International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-8)

Research Article | **August 2015**

Scale Invariant Feature Transform based Sign Language Recognition for Deaf People

Jadhav Narendra S
Department of ETC Engineering, SGGSIE&T,
Nanded, India

Abstract:

Sign language is one of the best methods used to communicate with deaf and dumb people and robots. The SIFT (Scale Invariant Feature transform) algorithm takes an image and transforms it into a collection of local feature vectors. SIFT method is used to extract all the features of certain types of signs, then formed theirs code book from all the features by using K means clustering and finally classified by using the multiclass support vector machine. This paper presentssystem of sign language recognition using bag-of-words and multiclass support vector machine (SVM) approach that uses Scale invariant feature of image.We are used American Sign Language (ASL) as an example of a gestural language(GL). The objective of the paper is to decode ASL Image into the appropriate alphabets.

Keywords: SIFT, ASL, SVM, GL, ISL, Bag of words, SL

I. INTRODUCTION

Loss of hearing can cause people to become isolated and lonely, having a tremendous effect on both their social and working life. Looking up the meaning of a sign is not a straightforward task. Sign Language is the well structured code gesture; every gesture has meaning assigned to it. Sign Language (SL) is the only means of communication for deaf people. With advancement of science and technology many techniques have been developed not only to minimize the problem of deaf people but also to implement it in different fields. It becomes difficult finding a well experienced and educated translator for the sign language every time and everywhere but human-computer interaction system for this can be installed anywhere possible. The motivation for developing such helpful application came from the fact that it would prove to be of utmost importance for socially aiding people and how it would help increasingly for social awareness as well. There are different categories of sign languages are there, Indian Sign Language, British Sign Language, American Sign Language etc[2]. In our approach, having encountered an unknown sign, the user can simply perform the sign in front of a webcam. Then, the system compares the input sign with videos of signs stored in the system database, and presents the most similar signs (and potentially also their English translations) to the user and also produce the audio of that specific character. The user can then view the results and decide which (if any) of those results is correct. It will not only benefit the deaf and dumb people of India but also could be used in various applications in the technology field.
The organization of the paper is as follows. In Section II,related works including various algorithms and techniques for sign language recognition. In Section III explain system functionality for recognition. In Section IV, the feature extraction technique. In Section V, the classification result for ASL character recognition. Finally, Section VI concludes the paper with future scope.

II. LITERATURE REVIEW

In last two decades, human hand gesture recognition provides a natural way to interact and communicate with machines has grabs much attention of many researchers around the globe. Various algorithms and techniques for recognizing hand gesture had been introduced by the researchers. Conventionally, for hand gesture recognition, the system should be consisting of four stages which are image acquisition, hand features extraction, processing extracted features and hand gesture recognition [4]. The block diagram shown in Figure.2 depicts the hand gesture recognition steps that are commonly applied by the researchers. Hand gesture recognition is a complex problem that has been dealt with in many different ways. Grobel and Assan (1996)[12] used HMMs to recognize isolated signs with 91.3% accuracy out of 262 sign vocabulary. They extracted the features from video recording of signers wearing colored gloves. Kjeldsen and Kender (1996) [10] suggest an algorithm of skin color segmentation in the HSV color space and use a back propagation neural network to recognize gestures from the segmented hand images. Hongo et al. (2000)[3] use a skin color segmentation technique in order to segment the region of interest and then recognize the gestures by extracting directional features and using linear discriminant analysis. Manresa et al. (2000)[11] propose a method of three main steps: (i) hand segmentation based on skin color information, (ii)tracking of the position and the orientation of the hand by using a pixel based tracking for the temporal update of the hand state and (iii) estimation of the hand state in order to extract several hand features to define a deterministic process of gesture recognition. Imagawa, Matsuo, Taniguchi, Arita, and Igi.(2000) [13] present " A local feature extraction technique is employed to detect hand shapes in sign language recognition". They used appearance based Eigen method to detect hand shapes. Using a clustering technique, they generate clusters of hand shapes on an Eigen space. They have achieved accuracy of around 93% recognition of 160

