Fuzzy Rule-Based Hand Gesture Recognition for Bengali Characters

Tanzila Ferdous Ayshee, Sadia Afrin, Quazi Ridwan Hasib, Md. Hossain, Rashedur M Rahman
Department of Electrical Engineering and Computer Science
North South University, Plot-15
Block-B, Bashundhara
Dhaka, Bangladesh
{bellaruby.ayshee, afrinsadia.nsu}@gmail.com, {quazi.hasib, myhossain}@yahoo.com, rashedur@northsouth.edu

ABSTRACT: *Sign Language, a language that uses a system of manual, facial, and other body movements as the means of communication, is the primary means of communication for people having speaking and hearing impairment. This paper uses image processing and fuzzy rule based system to develop an intelligent system which can act as an interpreter between the Bengali sign language and the spoken language. Initially the data is processed from raw images and then the rules are identified by measuring angles. Primarily, the system is tested only for two letters in Bengali.*

Keywords: Fuzzy Logic, Image Processing, Intelligent System, Sign Language

Received: 4 March 2014, Revised 11 April 2014, Accepted 18 April 2014

1. Introduction

People having speaking and hearing impairment are usually deprived of normal communication with other people in the society. Since they cannot talk or hear like normal people, they have to depend on some sort of visual communication in most of the time.

Sign Language is the primary means of communication among the people having speaking and hearing impairment. It is a visual language consisting of various signs, gestures, finger spelling and facial expressions. As like any other language it has also got grammar and vocabulary but uses visual modality for exchanging information [1]. But normal people are usually unaware of these grammars. So when mute or hard of hearing people try to express themselves to other people with the help of these sign language grammars, most normal people do not understand those. As a result it has been seen that communication of a person having speaking or hearing impairment are only limited within his/her family.

At this age of technology, the demand for a computer based system targeting this problem is highly demanding. However, researchers have been attacking the problem for quite some time now and the results are quite promising. Different languages have their own sign language like:

- American Sign Language (ASL)
- Brazilian Sign Language (LIBRAS)
- Australian Sign Language (Auslan)
- Bengali Sign Language (BaSL)

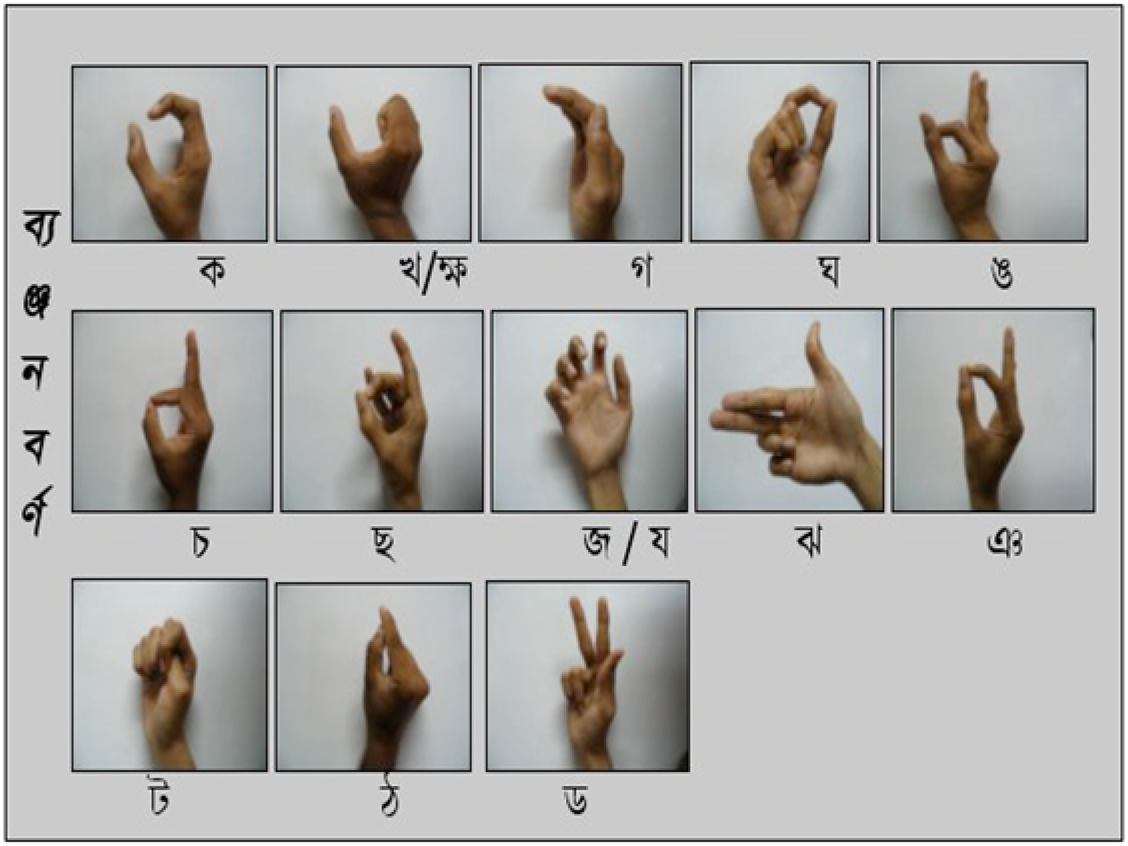

Figure 1. Some BaSL single hand gestures

In respect to Bangladesh, about 16 million people are living with a disability which includes speaking and hearing impairment [2]. Though there are some interesting technologies for speech recognition, not many researchers have been carried out in sign language recognition, especially in Bengali Sign Language Recognition. Our objective is to develop an intelligent system using image processing and fuzzy feature based rule which can act as an interpreter between the Bengali sign language and the spoken language dynamically.

