IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 3, Ser. I (May - June 2020), PP 14-19
www.iosrjournals.org

Indian Sign Language Character Recognition

Shravani K[1], Sree Lakshmi A[2], Sri GeethikaM[3], Dr.Sapna B Kulkarni[4]
[1]*Computer Science, RYM Engineering College, VTU University, Belagavi, India*
[2]*Computer Science, RYM Engineering College, VTU University, Belagavi, India*
[3]*Computer Science, RYM Engineering College, VTU University, Belagavi, India*
[4]*Computer Science, RYM Engineering College, VTU University, Belagavi, India*

Abstract: *Indian sign language (ISL) is one of the challenging topic as it is in rudimentary stage of its development, unlike American Sign Languages (ASL). This project aims at classification of Indian sign languages using machine learning models. There has been broad research on ASL and adequate data is available to analyse it. As, India is multi diverse country, there are severalregions and cultures which results in different variations of languages for communication. So, there are very limited standard data sets, which has variations and noises. ISL uses both hands to make gestures instead of one hand unlike ASL. It leads to occlusion of featuresand this is a major barrier for the lack of development in this field. This project aims at helping in then research of this field further by providing a data set of ISL. A data of sign language was created by us for alphabets and numeric. Later, the features will be extracted from the collected segmented data using image pre-processing and Bag of words model. Histograms are generated to map the alphabets with images. In the final step, these features will be fed to supervised models for the classification.*

Key Word*:Sign language; Gestures; Bag of words; Image processing; Classification; Support Vector Machine; Clustering; Visual words.*

Date of Submission: 25-04-2020 Date of Acceptance: 08-05-2020

I. Introduction

A gesture is a pattern which may be static, dynamic or both, and is a form of nonverbalcommunication in which bodily motions convey information. Communication is an important aspect when it comes to share or express information, feelings, and it brings people closer to each other with better understanding. When it comes to disabled persons for example deaf and dumb people, it becomes tougher for them to communicate using natural language. So, they use sign language to communicate with themselves and with entire world. But normal people find it difficult to understand sign language as they do not have mostly any prior education or experience in this. Sign language is composed of visual gestures and signs, which are usedby deaf and mute for their talking. It is a well-structured code gesture where every signhas a specific meaning allotted to it. These signs are not only used for alphabets or numeric but also for common expressions also for example greetings and sentences There are 143 existing different sign languages allover the world, mainly American Sign Language (ASL), British Sign Language, French SignLanguage, Japanese Sign Language, and Indian Sign Language (ISL) [8]. Every country has its own language, similarly, sign language is not a universal language and differs from countryto country. There has been a lot work already done on ASLrecognition as it is widely learnt language all over the globe. ASL uses single hand in the gesture representation and it is simple comparing to ISL. ISL uses boththe hands for gesture representation and it is complex comparing to ASL. Because of this reason, there is less research and development in this field.

This project goal is to take the simple step in connecting the social and communication bridge between regular people and the disabled people with the help of Indian Sign Language.A s our project only deals with alphabets and numeric in ISL, it can be extended to common expressions and also words which can be more effective for disabled and normal people in communication and understanding. As we live in a centurywhere India is developing at a rapid pace in terms of digital and technological advances,this project could be one of the steppingstones where technology meets humanity andhelp the hearing impaired and mute community.

II. Related Work

As discussed above, sign languages are a key of communication for disabled people. Karishma Dixit et al [4] presents a technique which perceives theIndian Sign Language (ISL) and converts into a normal text. This process comprises ofthree phases, to be specific a preparation stage, a testing stage and a classification stage.Combinational parameters of Hu invariant second and basic shape descriptors are made toframe another component vector for sign recognition. Classification is done using a a multi-class Support Vector Machine

(MSVM) model.The viability of theproposed technique is approved on a data set having 720 pictures. Trial results exhibitthat the proposed framework can effectively perceive hand signal with 96% accuracy rate.

Subhash Chand Agrawal et al [1] presents the recognition of two-handed Indiansign language (ISL). Their method consists of three stages mainly, Image segmentation, Feature extraction, Recognition. They used Otsu algorithm which is a automatic image thresholding method for segmentation step. In the next step, they used SIFT key points extraction algorithm and HOG descriptors. To classify all gestures, they used multi-class Support Vector Machine(MSVM) model and trained it with all features. Using this method, their model scored 93% accuracy rate in recognising gestures. Our method is also similar to them, but we use canny edge and speeded up robust features(SURF)key points algorithm instead of SIFT in segmentation phase.