words. Vogler and Metaxas (*1997*) used computer vision methods and HMMs to recognize continuous American Sign Language sentences with a vocabulary of 53 signs. They modeled context-dependent HMMs to alleviate the effects of movement epenthesis. An accuracy of 89.9% was observed. Triesch and Von der Malsburg (2001)[9] propose a computer vision system that is based on Elastic Graph Matching, which is extended in order to allow combinations of different feature types at the graph nodes. TimiOjalaet. al. (2002) [2] they suggested the method for texture classification using Local Binary Patterns. Rotation Invariant method is widely used for texture classification and recognition. Although it achieved a high accuracy of 96%, their system was limited only to 10 distinct signs. H. Desa, and W. Majid (2009) [5] developed an ASL finger spelling system using a Cyber glove, with the use of neural networks for data segmentation, feature classifier, and sign recognition. Using a tree-structured neural classifying vector quantize, 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.

An SL System was proposed by Kishore Kumar et al (2011) [4] to automatically recognize gestures of words and convert them to text or audio format. The system has a success rate of 91%. DeepikaTewari and Sanjay Kumar Srivastava (2012) proposed a method for the recognition of Indian Sign Language [5] in which gesture frames were taken with a digital camera. The recognition rate achieved was 80%. Pol Ashwin et al (2013) [13-14] proposed Sign Language Recognition System Using SIFT Based Approach for ASL.

Joe Naoum-Sawaya et al proposed a system for American Sign Language Recognition [6]. The
hands were detected by performing motion detection over a static background. The accuracy for this system is 96% in daylight with distinct backgrounds. Jaspreet Kaur et al (2012) uses modified SIFT algorithm for American Sign Language[7]. This system is 98.7% accurate for recognizing gestures.

III. SYSTEM FUNCTIONALITY

A vision based identification of the static signs of Indian sign language (ISL) alphabets and numerals. The signs considered for recognition includes the alphabets (A-Z) excluding J and H, and numerals (0-9). The system deals with images of bare hands, which allows the user to interact with the system in a natural way and in whatever environment he is comfortable with. A static sign is determined by certain configuration of the hand, while a dynamic sign comprises of one or more moving gesture *i.e.* a sequence of hand movements and configurations. As an initial phase, our algorithm focuses on static signs for the alphabets and numbers. The recognition of the alphabets j and h (dynamic signs) requires additional steps for identification which traces the trajectory of motion and extracts the shape of the trajectory. Fig 1 shows the ISL dataset.

Fig 1: ISL Dataset [14]

The system is designed to visually recognize static signs of the American Sign Language (ASL), all signs of ASL alphabets using bare hands. The user or signers are not required to wear any gloves or to use any devices to interact with the system. But, since different signers vary their hand shape size, and operation habit and so on, which bring more difficulties in recognition. The entire method consists of the following three main stages: Stage A: Image Preprocessing Stage B: Feature extraction using SIFT. Stage C: Gesture Recognition.

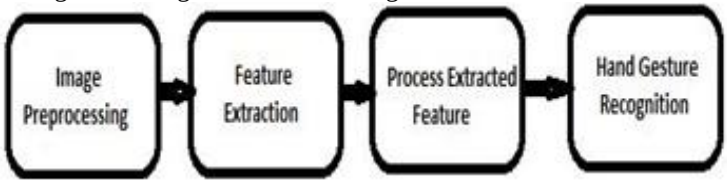

Fig.2 : Basic Block Diagram of Recognition System

1) Detection of scale space extrema

The first stage of key point detection is to identify locations and scales that can be repeatable assigned under differing views of the same object. Detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space[7] (Witkin,1983). Under a variety of Reasonable assumptions the only possible scale-space kernel is the Gaussian function. Therefore, the scale space of an image is defined as a function, L(X,Y,σ) that is produced from the convolution of a variable-scale Gaussian, G(X,Y, σ) with an input image, I(X,Y):

Fig. 3: ASL figure spelling of letter W, A, and V.

