Bayesian K-Nearest Neighbour based Redundancy Removal and Hand Gesture Recognition in Isolated Indian Sign Language without Materials Support

Rohit Agarwal
Department of Computer Engineering and Applications
GLA University, Mathura, India-281406
rohit.agrwal@gla.ac.in

Abstract—Indian sign language is used by deaf and dumb persons for communication. Physical materials are not needed for ISL. During this anticipated paper we tend to propose a unique vision-based (VB) methodology for the Indian language (ISL) recognition of isolated signs. The methodology proposed consists of 3 modules: preprocessing, extraction and classification of features. One signer hand is segmented from sign video in segmentation or preprocessing phase. Feature vector is extracted from feature extraction module and show the manual sign parameters used as input for classification. To decrease the computational complexity of the system, a redundant frame algorithmic rule has been applied. The experiment results demonstrate that proposed system achieves 90.3% recognition accuracy based on a lexicon of twenty-one signs of ISL.

Keywords: Skin color segmentation, Indian Sign language, Materials, Redundant frames removal, Bayesian KNN.

I. INTRODUCTION AND RELATED WORK

Sign language is the most expressive and natural way which is used by hearing and visually impaired people for communication. This can involve different hand shapes, movement and orientation of hands, arms, body and facial expression to communicate. Sign language has been developed for the deaf communities. Signing is also done by hearing impaired person who cannot speak. The language used in India is called Indian Sign Language (ISL). Sign language is drastically different from spoken language. In Sign language, a sign is composed of cheremes while in spoken language a word is composed of phonemes. Spoken language is the sequence of sound patterns whereas sign language can be sequential but also parallel due to high dimensionality: hand shape/orientation, hand location, facial expression and mouth/head movement [1].

Sign language is consisting of two types of parameters: manual and non-manual. Manual parameters use hand gestures with different hand shapes and orientation to deliver the message. Non-manual parameters consist of movement of the body, head, cheeks, mouth, and eyes which are used simultaneously to convey the speaker’s thought. Research in sign language and hand gestures mainly focused on two dimensions: isolated sign and continuous sign. In isolated sign there is only one sign that user performs while in continuous sign more than one signs one after the other. Continuous signing suffers from the problem of co-articulation i.e. a sign may be affected by preceding or succeeding sign [2].

For gesture data capturing in sign language recognition (SLR) there are two types of approaches [3] used: instrumented glove and vision based. Instrumented glove-based systems capture the data of hands through motion sensors which restrict the movement of signer. Vision based device captures hand movement
through video camera. The recorded video is then analyzed by artificial intelligence or image processing.

There are some problems related to sign language recognition. The signing speed can differ greatly, even if the same person performs the same sign twice, minor changes in hand position and velocity may happen. In addition, there are some difficult issues such as hand tracking or hand segmentation from context and setting, difference in lighting, occlusion, and location etc. [4]. Most of the works in isolated sign recognition has simplified the problem of segmentation and tracking by wearing devices on the hands, such as color gloves or markers to directly measure the location features.

The sign is performed in three phases [1]: Preparation, stroke and retraction. The movement of the hand toward intended location before the actual stroke of a sign (the preparation), desired movement of hands in performing in the sign (stroke), after finishing the sign or hand in the relax position (retraction). Stroke phase is most important phase and contain a lot of either irrelevant or redundant frames.

A classification approach was proposed in [12], in which, authors used Zernike moments to identify the orientation of the hand. In state-of-the-art work on isolated sign recognition, Aran et al. [5], Sign Tutor stages consist of a face and hand detector stage, analysis stage and the final sign classification. Signer wore colored gloves on hands to simplify the problems of hand detection, segmentation and occlusion. The most important part of the system is the Analysis and classification subsystem which tracks the hand using Kalman filter, extract various features and classifies the sign using Hidden Markov Model with a recognition rate 94.2% and 79.61%. Sandjaja and Macros [6] also used color-coded gloves to make human hand tracking easier, systems extract important features from the video using multi-color tracking algorithm and Hidden Markov Model for the classification. Multi-function extraction proposed by Quan [7] used hand which is the only considered object. These features include color histogram, Hu moments, Gabor wavelet, Fourier descriptor and SIFT features and classification support vector machine used. Nandy [4] proposed a method for the recognition of static isolated sign with a direction histogram and classification is performed using Euclidean-Distance and K-nearest neighbour with a 90% accuracy.