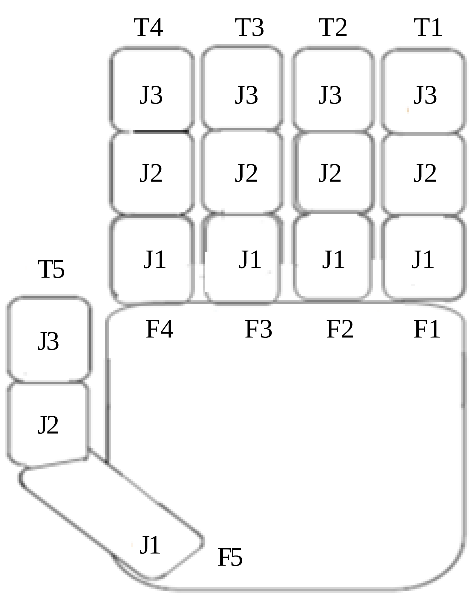

Figure 2. Sample hand marking fingers, joints and tips

We focus our ambition to a static variant of the gesture recognition problem which emphasizes on translating sign language character to regular Bengali character expressed by single hand gesture. Initially, the system is tested only for two letters in Bengali, which are ক and [illegible] from image inputs.

2. Related Works

Quite a few researches have been done on this issue using diverse methods. The earliest attempts at sign recognition were designed for alphanumeric input [3]. Though it did not directly recognize alphabets, it recognized a dialect which was easier to recognize with the sensor on the glove. The "*THETOS*" system translates the Polish text into Polish Sign Language using Natural Language Processing (NLP) [4] and [5]. A visual representation of Australian Sign Language (Auslan), reported in [6], uses 3D animated signer and is known as the "*Auslan Tuition System*" (Avatar). Glove based recognition strategy has been developed in Pakistani sign language, described in [7]. Bedregal, Costa and Demuro introduces a fuzzy rule-based method for the recognition of hand gestures acquired from a data glove, with an application to the recognition of some sample hand gestures of LIBRAS, the Brazilian Sign Language [8]. Bowden & Sarhadi present a method by which the one-state transitions of the English Language are projected into shape space for tracking and model prediction using a HMM like approach [9].

Using HMM, Thad Starner developed an unobtrusive single view camera system that can recognize a subset of American Sign Language (ASL) [10]. A video-based continuous sign language recognition system based on Hidden Markov Models (HMM), that aims for an automatic signer dependent recognition of sign language sentences, based on a lexicon of 97 signs of German Sign Language with one model for each sign [11].

Unlike any other approaches in recognizing the Bengali sign languages used by mute or hard of hearing people, the intelligent system that is developed and described all though out this paper used Image Processing and Fuzzy Feature based rule combined to accomplish the job.

Linguistic term	Notation
STRAIGHT	St
CURVE	Cv
BENT	Bt

Table 1. Linguistic Terms of Linguistic Variables Representing Finger Joints

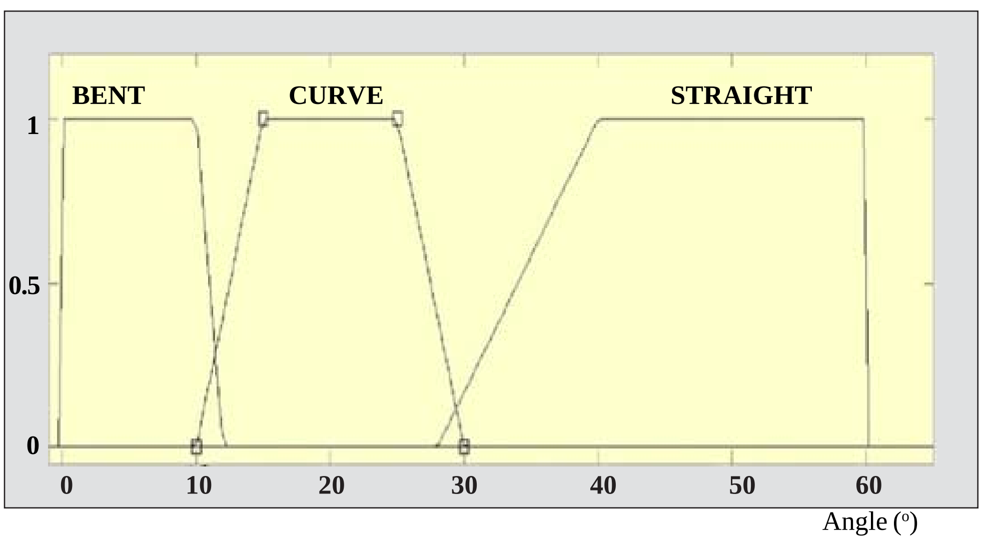

Figure 3. Fuzzification of the linguistic variable F5J2 in the thumb finger F5

3. Fuzzification

The proposed system works in two phases to perform the sign language verification. The fuzzification methods are defined in the first phase and the raw images are processed to identify the fuzzy rules in the second phase. Consider a sample hand as