$L(X, Y,\sigma) = G(X, Y,\sigma) * I(X,Y)$ (1)

To efficiently detect stable key point locations in cale space, [8] proposed Using scale-space extrema in the difference-of-Gaussian function convolved with the image, D(X,Y, σ) which can be computed from the difference of two nearby scales separated by a Constant multiplicative factor k:

$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)$

$= L(x, y, k, \sigma) - (L(x, y, \sigma)$

2) Locate maxima or minima in DOG image

In this step we coarsely locate the maxima or minima. We go through each pixel and check all its neighbors. The check is done within the current image, and also the one above and below it. In Fig.4 X marks the current pixel and green circles mark the neighbors. This way, a total of 26 checks are made**.** X is marked as a "key point" if it is the greatest or least of all 2 6 neighbors. To increase chances of matching and stability of the algorithm, we also find sub pixel maxima or minima using Taylor series approximation.

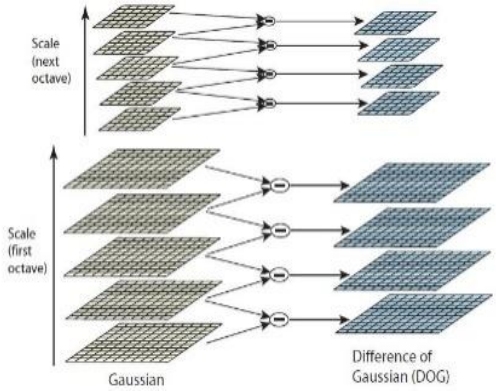

Fig. 3: SIFT Take image to next level we take original image and progressively blur out image. Then, resize the original image to half size and you generate blurred out images again

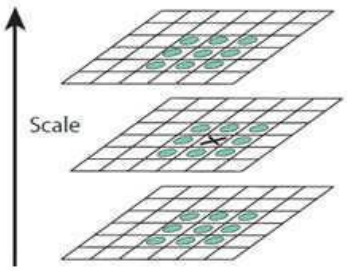

Fig. 4: Key point localization

3) Assigning an orientation to the key point

We already know the scale at which the key point was detected. The next thing is to assign an orientation to each key point. This orientation provides rotation invariance. The idea is to collect gradient directions and magnitudes around each key point. Gradient magnitudes and orientations are calculated using these formulae:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (3)$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \quad (4)$$

The magnitude and orientation is calculated for all pixels around the key point. Then, histogram is created for this.

4) Key point Descriptor

We want to generate a very unique fingerprint for the key point. We also want it to be relatively lenient when it is being compared against other key points. To do this, as shown in Fig.5 we take 16×16 window around the key point. This 16×16 window is broken into sixteen 4×4 windows. Within each 4×4 window, gradient magnitudes and orientations are calculated. These orientations are put into an 8 bin histogram. The amount added to the histogram bin depends on the magnitude of the gradient. This is done using a "Gaussian weighting function". Doing this for all 16 pixels, you would've "compiled" 16 totally random orientations into 8 predetermined bins. You do this for all sixteen 4×4 regions. So you end up with 4x4x8 = 128 numbers. These 128 numbers form the "feature vector". The key point is uniquely identified by this feature vector.

IV. GESTURE RECOGNITION

For Sign image matching, we saved the feature vectors for the training image set. When a New image is applied to the algorithm, preprocessing steps discussed in Section III are first performed. Then, we use SIFT algorithm to calculate the feature vectors for this input image. The minimum Euclidean distance between each feature vector of the query image and all the feature vectors of the database is found. The Gesture image having a feature vector with the minimum Euclidean distance to a feature vector of the query image is given a vote to be the right sign of alphabet. After we go over all the feature vectors of the query image giving votes to alphabet sign in the database, we observed that we always have the right sign of alphabet to be the one with the highest number of votes.

We decided to compare the highest vote (corresponding to the right ASL alphabet) and the second highest vote (corresponding to the most conflicting alphabet). If the difference between them is larger than a threshold, then there is a match and this match corresponds to the highest vote. If the difference is smaller than a threshold, then we declare a 'No Match'.