The model that has been proposed by Bhumika Gupta [5] perceives static picturesof the markedletters in order in the Indian Sign Language. Dissimilar to the letter setsin other sign languages via gestures like the American Sign Language and the ChineseSign language, the ISL letter uses both the double hands and as well single hand.Hence, itmakes ease the recognition of the gestures by categorizing them into single hand and doublehanded signature. For the two classes two different features have been used, namelyHOG and SIFT which are separated for a set of training images and are consolidated intoa single matrix. After which, the features of HOGand SIFT for the input test images arejoined with the feature matrices of HOG and SIFT.The resultant classification of the testimage is obtained by feeding k-Nearest neighbour classifier with computed correlationfor the matrices.

Figure 1. ISL gestures for numeric and alphabets

III. Implementation

The algorithm implemented in this paper recognizes Indian Sign Language gesturestaken from static pictures. The system comprises of several steps which are Imagecollection, Image pre-processing (segmentation), Feature extraction, Classification asshown in the figure 2. Bag of visual words (BoW) model has been implemented to classifythe images. Idea of BoW is adapted from Natural language processing (NLP). Inimage processing, BoW model concept can be called as “histogram-based representationof independent features”. So, an image can be viewed as a document in order to depictany gesture using the BoW model. Likewise, it is important to describe “words” in imagestoo. To accomplish this, the following three steps are normally included: feature description,and generation of codebooks (visual words). Using these codebooks, histogramscan be generated for all the images. Further, classification of images can be done usingSupport Vector Machine (SVM) model.

Image collection

As there is lack of research in this field, proper data set for ISL is not availablecurrently. So, we created data set of ISL which contains 35 classes, each class with 1200images. Total 35 classes comprise of alphabets (A-Z) and numeric (1-9) with 42000images. All images for each class are captured using webcam video. Each frame of thevideo is saved as an image. To reduce the noise, gestures were captured with a blackbackground. All gestures with classes can be seen in the figure 1.

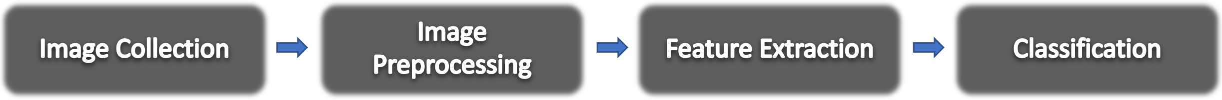
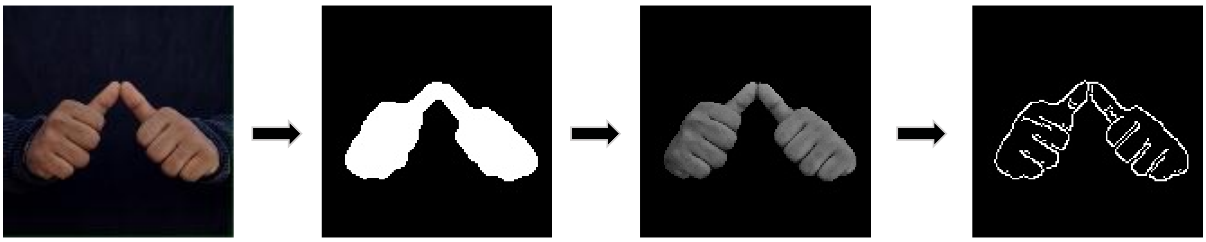


Figure 2. Overview of the system

Image pre-processing

In this phase, all images are pre-processed so that they can be used for featureextraction. This phase contains three steps which are image segmentation (skin masking),skin detection, edge detection. From the raw image (figure 3a) skin mask is generated byconverting the image to HSV colour space. The (H,S,V) range of all pixels from (0,40,30)to (43,255,254) are treated as skin pixels. Using the skin mask, skin can be segmented(see figure 3c). Finally, the Canny Edge technique [3] is used to detect and recognize thepresence of sharp discontinuities in an image, thus detecting the edges of the image (seefigure 3d).



a) Raw image b) Skin mask (gray scale) c) Skin (gray scale) d) Canny edge detection

Figure 3. Image Pre-processing

Feature extraction

Next phase is feature extraction. This phase contains three steps which are imagefeature detection, Clustering, code book generation for Bow model. Initially, forfeature detection we discussed to use Scale-invariant feature transform (SIFT)[7] algorithmto identify key features in the images. But we moved to speeded up robust features(SURF)[2] as SURF is a novel feature extraction algorithm and robust against scaling,rotation, variation and occlusion in viewpoint. SURF features of an image can be seen inthe figure 4.