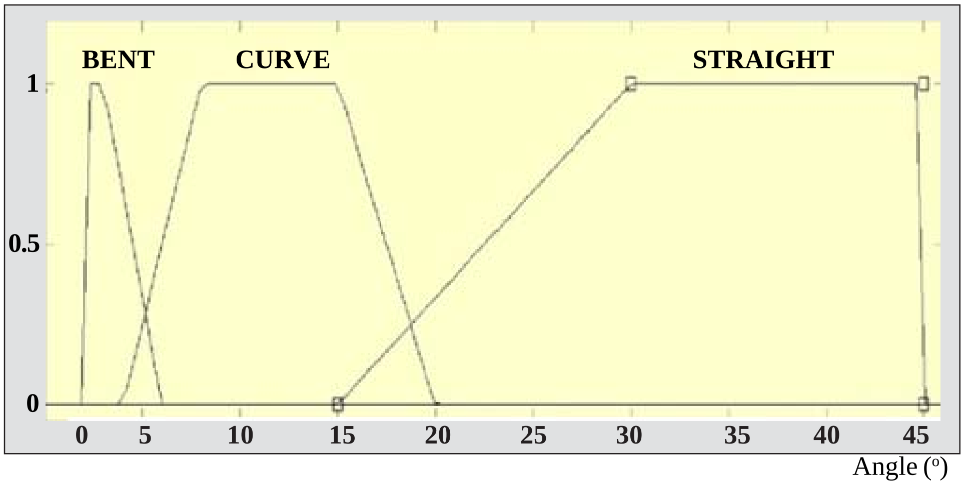

Figure 4. Fuzzification of the linguistic variables for J1 for the remaining fingers

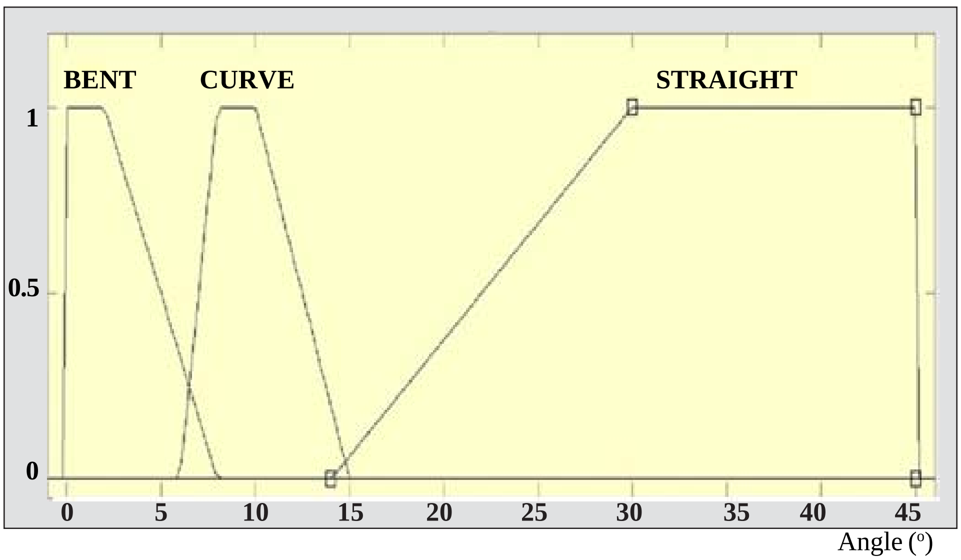

Figure 5. Fuzzification of the linguistic variable F5T5 when J2 is St

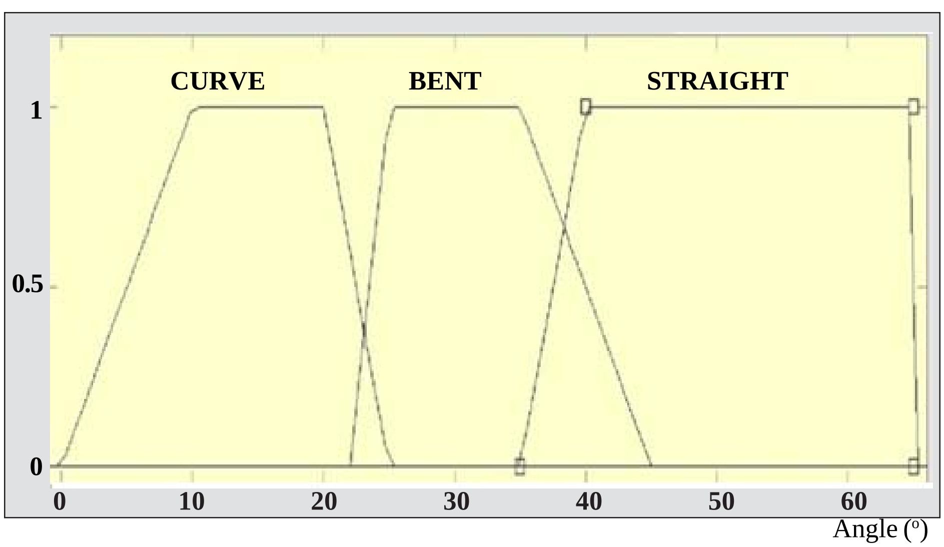

Figure 6. Fuzzification of the linguistic variables for J2 for the remaining fingers when J1 is St

shown in Figure 2. The fingers are labeled as: F1 (little finger), F2 (ring finger), F3 (middle finger), F4 (index finger) and F5