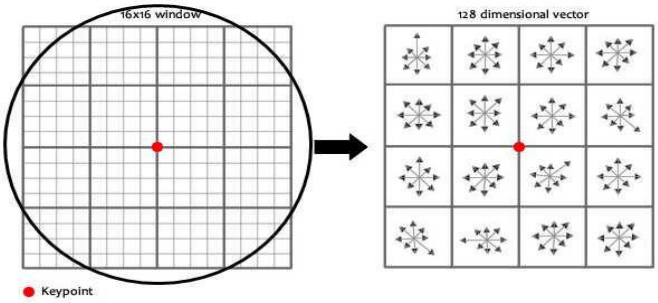

Fig.5 : Key Point Descriptor

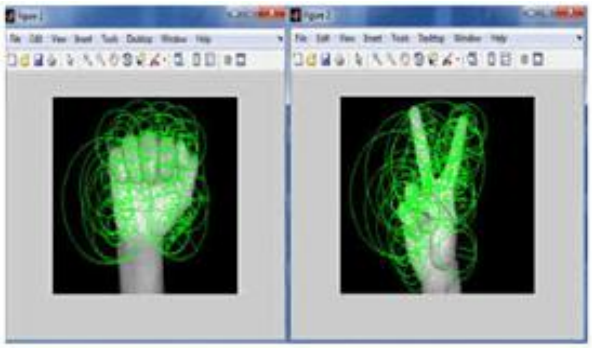

Fig. 6 : Sign A and V, Green circles of different sizes represent scale and we can detect interest point using Scale space of LOG.

V. EXPERIMENTAL RESULT

The hand gesture recognition system, presented in this paper is tested by using hand images from different people with varying shape, scale, rotation and size. The conclusions drawn based on the robustness of the features and Sign recognition. In our Experiment we take total seven gestures A, V, W, 1 etc. of hand of different people. In Fig.6 as we see circle of different sizes and each size correspond to scale we use σ value. This tells us we can detect interest point using scale space of LOG of image.

As shown Fig. 7 input image representing character V. When we calculate the feature descriptor of input image and it matches key points with the scaled image of character V present in the database. As shown in Fig. 8 input image representing character A. When we calculate the feature descriptor of input image and it matches key points with the scaled image of character A present in the database.

We have also seen that SIFT is successfully able to detect similarities between images, even though the image has went through transformation. We see for different images which has scale change, rotation version of image, also possible to recognize successfully.

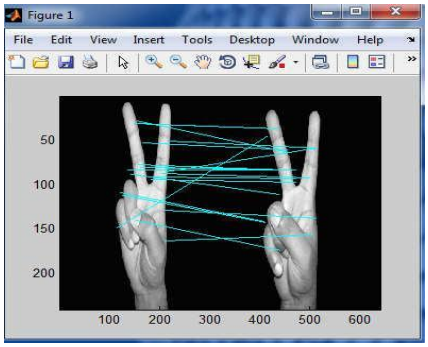

Fig.7 : Matched Database image Vs. Input image of ASL character V

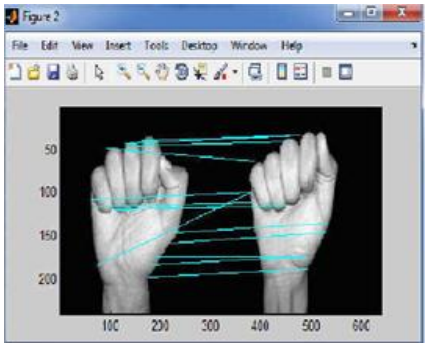

Fig. 8 : Matched Database image Vs. Input image of ASL character A

VI. CONCLUSION AND FUTURE WORK

We use SIFT algorithm for feature vector composition. The SIFT features described in our implementation have been computed at the edges which are invariant to scaling, rotation, addition of noise. These features are useful due to their distinctiveness, which enables the correct match for key points between different hand gestures. This makes the recognition system more practical since signers do not have to wear gloves that make the signing process natural. The system shows that the first stage can be useful for deaf persons or with speech disability for communicating with the rest of the people who do not know the language. As future work, it is planned to add to the system a learning process for dynamic signs. Also if we use Bag of Word approach for classification purpose then matching rate of signs possible to increase.

