ALPHABET RECOGNITION OF AMERICAN SIGN LANGUAGE: A HAND GESTURE RECOGNITION approach USING SIFT ALGORITHM

Nachamai. M[1]

[1]Department of Computer Science, Christ University, Bangalore, India
nachamai.m@christuniversity.in

ABSTRACT

This paper is a sincere attempt to recognize english alphabets as part of hand gesture recognition, using the SIFT algorithm. The novelty of this approach is, it is a space, size, illumination and rotation invariant approach. The approach has evolved to work well with both the standard American Sign Language (ASL) database and home-made database. The problem of alphabet recognition may seem to sound small but the intricacies involved in it cannot be solved using a single algorithm. Hand gesture recognition is a complicated task. A one stop solution is still not evolved for any recognition process. This paper has tried to approach this in a simple but efficient manner using the basic SIFT algorithm for recognition. The efficacy of the approach is proved well through the results obtained, invariably on both the datasets.

KEYWORDS

Gesture Recognition, American Sign Language, SIFT algorithm, Alphabet Recognition.

1. INTRODUCTION

Gestures are meaningful body movements which is capable of expressing something in a communication, although gesture finds a place to catalogue itself into non-verbal communication it prominently reaches well to the other end of communication. Gesture is motion of body that contains information [1]. Salutation to the flag is a well-known gesture, which means respect. The basic aim of a gesture is to convey information or interact with the environment. Based on the location of origination of gesture in the body, it can be categorized into hand and arm gestures, head and face gestures, and body gestures. This paper is the work on the first category i.e. hand gestures. The probable sub-divisions on the hand gestures are static gesture, dynamic gesture, and static and dynamic gesture [2]. Static gestures are the one in which a certain stationed pose is assumed. Dynamic includes a gesture movement that is defined. Gestures that involve and have an embedded pattern of both static and dynamic movements fall into the last category. Sign language is the best suited example of this category. Gesture research is termed as a complex research area, as there exists many-to-one mappings from concepts to gestures and gestures to concepts. The major drawback in pursuing research with gestures is that they are ambiguous and incompletely specified [3]. Gestures vary between individual as each one's convey of information may not be the same in a communication even though they ought to be so. When the same individual's case is taken it may not be the same, for the same meaning during different instances. With all these difficulties and ambiguities prevailing gesture research is still alive and numerous research work proceeds in this line. The reason could be the vast application areas it finds itself. One of the major look out is human computer interaction with gestures. One of the major applications includes navigating or interacting in a virtual or 3D environment [4]. The basic

intention is its gaining popularity, it helps in communicating to an electronic system or human placed from a distance. Once a gesture is made by the user the system must be able to identify the gesture made that is 'Gesture Recognition". The primary aim of this work is to create a system which can identify specific hand gestures and convey information to the system. The system works on dynamic gesture made representing the American Sign Language for alphabets.

2. REVIEW OF LITERATURE

Computers are invariably used by everyone extensively in today's world; one of the major areas of prominence is the human computer interface. The interaction between the human and machines is directly proportional to the utilization of the system. The better user friendly interface, scores a good usage statistics. Attempts in making a computer understand facial expressions, speech, and human gestures are paving to create a better human computer interaction [5]. Gestures are the non-verbal medium of communication. Hand gestures are probably the most common among the gestures. One of the applications of hand gestures is sign language. Sign language is the raw and the original form of communication that existed even before spoken language came into picture. Sign language is being used extensively by the hearing impaired, in the world of sports, in religious practices and also in work places [6]. In a number of jobs, globally considering, sign language is prominent part of the world. Recognition of gestures representing words is undoubtedly a difficult recognition task. Research is restricted to small scale systems in recognition of sign language.Real time tracking of gestures from hand movement is difficult than the face recognition[7]. A comparison has been done in [8] for still and moving image recognition. Sign language recognition has been extensively tried using mathematical models. [9] Explains the deployment of Support Vector Machine for sign language recognition.

3. METHODS AND MATERIALS

A gesture recognition system has four components motion modelling, motion analysis, machine learning and pattern recognition. Since the gestures considered are dynamic in nature a modelling and analysis of the same was a mandatory necessity. A gesture can convey information which is usually quantified as spatial information, pathic information, symbolic information and affective information. The work has attempted to identify the symbolic information displayed by the user. The ASL database used for identification consists of 26 English alphabets. The ASL database snapshot is shown in fig.1.

Figure 1. American Sign Language notation for the English alphabets [10]

3.1 Pre-processing

It becomes an unwritten rule or a mandate for any image processing system to go through a rigid pre-processing. This system also follows that mandate. The process starts with filtering noise, followed by an image adjust; histogram equalization and image normalization which has proved to be a good congregation structure for pre-processing images. The pre-processing also involves image subtraction of the background scene, since it would turn out to be an inconvincible problem to work with many noise interferences from the background objects. Image background is assumed to have subspaces, and reduced from the region of interest [11]. Simple sobel edge detection was applied to track the hand object on the screen. The maximum depth of the image was calculated and stored. The system does not store it as logical values of 0 and 1, the RGB image is continued in the system after pre-processing too.