In [13] ISL recognition is proposed by Optimized Neural Networks. The state-of-the-art system [14] uses Open Pose library, which assistances in making the skeleton of human physique and thus it delivers important points of the whole human body frame by frame.

Most of researcher had used the dataset of static signs where hand is the only object to simplify segmentation task. We proposed a vision-based approach for ISL recognition of isolated sign considering the sign's static and dynamic behavior. A redundancy removal algorithm is used to reduce the system's computational complexity. We have implemented to delete obsolete or redundant frames from the sign picture.

An automation system to recognize sign language in complex backgrounds with video of the signer and hand showing the sign is segmented and presented in [15]. Indian sign recognition area based on dynamic hand gesture recognition techniques in real-time situation is presented in [16]. Grid-based features [17] are used to
recognize hand poses & gestures from the Indian Sign Language (ISL) in real-time. Vision based ISL is proposed by Tyagi et al. [18].

This paper is organized as follow: Section II describes the proposed method. A narrative of the tests assumed and their results over 21 isolated sign vocabulary from the Indian Sign Language is detailed in Section III. The last section grants conclusion and future work.

II. PROPOSED APPROACH

The proposed block framework is depicted in Figure 1. It contains mainly of three components, preprocessing module, extraction of features and a module for classification. Hand region from sign video frames is detected in the preprocessing module. It is consisting of skin color segmentation, redundant frames removal and face elimination. Specific features such as hand shape, hand motion, and hand orientation are extracted in feature extraction module. Various sign groups are used in the classification module for evaluating a gestured input picture. The output with the highest probability is identified for recognizing the gesture and showing its significance.

A. Preprocessing

1) Segmentation and Noise Removal

Before applying segmentation, frames are extracted from the video of the sign. Skin color segmentation is used to detect human skin pixels from frames. In Figure 2(b), face and hand area from every frame of the video is segmented through skin color segmentation algorithm.

2) Redundant Frames Removal

A number of meaningless or redundant frames are in sign language video. We have applied an algorithm to remove those irrelevant frames. The steps of the approach are outlined below. We used the threshold A= 350 pixels for all the
experiments. We set this threshold based on the frame size of the video. Relevant frames are obtained by and duplicate frames will have reduced when following steps for each frame is repeated. In relevant frame pixels are moving. From Figure 2(c), it is clear that only six frames are taken. These six frames out of 100 are used for feature extraction.

### Redundant Frames Removal Algorithm

For each sign video with \( N \) frames:

1. Initialize first frame \( f_1 = 1 \), and counter \( m = 1 \).
2. For independently frame \( i = 2, \ldots, N \):
   
   a) Calculate variance image among frame \( S(i) \) and \( S(f_m) \).
   
   b) Extract no. of pixels \( (N_p) \) from the largest connected component of the image.
   
   c) If \( N_p > A \) (threshold), set \( m = m + 1 \), set \( f_m = i \).
   
   d) Set \( i = i + 1 \)

3) **Face Removal and Noise Elimination**

Removal of face area of the signer is the primary objective of this process. It is done by removing duplicate frames form skin detected images. We remove the face area from other frames, when first segmented frame is used as input. This progression is required in light of the fact that face and hand zones are skin shading zones while we require just the hand zone from the frame successions for extricating direction of the motion. Largest connected component (L) is detected from the first skin image and finally calculates the difference image between L and different frames in the sequence.

In this way, we get only the hand area in other frames. The difference image contains unexpected small skin regions. In order to eliminate these small skin regions (noises), erosion operation with structuring element \([1 1 1 1]\) has been performed. Figure 2(c) shows the outcome of this stage.
B. Selection of Features

System’s classification performance is depending upon selection of suitable features i.e. Hand orientation, hand motion and hand shape.