REFERENCES

[1] David G. Lowe, "Distinctive image features from scale-invariant key-points", International Journal of Computer Vision, Vol.60, No. 2, pp. 91-110, 2004.

[2] TimiOjala, MattiPietikainen and TopiMaenpaa, 2002 "Multi-resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, pp.971-987

[3] Hongo, H., Ohya, M., Yasumoto, M., Yamamoto, K., 2000. Face and hand gesture recognition for human–computer interaction. In: ICPR00: Fifteenth Interna- tional Conference on Pattern Recognition, Barcelona, Spain, pp.2921-2924.

[4] Hongo, H., Ohya, M., Yasumoto, M., Yamamoto, K., 2000. Face and hand gesture recognition for human–computer interaction. In: ICPR00: Fifteenth Interna- tional Conference on Pattern Recognition, Barcelona, Spain, pp.2921-2924.

[5] W. K. Chung, W. Xinyu, and Y. Xu. 2008 "A realtime hand gesture recognition based on Haar wavelet representation", in Proceedings of the IEEE International Conference on Robotics and Biomimetics, Washington, DC, USA, 2008, pp.336-341.

[6] M. P. Paulraj, S. Yaacob, H. Desa, and W. Majid. 2009 "Gesture recognition system for KodTangan Bahasa Melayu (KTBM) using neural network", in 5th International Colloquium on Signal Processing and ItsApplications, pp. 19-22

[7] Witkin, A.P. 1983. Scale-space filtering. In International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, pp. 1019-1022.

[8] Lowe, D.G. 1999. Object recognition from local scale-invariant features. In International Conference on ComputerVision, Corfu, Greece, pp. 1150-1157.

[9] Triesch, J., Von der Malsburg, C., 2001. A system for person-independent hand posture recognition against complex backgrounds. IEEE Transanctions on Pattern Analysis and Machine Intelligence 23 (12), 1449-1453

[10] Kjeldsen, R., Kender, J., 1996. Finding skin in colour images. In: IEEE Second International Conference on Automated Face and Gesture Recognition, Killington, VT, USA, pp.184-188

[11] Manresa, C., Varona, J., Mas, R., Perales, F.J., 2000. Real-time hand tracking and gesture recognition for

human–computer interaction. Electronic Letters on Computer Vision and Image Analysis, pp.1-7.

[12] K. Grobel, M. Assan. 1996 "Isolated sign language recognition using hidden markov models." International Conference on Systems - ICONS

[13] AshwinS.Pol, Dr. S.L.Nalbalwar, Prof. N.S. Jadhav,2013, "Sign Language Recognition Using ScaleInvariant Feature Transform and SVM", International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013

[14] AshwinS.Pol, Dr. S.L.Nalbalwar, Prof. N.S. Jadhav,2013, "Sign Language Recognition System Using SIFT Based Approach", International Conference on Advances in Computer and Information Technology, Goa, 31st March, 2013

[15] Imagawa, I., Matsuo, H., Taniguchi, R., Arita, D., Lu, S., Igi, S. 2000 : Recognition of local features for camera-based sign language recognition system. In: Proc. 15th International Conference on Pattern Recognition. Volume 4. pp.849-853

[16] Divya S , Kiruthika ,S Nivin Anton A L and PadmavathiSSegmentation, 2014, "Tracking And Feature ExtractionFor Indian Sign Language Recognition", International Journal on Computational Sciences & Applications (IJCSA) Vol.4, No.2

[17] ArtiThorat, VarshaSatpute, AratiNehe, TejashriAtreYogesh R Ngargoje,2014,"Indian Sign Language Recognition System for Deaf People", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 3

BIOGRAPHY

Narendra Jadhavwas born in Ahmedabad, India,in 1978. I received the B. Tech. and M. Tech. degree from the Dr. BATU, Lonere-Raigad, in 2000 and 2004 respectively.I am currently pursuing the Ph.D. degree with the Department of ETC Engineering, SGGSIE&T, Nanded-India. My research interests include Biomedical Signal Processing and VLSI.

