Journal of King Saud University – Computer and Information Sciences xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario

P.K. Athira, C.J. Sruthi *, A. Lijiya

Department of Computer Science and Engineering, National Institute of Technology Calicut, Kozhikode, India

ARTICLE INFO

Article history:
Received 19 November 2018
Revised 16 March 2019
Accepted 5 May 2019
Available online xxxx

Keywords:
Support vector machines
Zernike moments
Computer vision
Gesture recognition
Feature extraction
Co-articulation elimination

ABSTRACT

Due to the high population of hearing impaired and vocal disabled people in India, a sign language interpretation system is becoming highly important for minimizing their isolation in society. This paper proposes a signer independent novel vision-based gesture recognition system which is capable of recognizing single handed static and dynamic gestures, double-handed static gestures and finger spelling words of Indian Sign Language (ISL) from live video. The use of Zernike moments for key frame extraction reduces the computation speed to a large extent. It also proposes an improved method for co-articulation elimination in fingerspelling alphabets. The gesture recognition module comprises mainly three steps – Preprocessing, Feature Extraction, and Classification. In the preprocessing phase, the signs are extracted from a real-time video using skin color segmentation. An appropriate feature vector is extracted from the gesture sequence after co-articulation elimination phase. The obtained features are then used for classification using Support Vector Machine(SVM). The system successfully recognized finger spelling alphabets with 91% accuracy and single-handed dynamic words with 89% accuracy. The experimental results show that the system has a better recognition rate compared to some of the existing methods.

1. Introduction

A sign language converts letters, numbers, words, and sentences of a natural language to hand gestures facilitating vocal disabled people to communicate with the outside world. It uses hand gestures and facial expression instead of sound to convey messages. Sign language varies from country to country as well as region to region. Indian Sign Language abbreviated as ISL is the sign language used by Indian deaf and dump community. In India, it is estimated that above 2 million people are deaf and among them, one million deaf adults and more than half million deaf children use ISL (Neha et al., 2014).

Deafness and vocal disability bring them significant communication problems while accessing education, job, etc. Since the majority of normal people can't understand their language, communication using sign language is always limited in the deaf-dump community. As humans, they deserve to get all the help needed to live an ordinary life. One way to help them is by using advanced technology to overcome some of the difficulties they face. Sign language using hand gestures are helpful for establishing human-machine interaction, which can facilitate the communication between normal people and hard of hearing people with the machine as a mediator.

The objective of the proposed method is to recognize dynamic and finger spelling (Li and Greenspan, 2007) words of Indian sign language, using a vision-based approach. The signs captured under uniform background is considered for training set construction to avoid other skin colored objects in the background and to make the system more accurate. Initially, the co-articulation phase (Bhuyan et al., 2006Bhuyan et al., 2005) is detected and eliminated from the gesture sequence using an acceleration feature. The selected words and finger spelling alphabets do not use face cues so face region is detected and eliminated using Viola-Jones algorithm (Kumar et al., 2016). Then skin color based segmentation (Bhuyan et al., 2014) is done on the resultant image for hand region extraction. The input to the system can be either static gestures or dynamic gestures. Based on the type of gestures certain features are extracted. The obtained features are then classified using

* Corresponding author.
E-mail address: sruthi_p170043cs@nitc.ac.in (C.J. Sruthi).

Fully documented templates are available in the elsarticle package on CTAN.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2019.05.002
1319-1578/

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

multi-class SVM. The major contributions of this paper are as follows:

- A real-time signer independent and cost-effective approach for Indian sign language recognition.
- Single-handed static and dynamic gestures and double handed static gestures are identified successfully.
- The proposed system recognizes ISL gestures from mobile camera videos without any additional sensors to detect hand region.
- A method for co-articulation elimination in fingerspelling alphabets.

The following section shows a detailed report regarding existing works in gesture recognition or sign language recognition. This is followed by the proposed method and various phases of gesture recognition. Usefulness and viability of the proposed method are explained in the experimental result phase followed by a comparison of the proposed method with already existing methods.

2. Related Work

In gesture recognition system various technologies are used for capturing hand gestures, which includes a glove-based method and vision based method. In data glove based method (Wang and Popović, 2009) the sensors in the glove can detect the movement of hands and pass this information to the computer. This approach has high accuracy in gesture recognition but it is quite expensive and inconvenient to the user. An instrument glove based approach is proposed in Deora and Bajaj (2012) which can recognize the alphabets and numbers of ISL. In this method, the signer needs to wear a red and blue colored glove while data acquisition for easy hand segmentation. And the gesture recognition is done using PCA. The result of recognition is 94%. But the system focus only on static gestures and fails to recognize the signs where both hands overlap.

Vision based methods provide more user convenience. There are two different approaches to the vision-based method. Appearance-based approach and 3-D hand model-based approach. 3D hand model-based method (Cheng et al., 2016; Prisacariu and Reid, 2011) make use of 3D information of the body parts. Using this information several key parameters like palm position, joint angles, etc. can be obtained. This method requires huge storage space to deal with a large number of features. It will give better accuracy at higher computation rate. Various methods suggest depth cameras for data acquisition (Suarez and Murphy, 2012; Kapuscinski et al., 2015). Depth images can be captured using Kinect, ASUS Xtion, etc. Most of the works choose Kinect for data acquisition (Kapuscinski et al., 2015; Dong et al., 2015). A depth-based sign language recognition using SVM is proposed in Kim et al. (2015). In this method finger length, palm radius and hand direction are considered as features. And to recognize the hand gesture a decision tree of these features are constructed.