1) Hand Shape

The main components of the sign language are the hand shape. Analysis of hand shape is complicated task for the system since there are number of hand shapes in sign language. For the same, we have extracted three features: Eccentricity, Compactness and Solidity.

a) Eccentricity

It is defined by the proportion of the separation between the foci of the oval and the length of its significant hub. Its worth is somewhere in the range of 0 (Circle) and 1 (Line).

b) Compactness

It shows the picture closeness to its outline and its centre. Its greatest worth is 1 for circles and its worth declines for circular shapes [9]. It is characterized as:
compactness = \frac{4 \pi \text{Area}_{\text{image}}}{\text{Perimeter}^2_{\text{image}}} \tag{1}

\text{c) Solidity}

It tends to be determined as the proportion of the picture item's region to the region of the article's curved structure. It quantifies the thickness of an item [9].

\text{Solidity} = \frac{\text{Area}_{\text{image}}}{\text{Area}_{\text{convex hull}}} \tag{2}

2) Hand Motion

Hand motion plays an important role in developing the sign. Hand curve may change the sign. The signs which are static do not need examination of hand trajectory. The combination of the centroids \((x_{ci}, y_{ci})\) forms the raw trajectory of the gesture path.

\[ p = \{(x_{c1}, y_{c1}), (x_{c2}, y_{c2}), \ldots, (x_{cN}, y_{cN})\} \tag{3} \]

Where \(NF\) is the number of frames. Each motion in the gesture based communication is performed at various speeds. The speed [10] is determined between the two progressive centroid focuses partitioned when as far as the quantity of video outlines (\(N\)) as follows:

\[ V = \frac{\sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2}}{N} \tag{4} \]

3) Hand orientation

Direction is one of the most significant highlights which are utilized to prepare classifier and perceive sign. The orientation change can be determined as follows:

\[ \Theta = \arctan \left( \frac{y_c + 1 - y_c}{x_c + 1 - x_c} \right) \tag{5} \]

Signs may have different number of frames due to inconsistent hand movement or variable length of the sign. Our data set has minimum number of frames in a sign 15 (for static gestures) and maximum number of frames 35 (for dynamic gestures) after removing redundant frames. Therefore, we have decided to select 30 frames of each sign for feature extraction. For the minimum number of sign video, rest of values in a feature vector considered as 0. For each frame, above mentioned feature has been extracted and stored in a feature vector. Fixing of number of frames is needed because when we train Bayesian KNN classifier, all of the training sample must have the same length or number of frames.
C. Classification

In the proposed work, Bayesian KNN classifier has been utilized to characterize the hand motions among numerous classes. In learning phase, feature vector of all kinds of gestures is used for classification. Based on their properties of hand sign, it is classified in appropriate class.

In pattern recognition and regression K-nearest neighbour (KNN) algorithm is used for classifying the objects. In KNN, an unlabeled or test data is classified from one of N distinct classes. This arrangement is based on a training data set of \( \{ y_i, x_i \} \), for \( i = 1 \) to \( n \), where \( y_i \in \{ 1, \ldots, N \} \) denotes the class names and \( x_i \) denotes a set of feature space. The distance metric \( \rho \) \( (x_{n+1}, x_i) \) is used to determine nearest which is usually Euclidean distance. However, k-nearest neighbour algorithm used for classification has some drawbacks. The selection of \( k \) is a difficult task. The k-nearest neighbour method does not consider probabilistic interpretation in the classification which marks it problematic to incorporate in a consistent decision process [11].