3.2 Feature vector composition

The representation captures the hand shape, position of the hand, orientation and movement (if any). The region of interest [12] i.e. hand was identified, from where feature vector was to be framed. The feature vector composed for the American Sign Language standard database samples stored consists of .jpg files of existing database along with a few real-time or home-made images. The keypoints derived from the image are placed in an array. All image pixel values that are greater than zero are considered as keypoints and the keypoint array gets generated. The match performance based on similarity measures is not made for every point; instead a dimensionality reduction is done. It is taken as the final feature vector. Only retain the keypoints in which the ratio of the vector angles from the nearest to the second nearest neighbour is more.
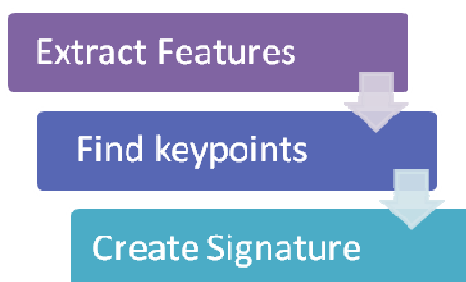
Figure 2. Basic flow of vector composition.

3.3 Methodology

The method or algorithm adopted is called Scale Invariant Feature Transform (SIFT). When considering images, variance is one major factor that comes when the image appears in a large screen. The image window size might be standard, but the image size within the window may vary in real-time. Basically, there are five types of common invariance that could be found in images, scale invariance, rotation invariance, illumination invariance, perspective invariance and affine transformations. As a basic and first step in building robust gesture recognition system the scale invariance, illumination invariance and rotation invariance is handled in this work. The SIFT algorithm helps in managing this invariance. The method followed is depicted in the figure 2. The feature extraction is done by first finding the key points. The scale and location variance are eliminated at this stage by sub pixel localization and edge elimination. Sub pixel elimination is done by down sampling the region of interest considered. The edge is identified by the Sobel edge detection method [13] and cropped. The signature images are derived from the image

gradients which are sampled over 16*16 array of locations in scale space, then an array of orientation histograms are drawn for the same. The figure 3 shows the image gradients and the keypoint descriptors derived for an image.
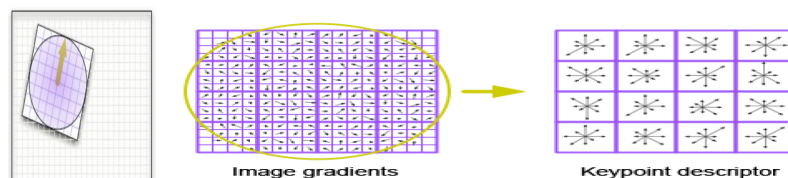

Figure 3. Signature image

The method adopted to achieve this scale invariance is scale space Difference of Gaussian (DOG method). The figure 4 shows the working of the DOG method. The scale works octave by octave.
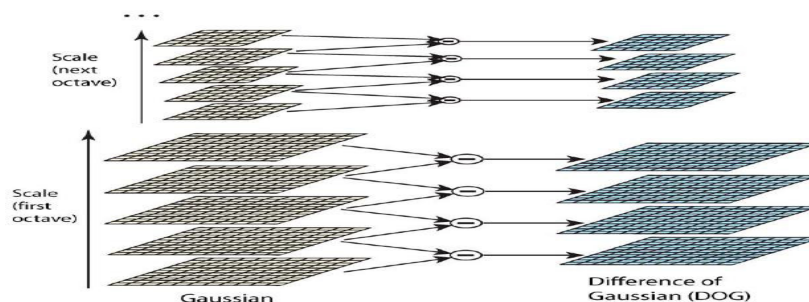

Figure 4. Difference of Gaussian

Every scale is taken in term of an octave and the change in between the next octave is calculated as a Gaussian function. Since the features are derived from the difference of the images, if the feature is repeatedly present in between difference of Gaussians it is scale invariant and it is retained. This paves way as a major key factor for the performance of the system.

3.4 Algorithm

The algorithm used is LOWE's SIFT algorithm [14].SIFT is an invariance algorithm and because of that feature its results are promising for real time as well as formatted images. The scale invariance is the main intention of selecting this algorithm. SIFT as defined by Lowe is a histogram of gradients. The algorithm packages keypoints in each pixel location as [row, col, scale]. The acos of keyvectors are sorted and the first 128 values are used as the feature vector for an image. The input image is compared with all its keypoints with the database image vectors where the nearest neighbour has angle less than the distance ratio, the keypoints are taken as matched. The maximum keypoints matched image is retrieved or recognized as that character. The algorithm executed is depicted in fig 5. The flow explains the step by step process of the algorithm implemented.