Next step is to cluster all these SURF features which are similar to make a visualvocabulary. It's unlikely possible to human to find all similar feature descriptors. So, K-Meansclustering algorithm can be useful in this step. K-means algorithm is a clusteringunsupervisedalgorithm that is widely used. It aims at partitioning n features to k numberof clusters and predicts a new feature belonging cluster by mean (centroid) of the clusters.As, we are using large data of SURF features of 42000 images, K-Means clustering takesmuch processing time and memory. So, we used mini batch K-Means which is similar toK-means but has advantages in terms of processing time and memory consumption. Onceall SURF features are trained to mini batch K-means, it clusters all the similar featuresinto a bag and number of clusters (visual words) are equal to k. We used k value 280 asthere are 35 classes to classify. Using this model, visual words for each image can bepredicted.

Next step in this phase to calculate histograms using these predicted visual words.Histogram can be calculated by finding the frequency of occurrence of each visual wordthat belonging to image in total visual words.The whole phase can be seen in the figure5.

Classification

Once all histograms are generated for total data set, the next step is classification.Before going to classification, it is important to divide the data for training and testing.So, we split the total data in 80:20 ratio such that each class has 960 images for trainingand 240 images for testing. Once the data is ready, the next step is to feed the trainingdata to machine learning model. We used Support Vector Machine to train and predict thedata with linear kernel. Similarly, other models can also be used for predictions such asK-Nearest

Neighbours, Convolution Neural Network, Logistic Regression classifier etc. Additionally, real time recognition also developed so that a user can predict the gestures using video feed in real time.

Figure 4. SURF key points and feature descriptors for an image.