Due to the limitations of glove-based, most of the research works focus on appearance-based approach with Discrete wavelet transform and Hidden Markov Model (Tripathi et al., 2015), artificial neural network (Adithya et al., 2013; Sharma et al., 2014), fingertip based gesture recognition (Kumar et al., 2016Hussain et al., 2014), scale invariant feature transform (Patil and Sinha, 2017), Zernike moments (Kalpana Sharma and Dutta, 2014Otiniano-Rodriguez et al., 2012), Fourier Descriptor (Nanivadekar and Kulkarni, 2014), B-spline (Geetha and Manjusha, 2012Geetha and Aswathi, 2013) etc. The main advantage of the appearance-based approach is low processing time and it has real-time performance because of the usage of 2-D image features.

Vision based methods involve various image processing and pattern classification techniques for sign language recognition. Lilha and Shivmurthy (2011) proposed a method by recognizing static as well as dynamic hand gestures of ISL. This method used Histogram of Orientation Gradient (HOG) and Histogram of Edge Frequency (HOEF) for extracting features and SVM for classification. The dynamic gestures H and J are recognized by considering either the initial or final frame of the gesture. They have achieved 98.1% accuracy using the HOEF feature. But the signer needs to wear a wristband to distinguish palm and forearm.

A vision based ISL character recognition method was proposed in Ashok Kumar Sahoo (2014). The system is designed to recognize only isolated signs. It can also be used in public places where different backgrounds are present. Structural features, local histogram features and direct pixel values of grayscaled images are extracted from gestural images for sign recognition. After extracting features kNN and neural network classifier are used to classify the hand gestures. They claim to have achieved 95% recognition rate on the single-handed dataset and 96% on the double-handed data set. A similar method is proposed in Sharma et al. (2014) using centroid techniques and direct pixel value to extract features.

Dixit and Jalal (2013) proposed an approach to recognize double handed ISL gestures. Hu invariant moment and structural shape descriptors are the features used. And Multiclass SVM (MSVM) is used for training and recognizing ISL gestures. Another paper (Otiniano-Rodriguez et al., 2012) is based on the comparison of Hu and Zernike moments of hand gestures. The Zernike and Hu Moments of the binary image are calculated and gesture recognition is done using SVM classifier. The method focused on static alphabets of ASL and obtained 96% accuracy. The paper concluded with the remarks that Zernike moments show high accuracy than Hu moments. In Kalpana Sharma and Dutta (2014) also analyzed the exact behavior of complex Zernike moments in the study of the capability of Zernike moments in ISL alphabets recognition. The authors claimed that Zernike moments up to order 10 is enough to recognize and reconstruct images.

A method for speech and gesture recognition of ASL was proposed by Kumar et al. (2016). In the proposed method input is a real-time video. Then the skin color segmentation is done using HSV color space, feature extraction of static gestures is done with Zernike moments and curve features for dynamic gestures. The classification phase used a multi-class SVM. The system makes use of a standard module called Sphnix for speech recognition. The authors claimed accuracy of 93% on static gestures and 100% accuracy on dynamic hand gestures. Imran et.al. also proposed a similar method on ASL for bare hand posture recognition followed by accurate palm and fingertip estimation based on the hand contour (Chen, 2012; Chen, 2008). This method gives satisfactory recognition rate. The main limitation of this method is that it focuses only on ASL alphabets recognition.

Bhuyan et al. (2008) proposed certain features to identify the gesture trajectory of dynamic gestures. The gesture recognition using these set of features result in 95% accuracy. Another method for continuous gestures of ISL is proposed in Tripathi et al. (2015). In this method, the continuous gestures are separated using the gradient method. It calculates the gradient of each frame and checks for overlapping between continuous frames. A more efficient gesture segmentation is done on (Li and Greenspan, 2007) using continuous dynamic programming. They have got an accuracy of 95%.

A novel method is proposed for recognizing unwanted gestural motions between meaningful gesturing in trajectory-based gesture recognition in Bhuyan et al. (2006). The key idea behind the co-articulation detection is that during co-articulation phase the hand moves quickly and in gestural phase, the hand motion is smooth and slow. The method uses Finite State Machine for recognition

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

of gestures with a finite number of keyframes and keyframe duration. The classification is done by matching the incoming gestures with states of FSM. The co-articulation is detected by considering the variation of acceleration in motion.

Recognition of single-handed dynamic gesture recognition is considered in Jalal (2015). In the preprocessing phase the skin color based detection is performed using YCbCr color space and then face region is eliminated for hand segmentation. The proposed method used circularity, extend convex deficiency, hand orientation, and motion detection code as features and multi-class SVM used for classification. They have got 90.4% recognition rate. Subhash et al. (2014) proposed a similar method but the trajectory of the sign is considered instead of motion detection code and got recognition rate 95.3%.

A vision-based approach is introduced by Zaki et al. for American sign language (ASL) recognition. In this work the kurtosis position, PCA, and motion chain code(MCC) are used as features for describing articulation point, hand shape, orientation and movement respectively (Zaki and Shaheen, 2011). Skin color thresholding followed by connected component identification is used for face hand detection and tracking(skin-blob tracking). Final recognition step is done using HMM classifier and an overall recognition error rate of 10.9% is achieved over RWTH-BOSTON-104 database. Ibrahim et al. proposed an automatic Arabic sign language recognition system which uses skin-blob tracking technique for hand segmentation and tracking (Ibrahim et al., 2018). This technique uses geometric features like the center of gravity of hand, motion velocity and orientation as a feature vector and Euclidean distance for recognition. This method was evaluated on a dataset of 30 isolated words and obtained a recognition rate of 97%.