4. EXPERIMENTAL RESULTS

The implementation of the algorithm was done in Matlab. The standard American Sign Language dataset was used. The dataset [15] comprised of all the 26 alphabets, and 10 alphabet repeat entries with difference in lighting and orientation. The 80% of the test sample was used for training and 20% for testing. The implementation gave 100% accuracy in identifying the test sample for this dataset. Sample images used for training are shown in the figure 6.
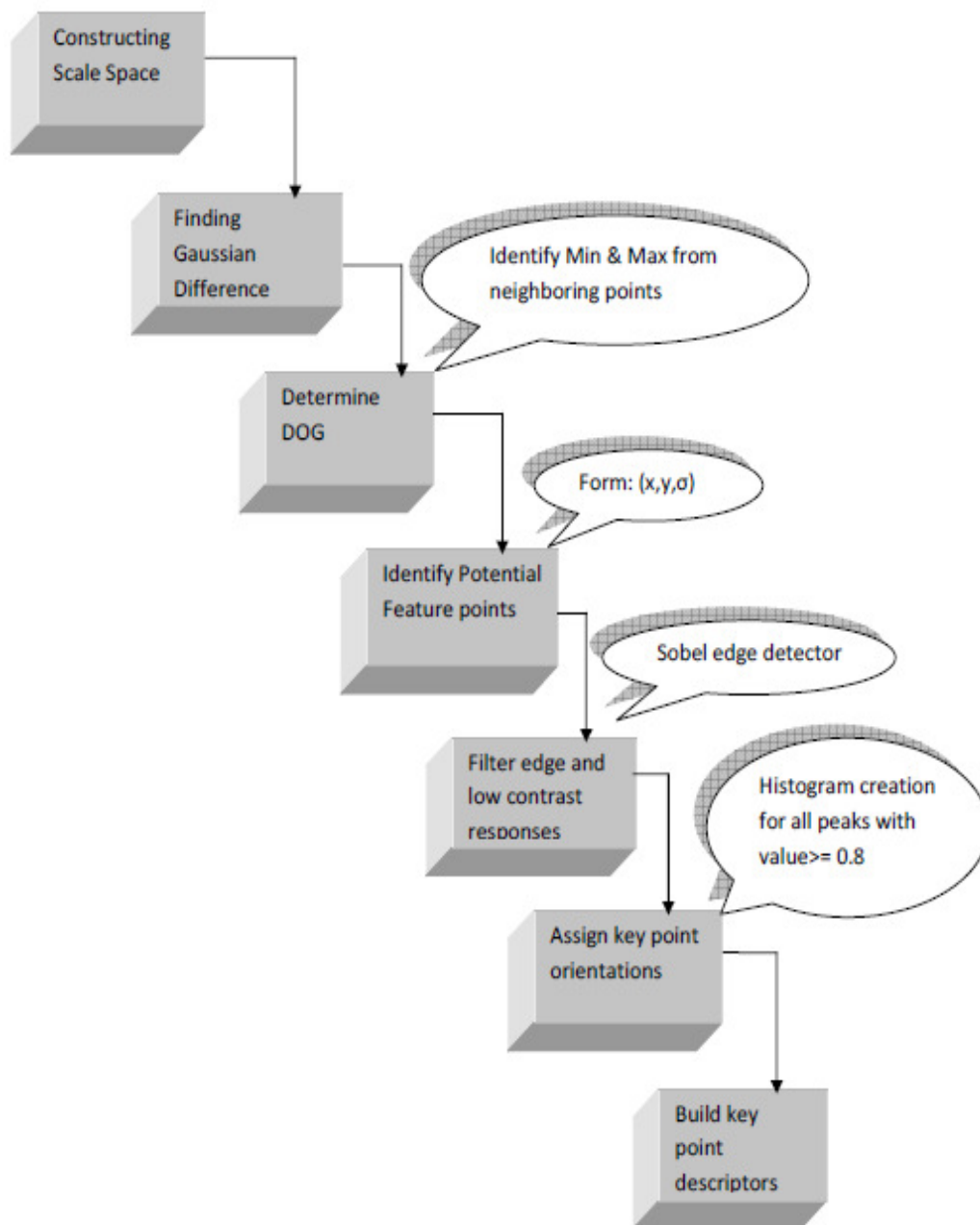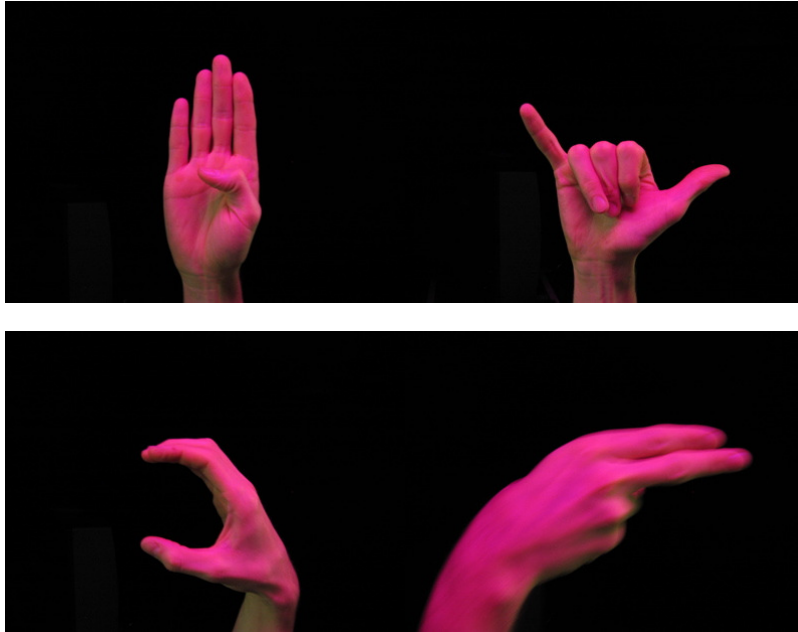
Figure 5. Flow of the method