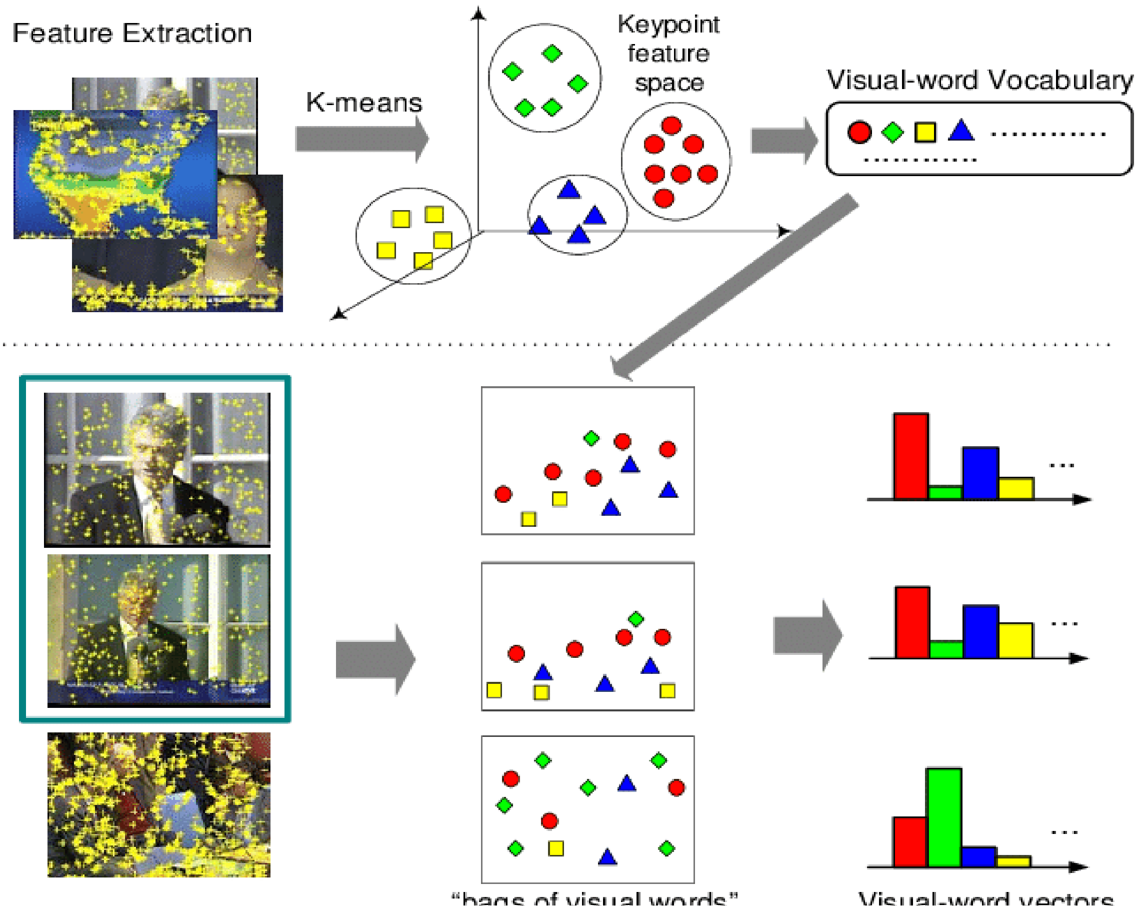


Figure 5. Bag of visual words representation. **Above:** Training for the K-means model using all features (SURF features) **Below:** Prediction of visual words and representing histograms for each image. [6]

IV. Results

Using Bow, integrated with robust SURF feature descriptors, the model scored 99% accuracy. The confusion matrix for the model can be seen in the figure 6. All labels have been predicted correctly by the SVM except label 2. Real time recognition prediction results can be seen in the figure 7. The precision, recall, f1 score was also 99.98%. But there could be slight biasing in the model prediction as data set has much similar images without variations for example in light and skin tone. So, using a large and variety of images in data set, this approach can be more robust for real world applications.

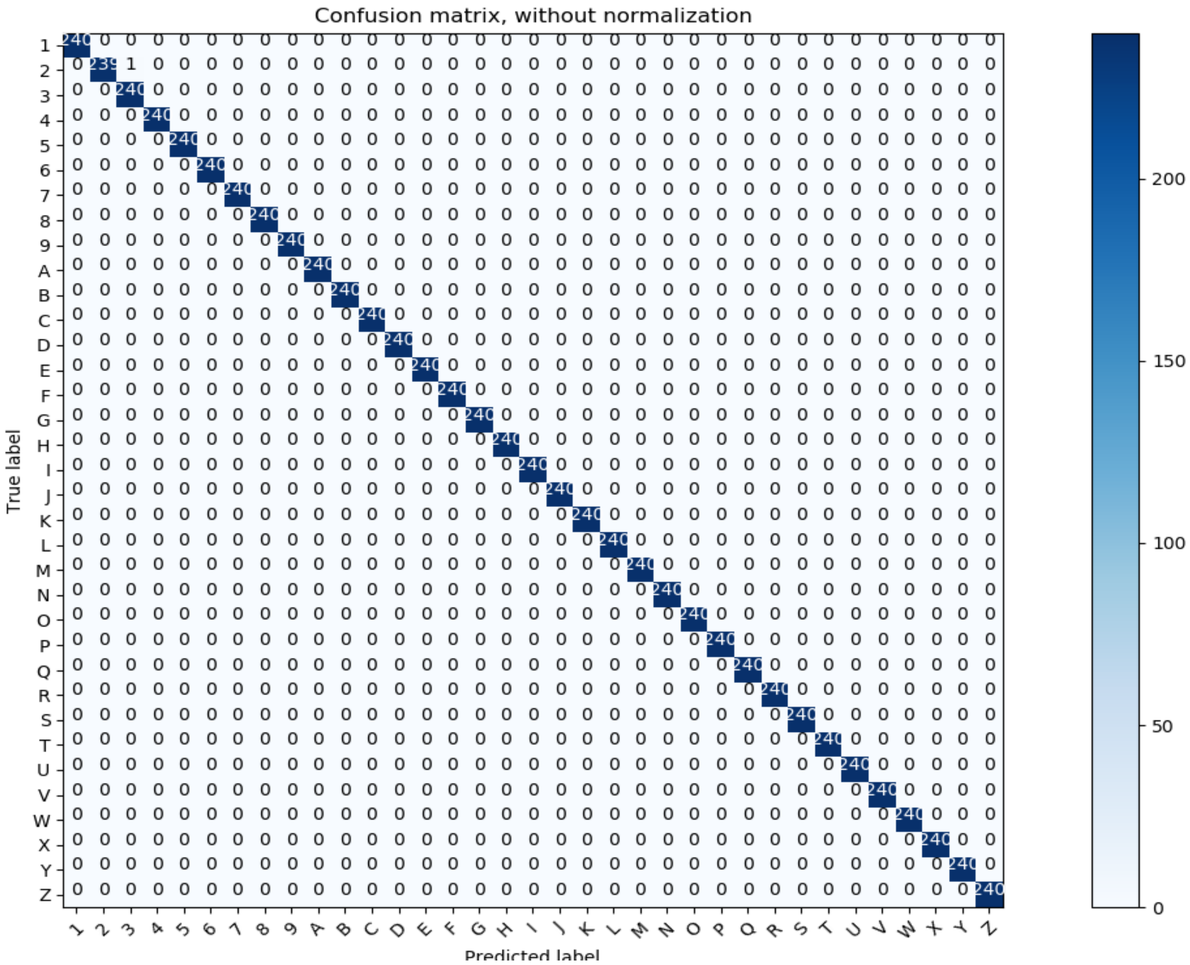
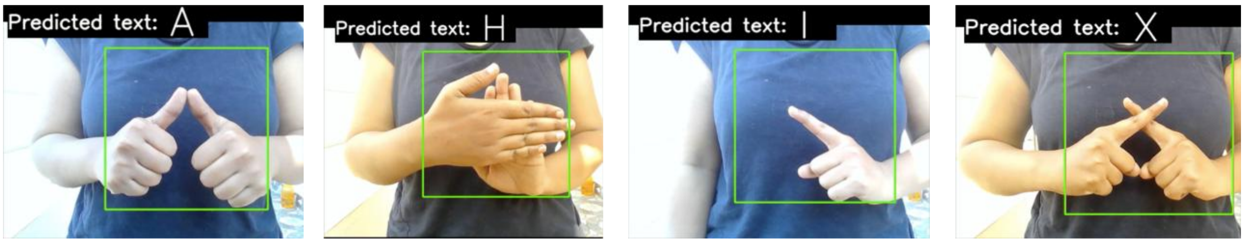