A segment and merge approach for continuous video recognition of American sign language was proposed by Kong et. al. In this work, data acquisition was done using a cyber glove and magnetic tracker. Continuous sign video was segmented based on velocity and direction angle into smaller sub-units. These subunits were then labeled SIGN or ME(movement epenthesis) by a Bayesian network. Subunits labeled ME are discarded after storing their location details. Remaining SIGN labeled subunits are then fed into two layer multi conditional random field (CRF). The first layer in CRF is a combination of four linear CRF with independent handshape, movement, orientation and location classifiers. The second layer combines the output of the previous layer and inputs this to semi Markov CRF for sign sentence recognition. a precision rate of 86.6% and a recall rate of 89.8% was achieved (Kong and Ranganath, 2014). Elakkiya et.al. proposed a new subunit sign modeling approach which addresses ambiguities during segmentation of hand and identifying epenthesis movement in long video sequences. This work uses enhanced dynamic programming with dynamic time wrapping and clustering methods using spatial and temporal features. Here subunit modeling of the long video sequence and epenthesis movement identification was done using subunit multi-stream parallel HMM (SMP-HMM) and minimum entropy clustering. They claim an accuracy of 98% and an average recognition time of 1.25 s (Elakkiya and Selvamani, 2018).

Kharate et. al. proposed a comparative analysis over three classifiers and feature descriptors like Fourier descriptor, 7 Hu moments, shape matrix, and chain code. Three classifiers analysed here for static alphabet and numeral recognition are Nearest Mean Classifier, K-Nearest Neighbour Classifier and Naive Baye Classifier (Kharate and Ghotkar, 2016). Selfie mode continuous ISL video recognition was proposed by Rao et. al. which uses a deep learning based approach to solve the problem. They claim an accuracy of 92.88% over 46 ISL signs (Rao et al., 2018). A Coupled HMM-based method which use the leap motion sensor and Kinect sensor for ISL was proposed by Kumar et, al. This can recognize dynamic isolated sign gestures with an accuracy of 90.80% (Kumar et al., 2017). Availability of ISL dataset for a global reference is unavailable. Nandy et al. in their work Recognition of isolated Indian sign language gesture in real time (Nandy et al., 2010) and Recognizing & interpreting Indian sign language gesture for human-robot interaction (Nandy et al., 2010) used ISL dataset which can be used for research in this domain as a reference dataset.

In spite of having very high computational and cost overhead 3D sign language recognition have very high recognition rate. This is because the introduction of the third dimension actually solves many challenges in 2D SLR by providing extra information about the depth. It includes problems caused due to camera position, variation in illumination, background complexity, occlusions, etc. Kumar et al. proposed an adaptive graph matching approach for Indian SLR (Kumar et al., 2018). In this work, undirected graphs were used to represent signs based on 3D position trajectories. They created a 3D template for mostly all Indian signs in ISL. Adaptive graph matching along with motion segmentation accomplished sign identification. Performance of this approach was evaluated with various datasets like HDM05, CMU, 3D Sign datasets and they claim an accuracy of 96% and above. Motionlet matching with adaptive kernels. Kishore et al. (2018) was also contributed by the same team Kumar at.el. In this work initially, each frame is segmented to obtain motion joints and non-motion joints. After extraction of motion joints, it undergoes a classification into one of four predetermined classes. This is phase 1 and it deals with the hand tracking. Phase 2 deals with motionlets which is intra finger variations. Finger joint relative distance and joint angle measurements are used to extract the shape and orientation of 3D motionlets. To find the similarity between a query sign and database sign three feature kernels were created for each sign based on finger shape orientation and trajectories. They claim an accuracy of 98.9%. In both methods discussed above data capturing setup was very complex which involved eight IR, one video camera and signer wearable reflective markers.

From the literature survey, it is evident that most of the existing works focus on static gestures. Dealing with dynamic gestures was very difficult because of which for better accuracy special wearable and sensing equipment was used with more cost and computational overhead. This points to the fact that a fast, low cost and simple system of ISR which can be easily used by common people is highly required. The proposed method tries to solve some of the limitations of existing methods. The appearance-based systems are more user-friendly and efficient in terms of processing time than glove-based systems and 3-D hand model-based method. So we are focusing on the appearance-based method which is an economic solution for a mobile application with less storage space required. By contrast to the existing methods, we propose a more accurate method using shape descriptor and trajectory based gesture recognition, for continuous and dynamic words with co-articulation elimination. The system can distinguish each sign irrespective of hand size and skin color features.

3. System Overview

The gesture recognition system comprises of gesture-to-speech conversion. The input to the system is a video and the output will be a word corresponding to the gesture. The system considers single handed dynamic gesture recognition and finger spelling gesture recognition with co-articulation elimination. Each of the recognition module comprises of preprocessing, feature extraction and classification phase. The features extracted from each gesture recognition module are different. The input to gesture recognition module can be either a dynamic gesture or sequence of alphabets. In the case of finger spelling alphabets or continuous gesture with a sequence of ISL words, two consecutive gestures are mostly separated with some transition movements which should not be considered for recognition called co-articulation. These movements need to be spotted and eliminated in order to recognize the gestures correctly. Starting and ending phase of gestures also include unwanted movements and hence correct gesture spotting is highly essential. In our proposed system the input video undergoes gesture spotting and co-articulation elimination before passing it to gesture recognition module. The overall flow diagram of the gesture recognition module is shown in Fig. 1.