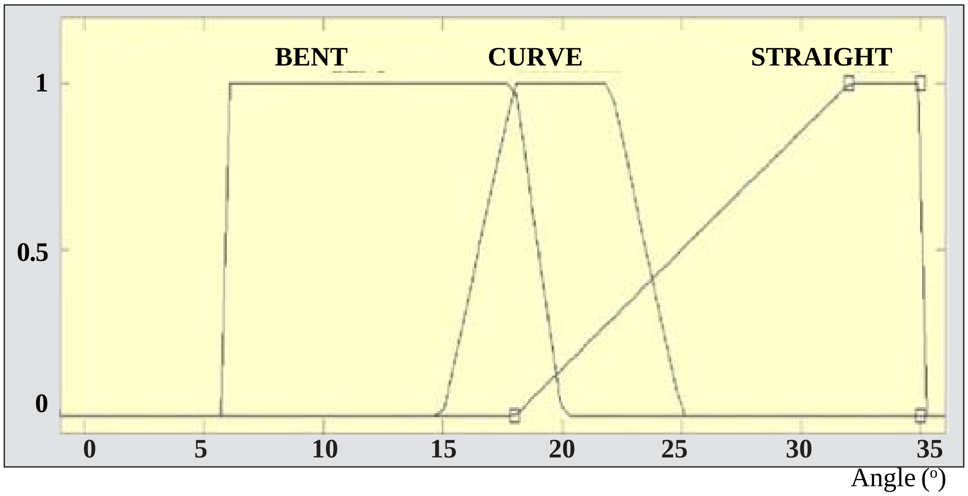

Figure 7. Fuzzification of the linguistic variables for J2 for the remaining fingers when J1 is Cv

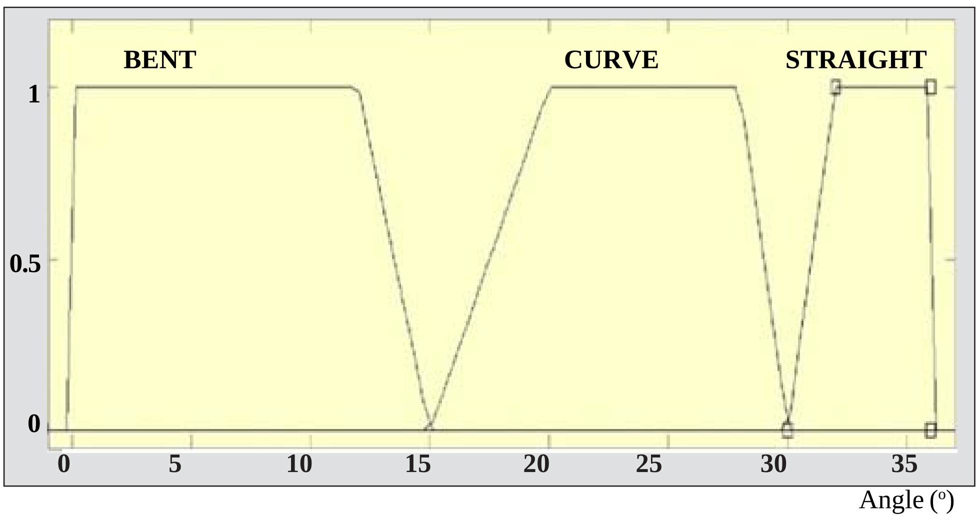

Figure 8. Fuzzification of the linguistic variables for J3 for the remaining fingers when J1 is St and J2 is Bt