Figure 6. Confusion matrix using SVM model



a) Alphabet A b) Alphabet H c) Alphabet I c) Alphabet X

Figure 7. Real time recognition of gestures results

V. Conclusion

ISL is a key for communication for deaf and dumb people in India. This papergives a detailed implementation for Indian sign language recognition using Bag of wordsmodel.In section 3, step wise implementation has been discussed which are image collection,image pre-processing, feature extraction (using K-means clustering, visual wordscollection) and Classification. Finally, results were presented of Bow model in section 4. Recognition for not only static images, but also real time recognition of gestures also developed. This project can also be extended for simple expressions and words in ISL including alphabets and numeric.

References

[1]. S. C. Agrawal, A. S. Jalal, and C. Bhatnagar. Recognition of indian sign language using feature fusion. In 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pages 1–5, 2012.

[2]. Herbert Bay, TinneTuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In Aleˇ sLeonardis, Horst Bischof, and Axel Pinz, editors, Computer Vision – ECCV 2006, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[3]. J. Canny. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(6):679–698, 1986.
[4]. K. Dixit and A. S. Jalal. Automatic indian sign language recognition system. In 2013 3rd IEEE International Advance Computing Conference (IACC), pages 883–887, 2013.
[5]. B. Gupta, P. Shukla, and A. Mittal. K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. In 2016 International Conference on Computer Communication and Informatics (ICCCI), pages 1–5, 2016.
[6]. Yu-Gang Jiang, Jun Yang, Chong-Wah Ngo, and Alexander Hauptmann. Representations of keypoint-based semantic concept detection: A comprehensive study. Multimedia, IEEE Transactions on, 12:42 – 53, 02 2010.
[7]. David G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60(2):91–110, November 2004.
[8]. Dutta, Kusumika& K, Satheesh& S, Anil & Sunny, Breeze. (2015). Double handed Indian Sign Language to speech and text. 374-377. 10.1109/ICIIP.2015.7414799.

Shravani K,etal. "Indian Sign Language Character Recognition." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22(3), (2020), pp. 14-19.