4. Proposed Method

As in the hierarchical representation of ISL(Ghotkar and Kharate, 2014) two basic classes of gestures, static and dynamic are considered as the input. The dynamic gestures are further classified into gestures with global motion and gestures with local motion. The system uses an external web camera for real-time video capturing which captures frames at a frame rate of 15 fps. From the set of input frames, the static and dynamic gestures are distinguished by their centroid change in successive frames. During the signing phase, the static gestures are identified based on the absence of centroid change in successive frames. If the centroid does not change in N number of frames it is considered as a static gesture. If this condition failed it is considered as either a dynamic gesture or a co-articulation region. The co-articulation region is identified using the acceleration feature. After the elimination of co-articulation phase, the gesture sequence is segmented. Based on the classification three different feature extraction techniques are considered. The motion trajectory (Bhuyan et al., 2008) of dynamic gestures (with global movement), shape-based recognition for dynamic gestures (with local movement) and Zernike moments calculation (Kalpana Sharma and Dutta, 2014Otiniano-Rodriguez et al., 2012) for static alphabets are considered in feature extraction phase. To identify the class of dynamic gesture a new threshold T_1 is considered. If the average centroid change is less than a lower threshold T_1, it is a gesture with a local hand movement. Otherwise, it is a dynamic gesture with global hand motion. The features obtained from the feature extraction phase are passed to SVM classifier for gesture recognition. The overall flow diagram of the proposed method is shown in Fig. 2.

4.1. Database Description

No standard dataset is available on ISL. So a new dataset is created that comprises of alphabets and dynamic words collected using an external web camera. A part of dataset has been collected from Rahmaniya HSS special school, Calicut. About 900 static images and 700 videos were considered for testing of alphabets and single-handed dynamic words, which are collected from seven

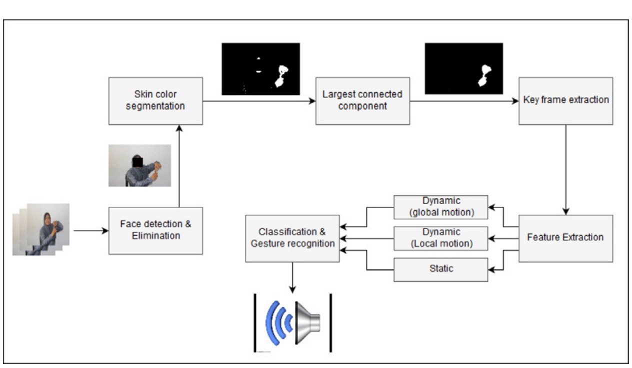

Fig. 1. Flow diagram of gesture recognition system.

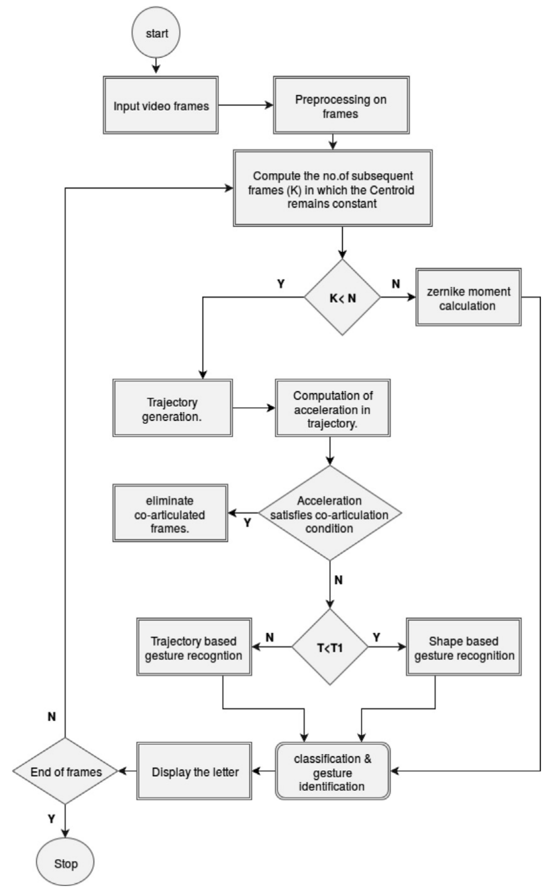

Fig. 2. Flow diagram of proposed method.

people. During the dataset collection, the user should wear at least half sleeve dress. The distance between the signer and the camera is adjusted to get the upper part of the body which is required for capturing dynamic gestures. Proper lighting condition and uniform background is considered for better results. Signers in different age groups and different genders are considered for signing to make the system useful to all. And signers have repeated the same signs multiple times in different positions and orientations for more accuracy.

4.2. Image Preprocessing & Hand Segmentation

The set of input frames are preprocessed for accurate hand region extraction using various image processing techniques. The preprocessing techniques such as face detection and elimination, hand segmentation, connected component extraction, and noise removal are considered for better hand segmentation. The ISL alphabets and lexicon considered in single-handed dynamic gesture recognition do not involve facial expressions or face occlusion. So the face region is detected and eliminated using Viola-Jones face detection algorithm (Yun and Peng, 2009). The input video sequence is composed of RGB values which are easily influenced by light. So to make it useful in most of the lighting conditions, the set of frames are then converted into YCbCr color space. YCbCr color space is less sensitive to light variations. From the skin color sampling results of various signers a skin color range is fixed in YCbCr color space and it is used for hand segmentation. The resultant frame may contain irrelevant skin color region. Large connected components in the frame are assumed to be the hand

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

region. All unwanted regions are removed by considering the largest area of hand region.

