চ(CHAW)	8	0	0	100
ছ(CHCHAW)	6	1	1	75
জ(Borgio JAW)	5	2	1	62.5
ঝ(JHAW)	6	2	0	75
ট(TAW)	8	0	0	100
ঠ(THAW)	7	1	0	87.5

ড(DAW)	6	0	2	75
ঢ(DHAW)	6	2	0	75
ত(TAW)	7	1	0	87.5
থ(THAW)	6	2	0	75
দ(DAW)	7	0	1	87.5
ধ(DHAW)	6	2	0	75
ন(Danto NAW)	7	0	1	87.5
প(PAW)	7	1	0	87.5
ফ(PHAW)	5	3	0	62.5
ব(BAW)	8	0	0	100
ভ(BHAW)	7	0	1	87.5
ম(MAW)	6	2	0	75
য(Antostho AW)	6	1	1	75
র(RAW)	7	1	0	75
ল(LAW)	7	1	0	87.5
শ(Donto SHAW)	6	2	0	75
হ(HAW)	7	0	1	87.5
ড(Dee Shoono RAW)	6	1	1	75
ঝ(Onooshshawr)	5	0	3	62.5
ঞ(Bishargo)	6	2	0	75

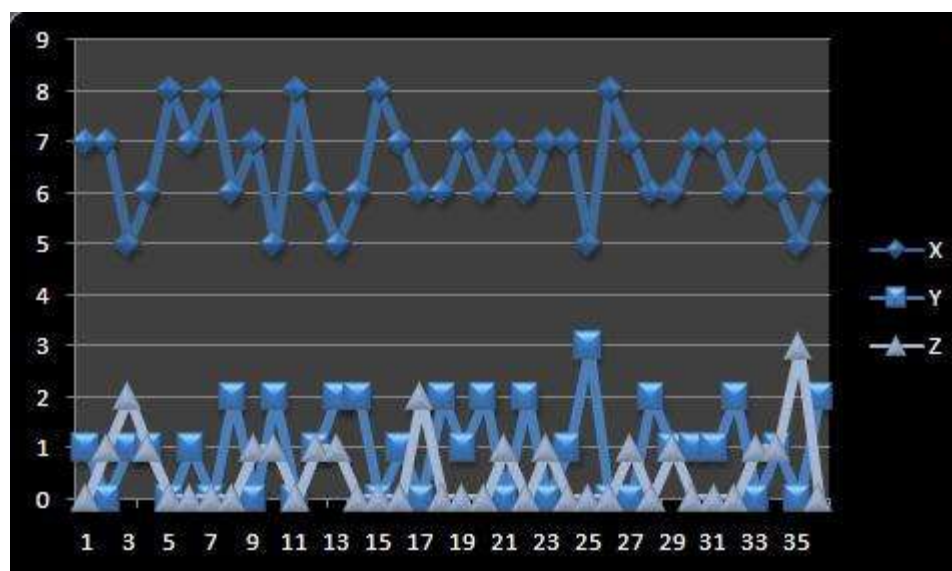


Figure 4: Comparisons among recognized, Unrecognized and wrong recognized letters (Here X: Recognized Letters, Y: Unrecognized Letters and Z: Wrong Recognized Letters).

VI. Conclusion

In this paper we present a simple model of static gestures of BSL alphabet recognition system using ANN with MATLAB tools. The system can perform static gesture training and recognition of BSL alphabet. The evaluation results show that the proposed method allows fast training and learning of new gestures and reliable recognition of the trained gestures afterwards. The average recognition accuracy of the system is 80.902%. Using more samples for training ANN may improve the performance of the system. Instead of ANN, other statistical recognition methods like Hidden Markov Model with more features may be used in future to improve recognition accuracy. The only limitation of our system is that, for learning NN, the feature vector should have integer values only. Future work will include extending the developed method to recognition of BSL with video based and interactive system.