The dataset-I i.e. matching the images from within the database itself was the fist experiment conducted. The sign character input was checking the match within the database. For Example the test input "B" is given as input, the relevant or most similar image in the database within dataset-I was found and retrieved. The figure 7, 8 and 9shows the Matlab command windows(matching process).
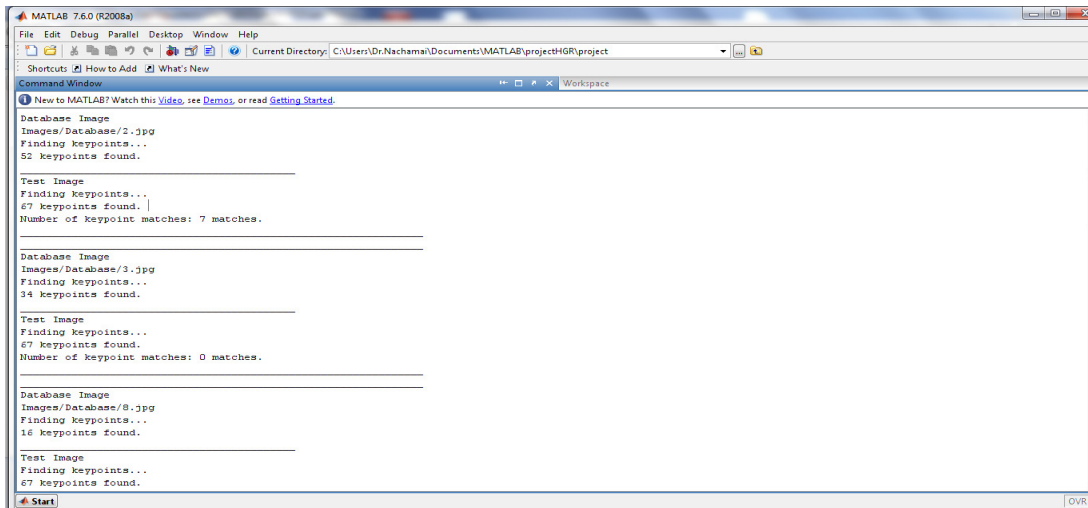