4.3. Region of interest extraction

Most of the existing works insists on wearing full sleeve dress while signing as a constraint (Subhash et al., 2014Jalal, 2015). And some of the works (Lilha and Shivmurthy, 2011) use wristband to identify the palm region. Here we use ROI algorithm for extracting palm region of the signer irrespective of any sleeve type or wristband. The algorithm is applied only if the height of bounding box is larger than a fixed height H. Otherwise it is assumed to be in full sleeve.

contour and convex hull (Chen, 2012Chen, 2012) is identified for centroid estimation and shape identification. Initially, the 0^{th} and 1^{st} order moments are calculated for getting the coordinates of the centroid. The 0^{th} and 1^{st} order moments are defined as (Geetha and Manjusha, 2012):

$$M_{00} = \sum_{x}\sum_{y} I(x,y) \tag{1}$$

$$M_{10} = \sum_{x}\sum_{y} x\, I(x,y) \quad M_{01} = \sum_{x}\sum_{y} y\, I(x,y) \tag{2}$$

Algorithm 1: ROI Algorithm

$N \leftarrow No.of frames$
for $i \leftarrow 0$ ***to*** N **do**
  Construct black image, B of size F_i /* F_i is i^{th} frame*/
  $\text{Rect}_i = SmallestBoundingBox(F_i)$
  ROI.width=Rect.width
  ROI.height=Fixed height
  G_i= Create a binary mask using B
  /* with all pixels inside the bounding box as white pixels*/
  ROI = BitwiseAND(F_i,G_i)
  return ROI

Another important thing is the elimination of the neck region which also possesses the skin color. The neck region is eliminated by extending the bounding box of face to a specific height and replacing it with a black image. The face and neck region elimination confirms the remaining largest skin color region to be the hand region. Implementation results of the algorithm are shown in Fig. 3.

Then centroid (x_c, y_c) is calculated.

$$x_c = \frac{M_{10}}{M_{00}} \; y_c = \frac{M_{01}}{M_{00}} \tag{3}$$

Algorithm 2: Centroid Calculation

Input: ROI
Output: Centroid of each frame.
$N \leftarrow No.of frames$
for $i \leftarrow 0$ ***to*** N **do**
  $\text{Rect}_i = SmallestBoundingBox(f_i)$
  (x, y) = Coordinates of Rect_i
  $(f_i.x, f_i.y)$ = centroid(bounded image in Rect_i)
  /* Calculating centroid based on moments.*/
  $(F_i.x, F_i.y)$ = (x+$f_i.x$,y+$f_i.y$)
  Centroid = $(F_i.x, F_i.y)$
  return Centroid
end

4.4. Centroid Estimation

After the image preprocessing steps the centroid of the binary image is estimated. The minimum bounding box of largest hand

The centroid from moments gives result by considering the No. of white pixel available in the preprocessed image and it is position invariant. So the movement of the bounding box is tracked to obtain centroid. Mapping of the centroid from moments and

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

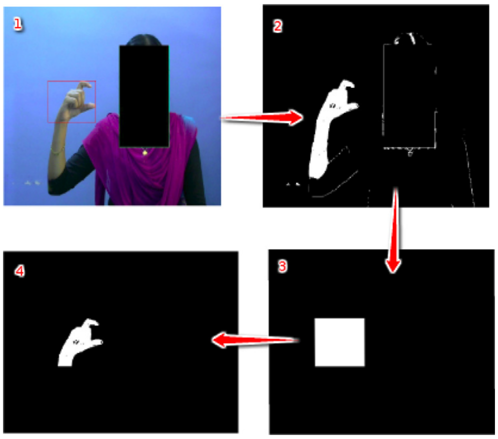

Fig. 3. 1) Bounding box identification 2) Skin color detection 3) Submask 4) Bitwise AND operation.

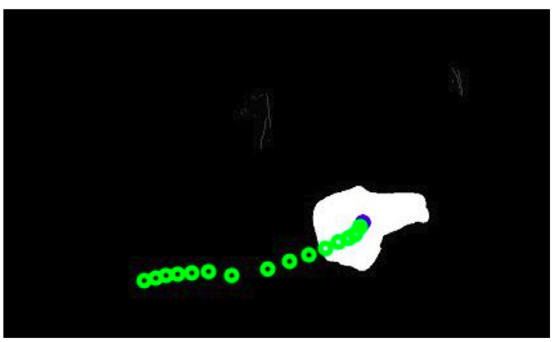

Fig. 4. Motion trajectory of the gesture 'Name'.

coordinates of the bounding box is used to get the required centroid. The centroid with respect to the movement of the hand is calculated using Algorithm 2. The sequence of the centroid, that forms trajectory is considered in trajectory-based gesture recognition. The implementation result of Algorithm 2 is shown in Fig. 4.

4.5. Key frame extraction

Uninformative frames of the gesture video can be identified by considering the frames that have no significant changes in their hand position or shape. The centroid of each frame is computed for identifying the positional change. The frames having a significant change in hand shape or centroid is considered as a keyframe. Only non-adjacent frames are considered for key frame extraction. It has been observed that in the input video of static alphabets the centroid does not change for at least N number of subsequent frames. Among that only $(N/2)^{th}$ frame is considered as a keyframe. In dynamic gestures, two thresholds are fixed for key frame extraction. One for the dynamic gesture with global motion and other for local motion. If the change of centroid between successive frames is larger than a threshold T_2, it is considered as a keyframe. Another threshold T_3 is set to find a significant change in shape. In dynamic gestures with local motion (shape-based dynamic gesture recognition), the shape descriptor Zernike moments are used to find the keyframes. If the Zernike Moments(ZM) difference between adjacent frames is greater than a value(value is set to 50) is considered as a keyframe.