To overcome these difficulties, Holmes and Adams [11] proposed a probabilistic structure for the KNN algorithm that considers uncertainty in \( k \) as well as interaction between neighbours. This method converts predictions on test set into probabilities in the range of [0 1]. Prediction probability \( P \) is \( [n \times q] \) matrix where \( n \) is the number of test points and \( q \) is the number of classes. \( P \) \( (i, j) \) gives the probability that the \( i^{th} \) test point belongs to class \( j \). The class which has the highest probability is considered for final result. They actualize the strategy as a block sequential algorithm where it is accepted that squares of information show up after some time which prompts the prescient appropriation for test information having the type of a probabilistic closest neighbor (PNN) earlier. The strategy doesn't require suspicions about the dissemination of the component vector or indicator factors.

III. Experiment Results and Analysis

The proposed structure has been verified on numerous video sequences having both static and dynamic one handed signs. Video of 21 signs from Indian Sign Language are collected from a video camera. The dataset is separated into two sets of training and testing. We used four videos of each sign in classification for training, and the rest used for testing. The proposed system is checked and trained on a single signer. The video which is used for training is not used for testing purpose. The accuracy of the proposed method is considered as follows:

\[
\text{Accuracy} = \left( \frac{\text{True classified gestures}}{\text{Total number of gestures}} \right)
\]  

(6)

Table 1 illustrates the recognition rate of each sign. The average recognition rate of 90.3% is achieved using the proposed model. The results show that proposed method of sign language recognition is effective and has high accuracy.
Snapshots of the system with three buttons are depicted in Figure 3 i.e. Train Bayesian KNN classifier, Test Video, Exit and Meaning of Sign. ‘Train BKNN’, is used for training with sample gestures. Through ‘Test Video button’, test video is entered. When we press ‘Meaning of sign’, then meaning of sign is shown in respective textbox. In figure 3(a) and (b), the system is producing correct result but figure 3(c) is giving incorrect results as original meaning of sign is “Afternoon”.

| S.No. | Sign          | No. of Signs | Correctly classified : No. of Signs | Accuracy (%) |
|-------|---------------|--------------|------------------------------------|--------------|
| 1     | Afternoon     | 15           | 13                                 | 86.7         |
| 2     | Back          | 12           | 12                                 | 100.0        |
| 3     | Bad           | 9            | 8                                  | 88.9         |
| 4     | Chest         | 11           | 9                                  | 81.8         |
| 5     | East          | 17           | 14                                 | 82.4         |
| 6     | Good          | 12           | 12                                 | 100.0        |
| 7     | Heart         | 15           | 13                                 | 86.7         |
| 8     | How           | 12           | 10                                 | 83.3         |
| 9     | How many      | 11           | 10                                 | 90.9         |
| 10    | How Much      | 10           | 9                                  | 90.0         |
| 11    | North         | 15           | 13                                 | 86.7         |
| 12    | Police        | 11           | 10                                 | 90.9         |
| 13    | Rat           | 11           | 11                                 | 100.0        |
| 14    | Right         | 14           | 14                                 | 100.0        |
| 15    | Sad           | 9            | 8                                  | 88.9         |
| 16    | Sorry         | 9            | 9                                  | 100.0        |
| 17    | Trouble       | 16           | 15                                 | 93.8         |
| 18    | West          | 11           | 10                                 | 90.9         |
| 19    | Whose         | 16           | 14                                 | 87.5         |
| 20    | Wrong         | 13           | 10                                 | 76.9         |
| 21    | Yesterday     | 10           | 10                                 | 100.0        |
| **Total** | **259**       | **234**      |                                    | **90.3**     |
IV. CONCLUSION AND FUTURE SCOPE

A novel VB sign language recognition system is proposed by using skin color segmentation, redundant frames removal, multiple features extraction and Bayesian KNN classifier. Evacuation of comparable or superfluous frames from video accelerates the framework and numerous highlights, for example, hand shape, hand direction and hand movement improve the exhibition of the framework. We have achieved up to 100% recognition rate. When two signs are similar for example “sad” and “heart”, system get confused and sometime gives the wrong meaning. Our upcoming effort will concentrate on the following
problems: (a) Incorporating non-manual parameters; (b) incorporating two handed signs to the sign dataset; (c) merge of classifiers for better recognition levels.