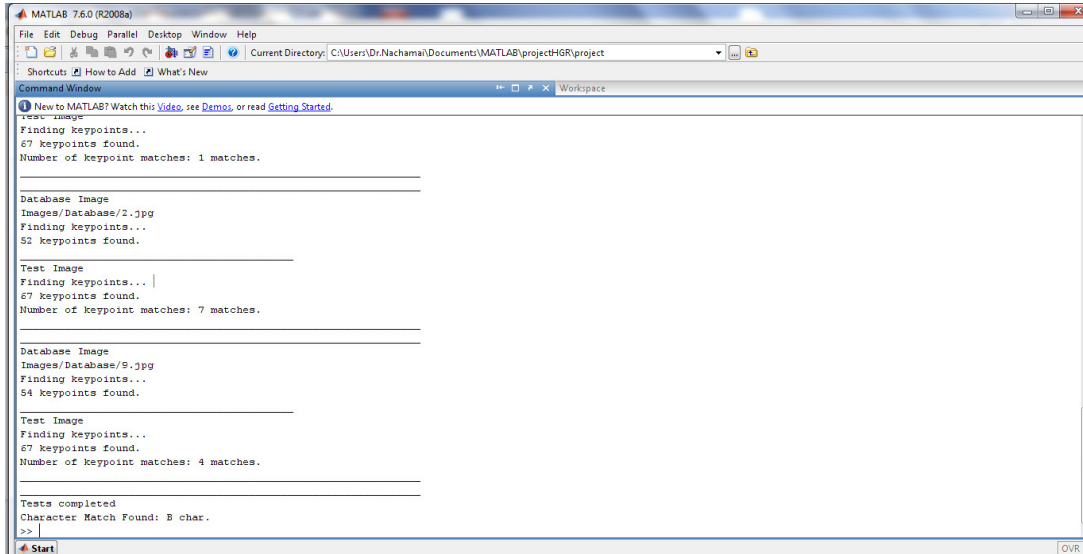Figure 7. Matlab window

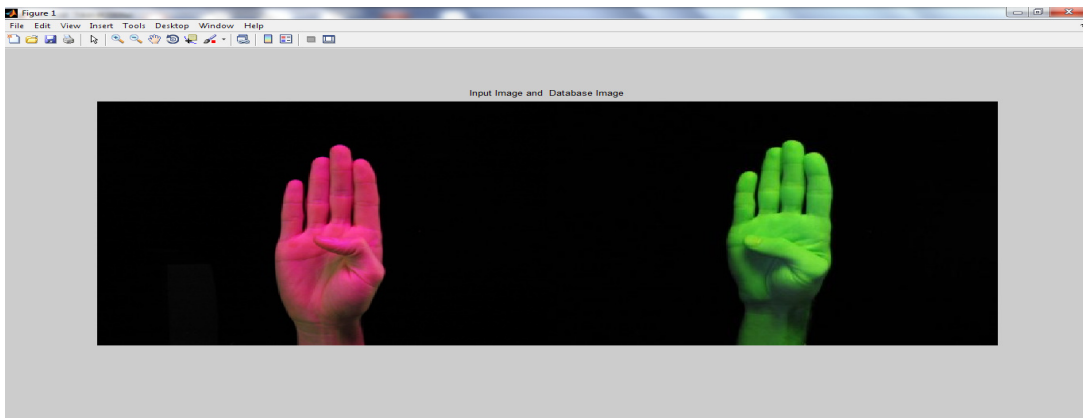and the retrieved equivalent image from the dataset.

Figure 8.Matlab window

Figure 9. "B" matched letter

From the figure 9 it is clear that even if the images do not match completely it terms of colour, and thumb finger orientation the system has found maximum key point matches with this image and has taken it as the match image from the database. It is clear from the images the thumb finger placement does not match equally.