The ZM difference is calculated using the following equation.

$$ZM\ Difference = \sum_{k=0}^{12} |ZM_i^A - ZM_i^B| \tag{4}$$

The keyframes obtained during shape based gesture recognition are shown in Fig.5. A twelve dimension vector is obtained by considering Zernike moments up to order 5, which is used in the shape identification phase. Zernike moments are constructed using a set of complex polynomials, which are the set of orthogonal polynomials defined on unit disc ($x^2+y^2 \leqslant 1$) (Kalpana Sharma and Dutta, 2014). The 2D Zernike moment A_{mn} with order m and angular dependence n is defined as:

$$A_{mn} = \frac{n+1}{\pi} \int_0^1 \int_0^{2\pi} R_{nm}(r) e^{-jm\theta} f(r,\theta) r\, dr d\theta \tag{5}$$

Where $j = \sqrt{-1}$, (r,θ) is defined over unit disc and R_{nm}(r) is the n^{th} Zernike radial polynomial.

$$0 \leqslant |m| \leqslant n,\ n - |m| = even \tag{6}$$

$$R_{nm}(r) = \sum_{k=0}^{\frac{n-|m|}{2}} (-1)^k \frac{(n-k)!\ (r)^{n-2k}}{k![n-2k+|m|/2]![n-2k-|m|/2]!} \tag{7}$$

4.6. Gesture spotting and Co-articulation detection

In continuous gesturing, one gesture follows the other in sequence. The phenomenon of influencing one gesture by the next in temporal sequence is called co-articulation (Yang et al., 2009). This is the goal of gesture spotting i.e. to locate the start point and end point of a gesture pattern, and to classify the gesture to one of predetermined gesture classes (Yang et al., 2009). The movement of the hand represents either a dynamic gesture or a co-articulation phase. Co-articulation is identified and eliminated using acceleration feature. Along with the acceleration condition, if the gesture satisfies a minimum requirement of the number of frames in a dynamic gesture, it is separated to dynamic block and recognized. It is observed that hand reaches a maximum acceleration during the co-articulation phase and reaches minimum acceleration during the gesture phase. The experiment shows that there is an increase in the acceleration of the hand between two gestures. This idea of the gradient of acceleration from one gesture to co-articulation and back to next gesture is used to eliminate co-articulation between two gestures in continuous gesture sentences.

4.6.1. Co-articulation detection between static gestures

Co-articulation between two static gestures are addressed in case of finger spelling alphabets. If a gesture does not change for more than N number of frames it indicates that it is a static gesture. This frame count N is determined during the training session. During static gesturing, the hand almost stay still for N frames and then moves with high acceleration to do the next static gesture, after which the acceleration falls to almost zero value again for the next static gesture. So the idea that co-articulation between

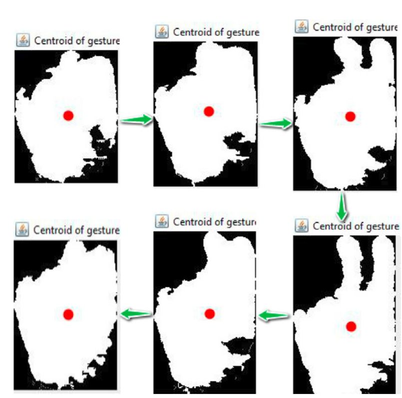

Fig. 5. Key frames of the dynamic gesture 'No' in shape based gesture recognition.

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

two static gestures occurs between two long pauses is used to implement the same. The results during the signing of continuous gesture TV are shown in Fig. 6. T and V are static gestures so the hand pauses in at least N number of frames.

4.6.2. Co-articulation detection between static and dynamic gestures

In case of a static gesture followed by a dynamic gesture, the signer keeps the hand in pause state in N number of frames. If that count is at least N it indicates the presence of static gesture. After the pause, the hand quickly moves to the starting of the next gesture. This quick movement is considered as co-articulation if the acceleration reached a maximum value at some point during the motion. Otherwise, the movement is considered as a dynamic gesture. Fig.7 shows a plot of acceleration during the continuous signing of a static gesture followed by a dynamic gesture.

4.6.3. Co-articulation detection between two dynamic gestures

In adjacent gestures having only global motion, the gesturing sequence forms motion trajectory for each of the dynamic gesture. After the completion of one trajectory, the hand pauses for a while and moves to the next gesture with high acceleration and again pauses before the next trajectory. This pause duration will be less than the static gesture time duration. The co-articulation phase is separated from both the dynamic gestures by spotting the minimum acceleration after the first dynamic gesture. If the next movement does not satisfy the co-articulation condition it is considered as the next dynamic gesture with no co-articulation region in between. Otherwise, it is a co-articulation phase. The co-articulation elimination and gesture recognition can be done using Gesture separation algorithm.

Algorithm 3: Gesture Separation Algorithm

Algorithm 3: Gesture Separation Algorithm

```
Input: Sequence of preprocessed frames after ROI extraction.
Output: Blocks of meaningful frames.
N ← No.offrames
for i ← 0 to N do
    Centroid[i] ← C_i
    Acceleration[i] ← acceleration(F_i, F_{i+1})
end
/* T is count of frames meeting static gesture conformation condition.*/
for i ← 0 to N do
    if Centroid[i] − Centroid[i+1] == Zero then
        count ← count + 1
        if count >= T then
            StaticBlock_i = F_{i−count} to F_i
            count ← 0
        end
    else if Acceleration[i]! = Min then
        Isdynamic ← true
        /* Min is the minimum value of acceleration reaches next.*/
        while (j == Min) do
            if acceleration[j] == Max then
                Isdynamic ← false
            end
            j ← j + 1
        end
        /* Max is the maximum value of acceleration during gesturing .*/
        loc ← j
        if Isdynamic then
            DynamicBlock_i ← F_i to F_loc
        end
        else
            CoArtblock_i ← F_i to F_loc
        end
end
```

Fig. 6. Co-articulation between the gesture TV.