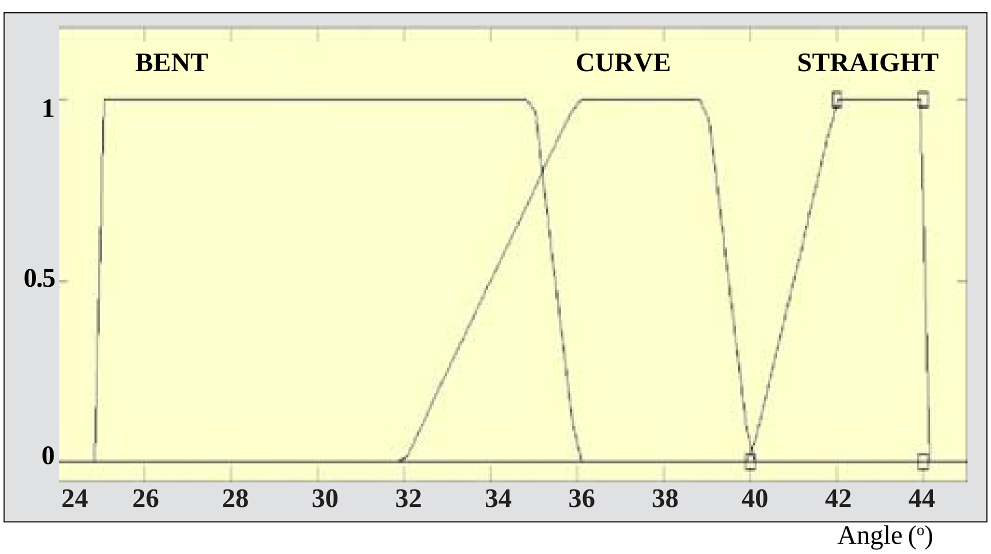

Figure 9. Fuzzification of the linguistic variables for tips of the remaining fingers when J1 is Cv and J2 is Cv to find the position of J3

(thumb). Tips of the fingers are labeled as T1 (little finger), T2 (ring finger), T3 (middle finger), T4 (index finger) and T5 (thumb). The joints in the fingers are labeled as J1 (the knuckle), J2 and J3, for each finger [8].

Each finger position corresponds to a linguistic variable, whose values are linguistic terms, representing typical angles of joints. For the joints in the fingers (linguistic variables F1J1, F1J2, F1J3 etc.) the linguistic terms given in Table I are: STRAIGHT (St), CURVED (Cv) and BENT (Bt) [8].

Here, 27 possible finger configurations are considered. The 27 possible finger configurations determine 27 inference rules that calculate membership degree of each finger to each configuration [8]. For example, If (F1J1 is STRAIGHT) and (F1J2 is CURVED) and (F1J3 is CURVED) then (F1 is StCvCv) [8]

Now for F5, i.e. the thumb finger, there is no noticeable movement for J1. So we always consider it as St. Also, angles for joints of thumb finger is different from other fingers. Figure 3-10. show the fuzzification of the linguistic variables for different fingers with different configurations. Here, X- axis represents the angles (from center) and Y-axis represents the degree of membership.

In the particular case of F5J2, if F5J2 is St, J3 does not show any significant difference in angles with change of position. So in that case, we skip measuring angles for F5J3 and measure the angle for F5T5, i.e. tips of the finger to determine the position of F5J3.

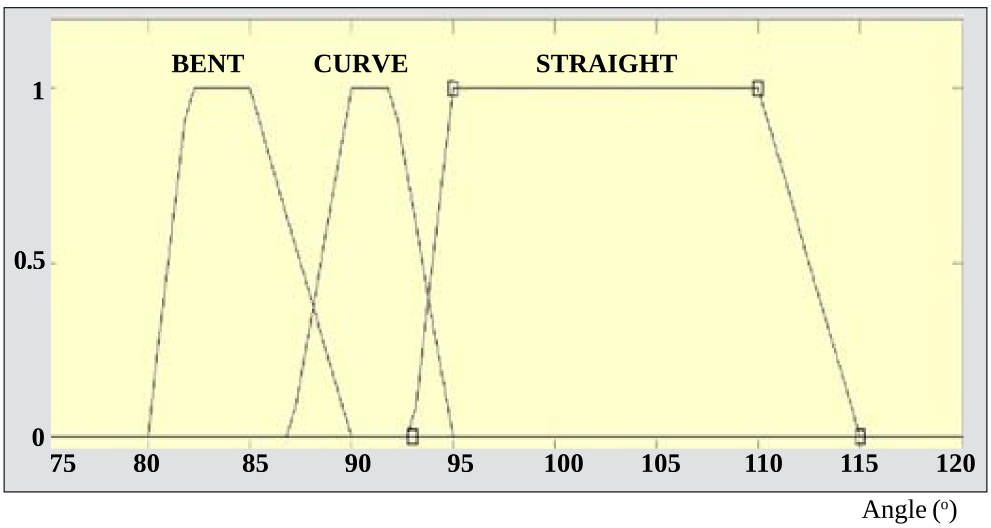

Figure 10. Fuzzification of the linguistic variables for tips of the remaining fingers when J1 is St and J2 is St to find the position of J3

Again, there are different angles for the joint J2 for different positions of J1 for the fingers F1, F2, F3 and F4. So the membership graphs of J2 are different for different positions of J1. Similarly, there are different membership graphs for J3 for different positions of J1 and J2. Again, sometimes J3 does not show any significant change of angles with the change of its positions. In those cases, angles with tips of the fingers are measured from the center and thus positions of J3 are determined.

4. Image Processing

As the inputs are raw images, they need to be converted into suitable form of data before any sort of calculation can be done on them. So the second phase of our work involves some image processing. The block diagram in Figure 11 gives an overview of the image processing system.

4.1 Initialization

Images, when fed to the system are first rendered, then scaled to 480 * 381 pixels for standardization and finally converted into corresponding 480 * 381 2d matrixes according to the pixel colors. Then the images are converted to grayscale so that further processing can be done.