REFERENCES

[1] G.A. Ten Holt, A. J. Van Doorn, M. J. T. Reinders, E. A. Hendriks, and H. De Ridder, “Human-Inspired Search for Redundancy in Automatic Sign Language Recognition,” ACM Trans. Applied Perception, vol. 8, Article 15, 15 pages, 2011.

[2] P. Dreuw, D. Stein, T. Deselaers, D. Rybach, M. Zahedi, J. Burgeroth, and H. Ney, “Spoken Language Processing Techniques for Sign Language Recognition and Translation,” Technology and Disability, vol. 20, pp. 121-133, 2008.

[3] H.D. Yang, S. Sclaroff, and S.W. Lee, “Sign Language Spotting with a Threshold Model Based on Conditional Random Fields,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, pp. 1264-1277, 2009.

[4] A. Nandy, J. S. Prasad, S. Mondal, P. Chakraborthy, and G.C. Nandi, “Recognition of Isolated Indian Sign Language Gesture in Real Time,” BAIP, Springer LNCS-CCIS, Vol. 70, pp. 102-107, 2010.

[5] O. Aran, I. Ari, L. Akarun, B. Sankur, A. Benoit, A. Caplier, P. Campr, A. H. Carrillo, and F. Xavier Fanard, “SignTutor: An Interactive System for Sign Language Tutoring,” IEEE feature article, pp. 81-93, 2009.

[6] I. N. Sandjaja, and N. Marcos, “Sign language number recognition,” Fifth International Joint Conference on INC, IMS,IEEE, pp. 1503-1508, 2009.

[7] Y. Quan, “Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling,” 5th IEEE Conference on Industrial Electronics and Applications, pp. 1537-1542, 2010.

[8] D. Chai, and K. N. Ngan, “Segmentation Using Skin-Color Map in Videophone Applications,” IEEE Trans. Circuits and Systems for Video Technology, vol.9, pp. 551-564, 1999.

[9] N. Jamil, Z. A. Bakar, and T. M. T. Sembok, “Image Retrieval of Songket Motifs using Simple Shape Descriptors.,” IEEE Proceedings of the Geometric Modeling and Imaging-New Trends (GMAI’06), pp. 171-176, 2006.

[10] A. A. Hamadi, M. Elmezain, and B. Michaelis, “Hand Gesture Recognition Based on Combined Features Extraction,” International Journal of Information and Mathematical Science vol.6, 2010.

[11] C. C. Holmes and N. M. Adams, “A probabilistic nearest neighbour method for statistical pattern recognition,” J Royal Statistical Soc B, vol. 64, pp. 295–306, 2002.

[12] Kumar, Anup, Karun Thankachan, and Mevin M. Dominic. “Sign language recognition." In 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 422-428. IEEE, 2016.

[13] Hore, Sirshendu, Sankhadeep Chatterjee, V. Santhi, Nilanjan Dey, Amira S. Ashour, Valentina Emilia Balas, and Fajian Shi. "Indian sign language recognition using optimized neural networks." In Information Technology and Intelligent Transportation Systems, pp. 553-563. Springer, Cham, 2017.

[14] Malhotra, Pooja, Chirag K. Mami, Nikhil V. Sankpal, and Hardik R. Thakkar. "Indian Sign Language Recognition System using Openpose." i-Manager's Journal on Computer Science 7, no. 2 (2019): 43.

[15] Reshna, S., and M. Jayaraju. "Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM." In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 386-390. IEEE, 2017.

[16] Rahja, J. L., A. Mishra, and Ankit Chaudhary. "Indian sign language recognition using SVM." Pattern Recognition and Image Analysis 26, no. 2 (2016): 434-441.

[17] Shenoy, Kartik, Tejas Dastane, Varun Rao, and Devendra Vyavaharkar. "Real-time Indian sign language (ISL) recognition." In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-9. IEEE, 2018.

[18] Tyagi, Akansha, and Sandhya Bansal. "Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System: A Review." Computational Methods and Data Engineering (2020): 39-53.