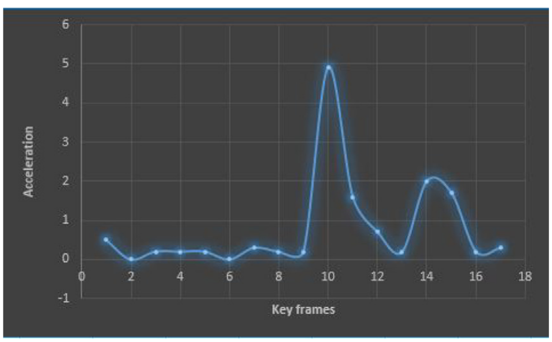

Fig. 7. Acceleration plot for continuous gestures.

4.7. Feature Extraction

In single-handed dynamic gestures (with global hand movement) hand shape, hand motion, hand location, and hand orientation are considered as features. If a global motion is there, it is found that the hand forms a trajectory. So trajectory based gesture recognition is selected (Kumar et al., 2016). A feature vector of dimension 6 is considered in this module.

feature vector = [Hand shape, Trajectory length, average speed, number of significant curves, the number of points of minima, palm orientation].

- Hand shape:The Zernike moments of keyframes is calculated for shape identification.
- Number of significant curves: The orientation change at each keyframe is considered if the orientation change is $>= 45^{\circ}$.
- Trajectory length: The total length of the trajectory is calculated by summing up the distance between keyframes.
- Average speed over the whole trajectory: If more is the number of direction changes lower will be the average speed.
- Number of points of minima: The speed of the hand becomes very low compared to some threshold T_l at points having sharp changes in trajectory.
- Palm orientation: It is the measure of palm orientation while signing.

The types of gestures with local motion is recognized using a shape-based recognition process. From the experimental results the words that failed to satisfy the threshold have no positional change, but only shape change. Zernike moments of each such frames are calculated. The feature vector includes hand shape, average speed, and palm orientation. And the dimension of the feature vector depends on the number of keyframes.

In static alphabets recognition, Zernike moments up to 5^{th} order and repetition up to 5 are considered. The resultant feature vector of dimension 12 is used for classification. The magnitude of Zernike moments is found as a good shape descriptor and that gives high accuracy even at order 5. Among ISL alphabets J and H are dynamic. They are recognized using 6-dimension feature set as in trajectory-based gesture recognition.

4.8. Classification

The proposed method gives much emphasis on preprocessing and feature extraction. Next important phase is the classification in which gestures are correctly classified into corresponding gesture classes based on the calculated features. Here we use Support Vector Machine classifier(SVM) for classification which finds the optimal separating hyperplane between classes based on supervised learning on the training data. Given a set of training examples each marked as belonging to one of the two categories an SVM training algorithm builds a model that predicts whether a new example falls into one category or other. Many research works (Jalal, 2015Subhash et al., 2014) suggests SVM classifier is a good choice for better gesture recognition. Three classification models are built, one for Zernike moments, next for trajectory-based recognition and last one for shape-based recognition of words. The recognition of the test image is done based on the training set. Initially, images in the dataset are trained using multi-class C-SVC (Chang and Lin, 2011) having radial basis function kernel with one against all strategy. When a new image is given in the testing phase, preprocessing is applied to the input image then the feature vector is formed. Through the classifier, the most matching gesture is recognized and displayed its meaning. The training and classification with SVM classifier are done using JAVA Machine Learning library JAVAML (Java-ML, 2019).

5. Experimental Results

Skin color segmentation using YCbCr colorspace gave better results on a uniform background and under proper lighting conditions. The performance of the gestures is obtained by varying the size of the dataset. The accuracy of the system is calculated using the equation.

$$Accuracy = \frac{Correctly\ classified\ gestures}{Total\ no.\ of\ gestures}\ X\ 100 \quad (8)$$

In single-handed dynamic gesture recognition the processing time is reduced by considering only keyframes in the feature extraction phase. The graph plotted using the number of keyframes against total frames is shown in Fig. 8. The accuracy of single-handed dynamic gesture recognition is 89%. The Gesture, 'Right' is giving less accuracy rate compared to other words due to the movement of a finger rather than whole hand movement. All other words are giving a good recognition rate in trajectory-based recognition and shape based recognition. Fig. 9 shows the recognition rate of single-handed dynamic words.

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002

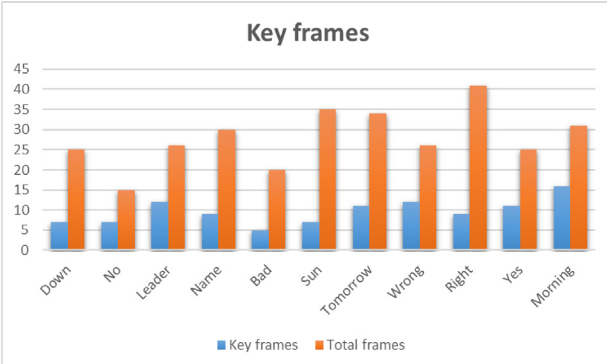

Fig. 8. No. of Keyframes extracted per video.