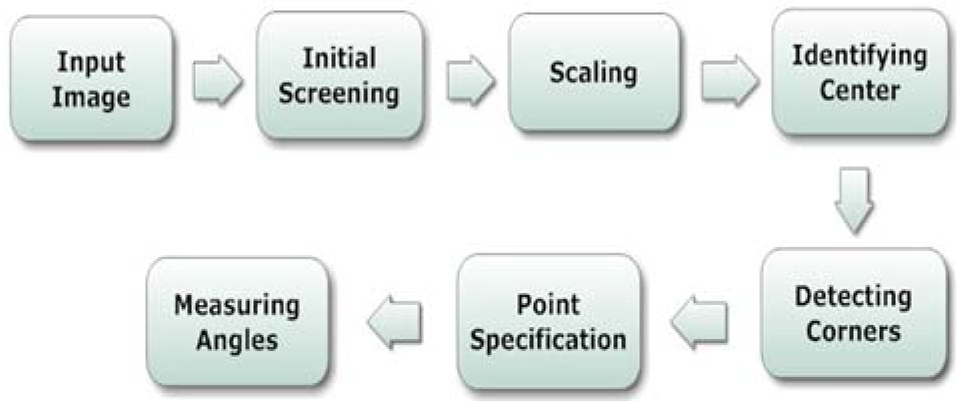

Figure 11. Block Diagram of the Image Processing System

4.2 Identifying Center

After the initialization, we determined the center of the image, which is approximately the palm of the hand, so that we can measure the angle of finger joints from that point and thus identify the specified fuzzy rule for the hand gesture.

4.3 Point Specification by Harris Algorithm

The basic task in this step is to simply detect the finger joints (i.e. J1, J2 and J3) from the processed raw images. We tried to use corner detection techniques to do this task in MATLAB since corners are common interest points in an image.

Corner detection is good for obtaining image features for object tracking and recognition especially for three-dimensional objects from two-dimensional images. Humans understand "*a corner*" easily, but for algorithm we need a more mathematical detection. From various corner detection algorithms such as Moravec, Trajkovic, Wang and Brady, SUSAN etc. [13], we choose to use Harris Algorithm since it was giving better results in finding joints of the fingers.

Harris corner detector is based on the local auto-correlation function of a signal [14]. The local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. Given a shift (x, y) and a point the auto-correlation function is defined as

Hence for small shift, E can be written as

$$\begin{aligned} E_{x,y} &= \sum_{u,v} W_{u,v} [I_{x+u,y+v} - I_{u,v}]^2 \\ &= \sum_{u,v} W_{u,v} [xX + yY + O(x^2, y^2)]^2 \end{aligned} \tag{1}$$

where I denote the image intensities and w specifies the image window. Here, first gradients are approximated by

$$X = \mathrm{I} \otimes (-1, 0, 1) \approx \frac{\delta I}{\delta x}$$

$$Y = \mathrm{I} \otimes (-1, 0, 1) \approx \frac{\delta I}{\delta y}$$

Hence for small shift, E can be written as

$$E(x, \mathrm{y}) = Ax^2 + 2Cxy + By^2 \tag{2}$$

where

$$A = X^2 \otimes w$$

$$B = Y^2 \otimes w$$

$$C = (XY) \otimes w$$

The response is noisy because the window is binary and rectangular, so a smooth circular window is used, for example a Gaussian:

$$W_{u,v} = exp - (u^2 + v^2)\, 2\sigma^2 \tag{2}$$

The operator responds too readily to edges because only the minimum of E is taken into account. To avoid this, the corner measure is reformulated to make use of the variation of E with the direction of shift.

The change, E, for the small shift (x, y) can be concisely written as

$$E\,(x, \mathrm{y}) = (x, \mathrm{y})\, M\,(x, \mathrm{y})^T \tag{4}$$

where the 2 × 2 symmetric matrix M is

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \tag{5}$$

Let α_1 and α_2 be the eigenvalues of M, then we have three cases to consider:

1) Both eigenvalues are high: Interest point (corner)

2) One eigenvalue is high: contour (edge)

3) Both eigenvalues are small: uniform region (constant intensity).

According to Harris algorithm,

- M is symmetric and positive definite that is α_1 and α_2 are > 0
- $\alpha_1 \alpha_2 = \det(M) = \mathrm{AB} - \mathrm{C}^2$
- $\alpha_1 + \alpha_2 = \mathrm{trace}\,(M) = \mathrm{A} + B$
- Corner Response, $R = \det M - k\,(\mathrm{trace}\ M)^2$; where k is an empirical constant, $k = 0.04 - 0.06$

The steps for finding the corner points are as follows:

Step 1)	Find points with large corner response function R (R > threshold)
Step 2)	Take the points of local maxima of R

This algorithm successfully determines the joints of the fingers as shown in Figure 12 (c) and Figure 12 (d). We specify the minimum and maximum number of points to be detected. These points are used to determine the positions of the finger joints by finding angles and thus to categorize the finger configuration.

4.4 Measuring Angles

At this point when the all the corner points that is, the finger joints were found out, the angles of those points are measured with respect to the center to determine their positions. To find the angle between two points, we consider the points as two vectors. Vectors are nothing more than lists of numbers. The number of numbers in each list (each vector), is called the dimension of the vector. The numbers inside the list are named scalars. Scalars can be almost anything that can be added or multiplied.