References

- [1] Stephan Liwicki, Mark Everingham, "Automatic Recognition of Finger spelled Words in British Sign Language," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 50-57, Miami, FL, USA, June, 2009.
- [2] T. Takahashi, F. Kishino, "Gesture coding based in experiments with a hand gesture interface device," ACM SIGCHI Bulletin, Vol. 23, pp. 67-73, 1991.
- [3] K. Murakami, H. Taguchi, "Gesture recognition using recurrent neural networks," in Proceedings of the Conference on Human Factors and Computing Systems, pp. 237-242, 1991.
- [4] J. Kramer, L. J. Leifer, "A 'talking glove' for nonverbal deaf individuals," Technical Report No. CDR 19900312, Center for Design Research, Stanford University, U.S.A., 1990.
- [5] J. Kramer, "The TalkingGlove (RTM): hand-gesture-to-speech using an instrumented glove and a tree-structured neural classifying vector quantizer," Ph.D. Thesis, Department of Mechanical Engineering, Stanford University, U.S.A., 1996.
- [6] S. Begum, M. Hasanuzzaman, "Computer Vision-based Bangladeshi Sign Language Recognition System" The 12th International Conference of Computers and Information Technology (ICIT) '09, Bangladesh.
- [7] H. Zabiri, N. Mazuki, "A Black-Box Approach in Modeling Valve Stiction," International Journal of Mathematical, Physical and Engineering Sciences 4:1 2010, 2010.
- [8] Principe, J., "Neural Networks and Adaptive Systems", John Wiley and Sons, NY, 1999.
- [9] A. Sozen, E. Arcaklioglu, "Solar potential in Turkey," Applied Energy, 80(1), pp. 35-45, 2004.
- [10] V. O. Oladokun, A. T. Adebajo, O. E. Charles-Owaba, "Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course," The Pacific Journal of Science and Technology, vol. 9, no. 1, 2008.
- [11] J. Lawrence, "Introduction to Neural Network: Design, Theory and Application", 6th edition Nevada City, CA: California Scientific Software, 1994.
- [12] T. Mitchell, "Machine Learning", 1st edition New York: McGrawHill Science, 1997.
- [13] D.E. Rumelhart, J. L. McClelland, "Parallel distributed processing: Explorations in the microstructure of cognition", Cambridge, Mass: MIT Press, 1986.
- [14] P. S. Georgilakis, N.D. Hatzigryriou, A. D. Doulamis, N. D. Doulamis, S.D. Kollias, "A Neural Network Framework for Predicting Transformer Core Losses," Proceedings of the 21st 1999 IEEE International Conference on Power Industry Computer Applications, PICA -99, pp.301-308, 1999.
- [15] www.mathworks.com/access/helpdesk/help/techdoc/ref/. Accessed 23/02/2012
- [16] Md. Aktaruzzaman, Md. Farukuzzaman Khan and M. Ekin Uddin "Recognition of Offline Cursive Bengali Handwritten Numerals using ANN," Journal of the Peoples University of Bangladesh, vol. 4, N0.1, pp18-28, Bangladesh, July 2009.
- [17] Bekir Karlik and A. Vehbi Olgac "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks" International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (1): Issue (4), (2011).
- [18] <http://answers.yahoo.com/question/index?qid=1006041717112>. Accessed 23/01/2012
- [19] en.wikipedia.org/wiki/Distance. Accessed 23/12/2011

Authors

Md. Atiqur Rahman has been serving as a lecturer with the Department of Computer Science and Engineering (CSE), University of Information Technology and Sciences (UITS), Dhaka. He obtained his B.Sc. and M.Sc. degree in Computer Science and Engineering from Islamic University, Kushtia. His research interest includes Bioinformatics, Image Processing, pattern Recognition, Wireless Communication.

Dr. Ahsan-Ul-Ambia has been serving as an Associate Professor with the Department of Computer Science and Engineering (CSE), Islamic University, Kushtia. He joined at Islamic University in October 1999 after completion his M.Sc. Degree from the Department of Applied Physics and Electronics, University of Rajshahi; Rajshahi. His research areas include Biomedical Image and Signal Processing. Dr. Ambia has several papers related to these areas published in national and international journals as well as in referred conference proceedings.

Md. Ibrahim Abdullah has been serving as an Associate Professor with the Department of Computer Science and Engineering (CSE), Islamic University, Kushtia. He joined at Islamic University after completion his M.Sc. Degree from the Department of Applied Physics and Electronics, University of Rajshahi, Rajshahi.. Right now he teaches courses on Wireless Communication, Graphics, and Data Communication etc. Md. Ibrahim Abdullah has several papers related to these areas published in national and international journals as well as in referred conference proceedings.

Sujit Kumar Mondal has been serving as an Associate Professor with the Department of Computer Science and Engineering (CSE), Islamic University, Kushtia. He joined at Islamic University after completion his M.Sc. Degree from the Department of Applied Physics and Electronics, University of Rajshahi, Rajshahi.. Right now he teaches courses on Wireless Communication, Graphics, and Data Communication etc. Mr. Mondal has several papers related to these areas published in national and international journals as well as in referred conference proceedings.