The second set of experiments was conducted with home-made images. The images was taken with a webcam and used for testing the vulnerability and performance of the system. The figure 10, 11 and 12 shows the test sample taken from an user for the same letter" B", and matched.
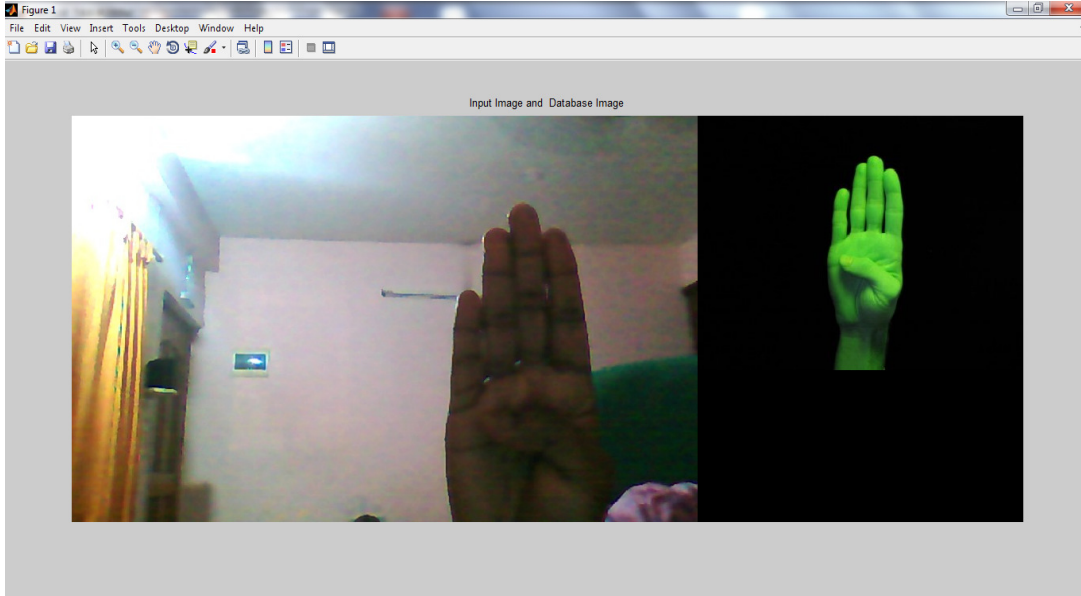
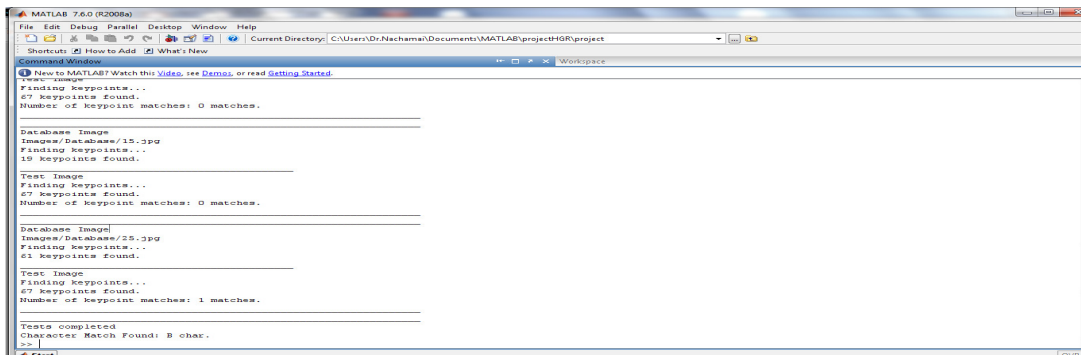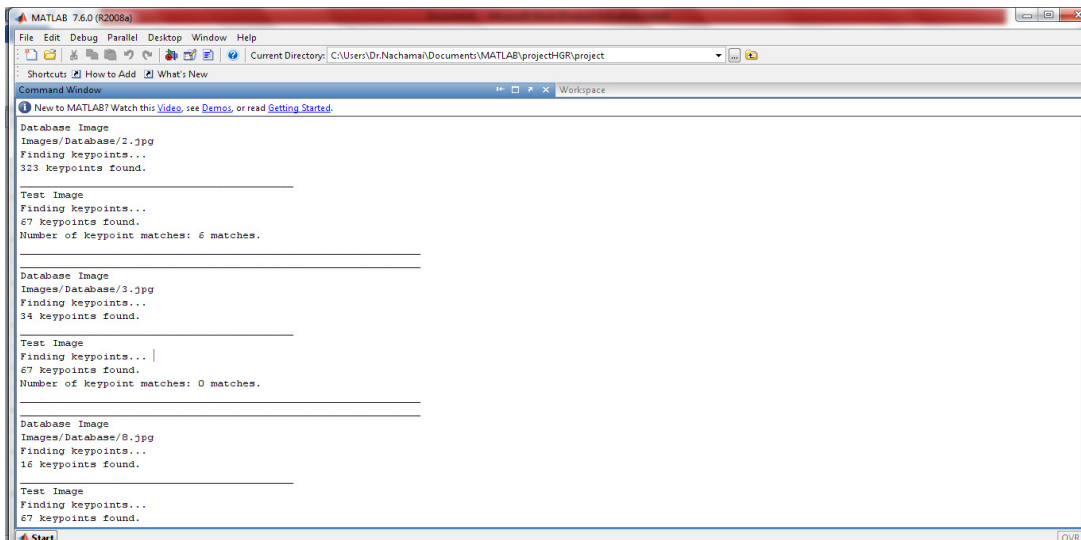Figure 10. Home-made database image match for letter "B".

Figure 11.MatlabWindow

Figure 12.Matlab Window

This test sample has a different background, lighting, scale, and illumination variance when compared to the dataset-I. The system has calculated maximum keypoints for the character and identified it.

The experiments were extended with different orientations, less significant empty spaces, and different hand sizes. These extended experiment results are shown below.
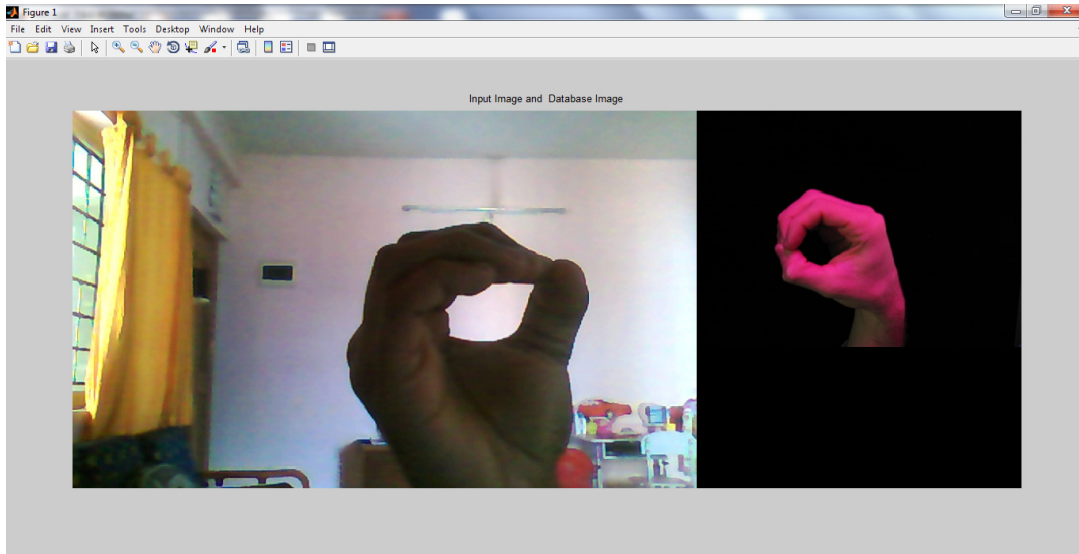