From Pythagorean Theorem, we know that the dot product of a vector with itself is related to the length of the vector. For example, the length of the vector p, or

$$\|p\| = \sqrt{p} \cdot \sqrt{p} \tag{6}$$

Again, given vectors p and q, the angle between them is given by:

$$\theta = \arccos\left(\frac{p.q}{\|p\|\,\|q\|}\right) \tag{7}$$

In our system, the center of the image is considered as the first vector and the joints of the fingers found by applying the Harris algorithm are considered as the other vector and angles are measured using (7). These angles are used to determine the position of a specific finger joint which shows the way to determine the finger configurations and finally leads to the recognition of the letter.

5. Recognition Process

We illustrate our proposed method by explaining the recognition process of the Bengali letters ক and চ.

First we define the finger configurations for the two letters according to the method described in section 3 of this paper. Combining the set of finger configurations we define the hand configuration for each letter. The hand configurations are basically the fuzzy rules. Then the raw input images are processed and other calculations are done according to the steps described in section 4 of this paper. By comparing the results found from the fuzzification and the image processing phase, a specific character is recognized.

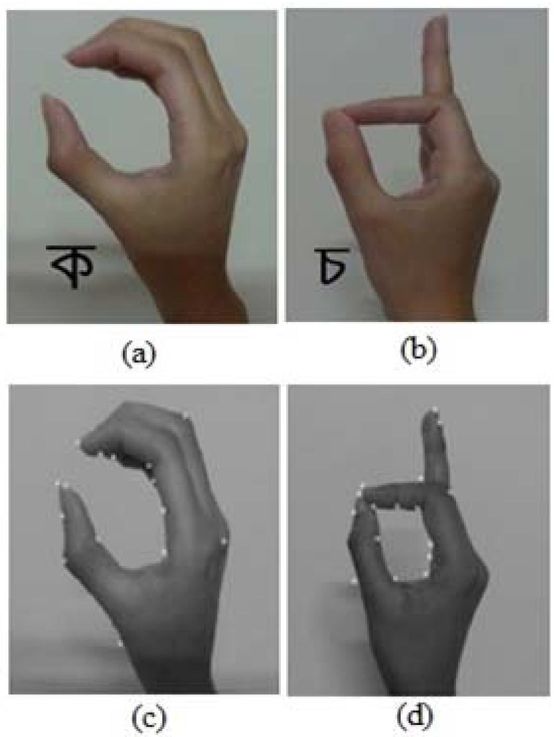

Figure 12. (a) Hand gesture of ক, (b) Hand gesture of চ, (c) Joints determined by Harris Algorithm for ক, (d) Joints determined by Harris Algorithm for চ

For ক in Figure 12 (a), we have finger configuration: (F1J1 is CURVED) and (F1J2 is CURVED) and (F1J3 is CURVED) so (F1 is CvCvCv). Similarly, F2, F3 and F4 are also CvCvCv. (F5J1 is STRAIGHT) and (F5J2 is STRAIGHT) and (F5J3 is CURVED) so (F5 is StStCv).

Likewise for চ in Figure 12 (b), we have finger configuration: (F1J1 is STRAIGHT) and (F1J2 is STRAIGHT) and (F1J3 is STRAIGHT) so (F1 is StStSt). Similarly, F2 and F3 are also StStSt. (F4J1 is STRAIGHT) and (F4J2 is BENT) and (F4J3 is STRAIGHT) so (F4 is StBtSt). (F5J1 is STRAIGHT) and (F5J2 is STRAIGHT) and (F5J3 is CURVED) so (F5 is StStCv).

Now, the hand configurations, i.e. the fuzzy rules are: If (F1 is CvCvCv) and (F2 is CvCvCv) and (F3 is CvCvCv) and (F4 is CvCvCv) and (F5 is StStCv) then the letter is. If (F1 is StStSt) and (F2 is StStSt) and (F3 is StStSt) and (F4 is StBtSt) and (F5 is StStCv) then the letter is.

For ক in Figure 12 (a), angle of F4J1 from center is found to be 10º. From Figure 5, it is seen that at 10º, membership degree of

Bent and Straight is 0, whereas membership degree of curve is 1. So, for index finger i.e. F4, F4J1 is Cv in case of ক. Now, J2 gives different angle ranges depending on the position of J1. As for ক, F4J1 is Cv, we evaluate F4J2 according to Figure 7. The angle for F4J2 from center for Figure 12(a) is found to be 32º which means F4J2 is Cv. Now again, since J1 and J2 both are Cv, we measured J3 accordingly for F4. But it is observed that J3 does not show any significant angle change when both J1 and J3 are curved. So in this case, we measure the angle of the tip of the finger F4, i.e. F4T4 from the center to find the position of J3. It is found that the angle of F4T4 is 38º from the center which corresponds that J3 is also Cv as shown in Figure 9. In this way, we get the finger configuration for F4 in case of the image given in Figure 12 (a) and found that F4 is CvCvCv. Similarly, we found the finger configurations for other fingers and combining all the five fingers, we determined the hand configuration. The measured hand configuration is matched with the defined fuzzy rule and thus the corresponding letter is recognized.