Figure 13. Home-made database image match for letter "O".

This figure 13 shows that the system was able to identify the letter "O", the test sample is a left hand image and the training sample was a right hand image. So the system is robust enough in finding the sign in whatever hand it is made.
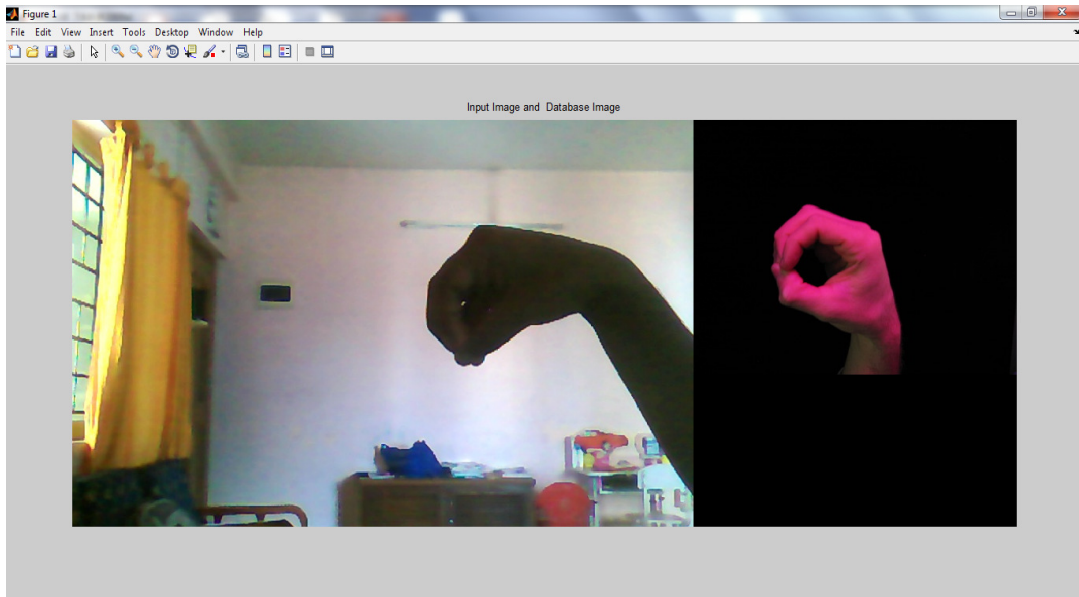

Figure 14. Home-made database image match for letter "O" [Different orientation].

The figure 14 shows that the system is capable of handling a completely mishap orientation in an image. Considering the training image is in an orientation of X-axis the test image is on an orientation of Y-axis.
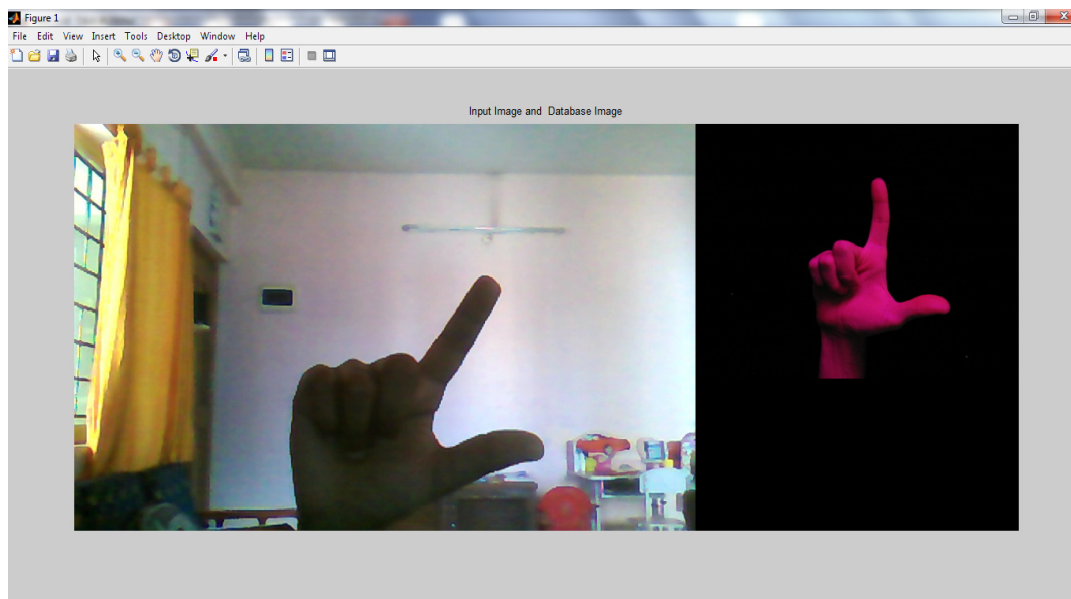