6. Results and Limitations

In our test data, total number of input images was 22. Among them, 10 images were gestures of and 12 images were gestures of of two different people. Since we have used raw images of hands, we have only been able to get the angles of fingers from one viewpoint. All the fingers are not visible from one sight of the images. For example, in Figure 12 (a), the fingers F1, F2 and F3 are not clearly visible. So it is not possible to find their positions from that image. In our system, such cases are handled by defining another rule. The rule says that if all the fingers are not visible, the configuration of the unseen fingers are defined same as the configuration of the last finger measured. For example, in case of Figure 12 (a), we started measuring angles from finger F5, and then F4. Now since other fingers are not visible, configuration of F1, F2 and F3 are defined same as F4. Similarly, in case of Figure 12 (b), the last measured finger is F3 and the fingers F1 and F2 are not visible. So the configuration of F1 and F2 will be same as F3. This method works well to determine the two letters, which we used to test, but this is not an optimum solution. This method may not work in case of other letters in Bengali.

A best possible solution to this problem can be obtained by getting more than one image for each gesture from different viewpoints so that all fingers are visible. Then combining all the data from all the images, a letter can be recognized. In that case, we will need to define different membership functions for each finger joint for different viewpoints since angles will change significantly if viewpoint changes.

More efficient output of letter recognition can be obtained by using video input in this system. If video input is used, it will be easier to get views of hand gestures from different sights. Also this method can be very effective for letter or expression recognition if input is taken from data glove. Additionally, the limitations of our proposed system can be minimized by advanced development of computer vision technology.

7. Conclusion

In this paper, we proposed a fuzzy method for Bengali Sign Language recognition. The related methods like Image processing were also investigated and used. The system was implemented using MATLAB. For our experiment, the gestures of two Bengali letter ক and were used. Initial experimentation indicated promising results.

We conclude that local images of hand and finger can be treated as clues for sign language as they are good enough to provide feature for discrimination. In fact, the recognition of sign language is a challenging task. Different positions of hand and finger represent different letters of sign language. There might be some similarity in them in terms of finger position and gesture but definitely each hand configuration combining all the fingers will represent one unique letter or expression. This is an interesting phenomenon and may be an important clue for the future research to improve the language performance. More work should be conducted to validate this technique on BaSL system as done in case of other language systems.

References

[1] Armstrong, D. F., Stokoe, W. C., Wilcox, S. E. (1995). Gesture and the nature of language. Australia: Cambridge University Press.

[2] Center for Disability in Development, [Online]. Available: http://www.cdd.org.bd/index.php?option=com_content&view=article&id=68&Itemid=512. [Accessed: Jan. 20, 2013].

[3] Grimes, G. J. (1983). Digital Data Entry Glove Interface Device, Bell Telephone Laboratories, Murray Hill, N.J., US Patent 4, 414, 537, Nov. 8.

[4] Suszczanska, N., Szmal, P., Kulikow, S. (2005). Continuous text translation using text modeling in the Thetos system, *In*: Proc. *International Journal of Computational Intelligence*., 1 (4) 255-258.

[5] Binh, N. D., Ejima, T. (2005). Hand Gesture Recognition Using Fuzzy Neural Network, *In*: Proc. ICGST Conf. Graphics, Vision and Image Proces., p. 1–6.

[6] Yeates, S., Holden, E. J., Owens, R. (2003). An animated Auslan Tuition System, *Machine Graphics & Vision International Journal*., 12 (2) 203 – 214, Feb.

[7] Alvi, A. K., et al. (2005). Pakistan Sign Language Recognition Using Statistical Template Matching, *World Academy of Science, Engineering and Technology*, p. 52-54, Cairo.

[8] Bedregal, B. C. R., Costa, A. C. R., Dimuro, G. P. (2006). Fuzzy rule-based hand gesture recognition, in Artificial Intelligence in Theory And Practice (M. Bramer, ed.), IFIP, 217, p. 285-294, Springer, Boston.

[9] Bowden, R., Sarhadi, M. (2000). Building temporal models for gesture recognition, *In*: Proc. British Machine Vision Conference, p. 32-41.

[10] Starner, T. (1995). Visual Recognition of American Sign Language Using Hidden Markov Models, Master's thesis, MIT, Media Laboratory, Feb.

[11] Bauer, B., Hienz, H., Kraiss, K. (2000). Video-Based Continuous Sign Language Recognition Using Statistical Methods, *In*: Proc. International Conference on Pattern Recognition, *Pattern Recognition and Neural Networks*, 2, p. 463-466, Sep. 3-7.

[12] Rahim, F. M., Mursalin, T.E., Sultana, N. (2010). Intelligent Sign Language Verification System – Using Image Processing, Clustering and Neural Network Concepts, *International Journal of Engineering Computer Science and Mathematics*, 1 (1) 43-56, January-June.

[13] Corner Detection [Online]. Available: http://en.wikipedia.org/wiki/Corner_detection. [Accessed: Jan. 23, 2013].

[14] Harris, C., Stephens, M. (1988). A Combined Corner and Edge Detector, *In*: Proc. Alvey Vision Conf., Univ. Manchester, p. 147-151.