Figure 15. Home-made database image match for letter "L"[Small size hand].

This figure 15 shows that the system works perfectly fine with any size of a hand. The hand samples in the figure 14 and figure 15 was taken from a six year old child. The system is capable of working with and can track any sized hands for the sign.

5. CONCLUSIONS AND FUTURE WORK

The attempt on building a robust system for recognizing American Sign Language which happens to be an unavoidable part of hand gestures has yielded successful results. The system has shown good performance with the database of American Sign Language. The system has also qualitatively attempted recognition of real time images. The future enhancement to the current work could include the other invariance. The system can also to be tested on other databases and with two hand signs.

REFERENCES

[1]   G.Kurtenbach & E.A.Hulteen, (1990) Gestures in Human-Computer Communication,The Art of Human-Computer Interface Design. Reading, Mass.: Addison-Wesley Publications.
[2]   Noor A. Ibraheem&Rafiqul Z. Khan, (2012) "Vision Based Gesture Recognition Using Neural Networks Approaches: A Review", International Journal of human Computer Interaction (IJHCI), Vol. 3, No. 1.
[3]   ZannatulFerdousi, (2008) Design and Development of a Real-time Gesture Recognition System , U.M.I Publishers.
[4]   H. Dardas & Mohammad Alhaj, (year) "Hand Gesture Interaction with a 3D Virtual Environment", The Research Bulletin of Jordan ACM, Vol. 2, No. 3, pp 86-94.
[5]   Ming-Hsuan Yang & Ahuja, Narendra, (2001)," Face Detection and Gesture Recognition for Human-Computer Interaction", Vol. 1, Springer publications.
[6]   Sanjay Meena,(2011), "A Study on Hand Gesture Recognition Technique", National Institute Of Technology, India(thesis-M.Tech).

[7]   Thad Starner & Alex Pentland ,(1995), Real-Time American Sign Language Recognition from Video Using Hidden Markov Models, (IEEE), pp. 265-271.

[8]   D. Kumarage, S. Fernando, P. Fernando, D. Madushanka& R. Samarasinghe, (2011), "Real-time Sign Language Gesture Recognition Using Still-Image Comparison & Motion Recognition", 6th International Conference on Industrial and Information Systems(ICIIS), Sri Lanka.

[9]   M.S. Sinith, Soorej G Kamal, Nisha B.,Nayana S, KiranSurendran, & Jith P S, (2012), " Sign Language Recognition Using Support Vector Machine", International Conference on Advances in Computing and Communications (IEEE), pp.122-126.

[10] Vaishali, S. Kulkarni, & Dr. S.D. Lokhande(2010), Appearance based recognition of American Sign Language using gesture segmentation (IJCSE), Vol. 2, No. 3, pp. 560-565.

[11] Md. Hasanuzzaman, 1S. M. Tareeq, M. A. Bhuiyan, Y. Shirai & H. Ueno, (2009), "ASL characters recognition using person specific subspace method", Journal of Science and Technology, Vol. 4,No. 2, pp. 39-53. [12]Ulka S. Rokade,DharmpalDoye, & ManeshKokare, (2009), "Hand Gesture Recognition Using Object BasedKey Frame Selection", International Conference on Digital Image Processing(IEEE), pp-288-291.

[13] N. C. Woods, O.B. Longe, & A.B.C. Roberts, (2012), "A Sobel Edge Detection Algorithm Based System for Analyzing and Classifying Image Based Spam ", Journal of Emerging Trends in Computing and Information Sciences ,Vol. 3, No. 4, pp. 506-512,

[14] David G. Lowe, (2004), "Distinctive Image Features from Scale-Invariant Keypoints", IJ computer Vision, pp. 1-24.

[15] A.L.C. Barczak, N.H. Reyes, M. Abastillas, A. Piccio & T. Susnjak, (2011), "A New 2D Static Hand Gesture Colour Image Dataset for ASL Gestures", Vol. 15, Res. Lett. Inf. Math. Sci., pp. 12–20.