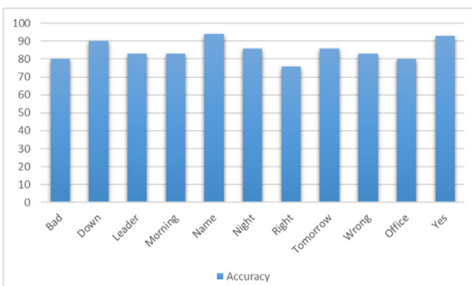

Fig. 9. Recognition rate of single-handed dynamic words.

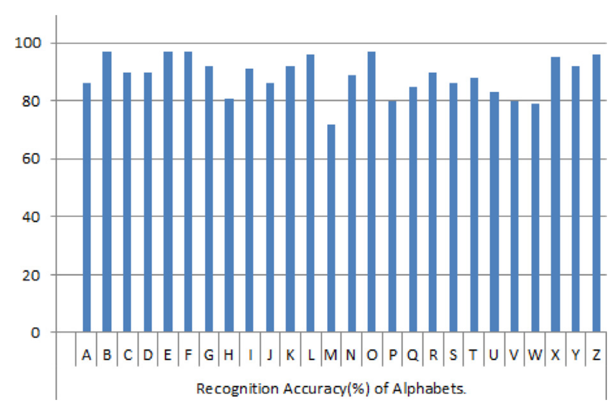

Fig. 10. Recognition rate of Finger spelling alphabets.

In continuous signing 24 static alphabets and 2 dynamic alphabets (J,H) are considered. Recognition using Zernike moments give better results on static gestures. The gesture recognition rate alphabets is shown in Fig. 10.

Due to the shape similarity of some alphabets [(M, N), (C, L)], M and C got less recognition rate. The accuracy of fingerspelling gesture recognition is 91%. Co-articulation detection and elimination are done with an accuracy of 100% between two static gestures, static and dynamic gestures and between two dynamic gestures on the test datasets.

6. Comparison with other methods

In the proposed method for ISL recognition, YCbCr color space is used for skin color segmentation, Viola-Jones method for face elimination, Zernike moments for feature extraction and multi-class SVM for classification. The proposed method has comparatively

Table 1
Comparison with other methods.

Ref.	Segmentation	Feature Vector	Classifier	Dataset (class,data)	Static Acc	Dynamic Acc	Remarks
Geetha and Manjusha (2012)	Background subtraction and connected component	B-Spline Approximation	SVM	(29,290)	≈ 90	–	Only static alphabets, complex fitting approximation
Rekha et al. (2011)	YCbCr	Principal curvature, 2-D wavelet decomposition, finger count	Multi-class SVM, Dynamic time wraping DTW	(23,230), (3,60)	86.3	77.2	Small dataset, principal curvature feature less stable in case of dynamic gestures.
Kalpana Sharma and Dutta (2014)	N/A	Zernike moments	SMO	(5,720)	94.4	–	Recognize only five static alphabets.
Bhuyan et al. (2008)	N/A	Trajectory length, no: of significant curve, speed, standard deviation, No: of minima.	Euclidean distance	(10,585)	–	95.8	Only dynamic gestures with global motion, vocabulary of only ten gestures.
Lilha and Shivmurthy (2011)	N/A	Histogram of Edge Frequency	SVM	(26,1560)	98	–	Use of wrist band
Bhuyan et al. (2014)	YCbCr	Position, orientation, length of ellipse least-squares fitted on motion trajectory	CRF	(10,500)	–	90	Only single-handed dynamic trajectory based numerals
Proposed method	YCbCr	Zernike moments & curve features	SVM	(24,786), (13,165)	90.1	89	Relatively high accuracy for static, dynamic and continious gesture sequences.

good accuracy and also it does not require any wrist band to recognize the palm region. Literature shows a number of different combinations of feature sets, segmentation methods and classifiers which show different recognition rates. Some of them are summarised in Table 1. In the table, the method for recognizing 26 English alphabets using the histogram of edge frequency and SVM (Lilha and Shivmurthy, 2011) shows an accuracy of 98%, but they use a wrist band to identify the palm region and must wear at least half sleeve dress. Similarly, Sharma et al. proposed a method for recognizing five static symbols with Zernike moments and Sequential Minimal Optimizer (SMO) (Kalpana Sharma and Dutta, 2014). This method achieved very good accuracy of around 94.4% but was proposed for only 5 static symbols. Bhuyan et al. proposed two methods for single-handed dynamic gesture recognition involving global motion (Bhuyan et al., 2008Bhuyan et al., 2014). Ten dynamic gestures which involve the movement of complete hand was the problem focus in both the cases which was proposed for human-computer interaction rather than sign language recognition. Comparison with other methods is done on different datasets as there are no standard datasets available for ISL. It is evident from the table that higher recognition rate for certain methods is either because of small vocabulary, small dataset or use of a wristband. In general, the proposed method which can deal with static gestures, dynamic gestures and co-articulation points together with relatively high recognition rate is an improvement over many of the existing methods.

7. Conclusion

The method follows a vision-based gesture recognition system to recognize static, dynamic and finger spelling words of ISL. This approach is very economical and can be implemented even with a mobile camera which makes it very user-friendly to be used by a common man. Keyframe extraction module in this work speeds up the computation and can be used in real time ISL recognition. Co-articulation detection between two static gestures, static and dynamic gestures and between two dynamic gestures are addressed using a gradient of acceleration approach and obtained 100% accuracy with test dataset. Introduction of trajectory-based method produced remarkable results in dynamic gesture recognition. Due to the lack of availability of dataset in ISL a new dataset is created which includes static gestures, dynamic gestures, and finger spelling alphabets. Experiment results show good recognition accuracy for static and dynamic gestures, finger spelling words and co-articulation detection and elimination.

7.1. Future work

The method can be enhanced by considering dataset captured under cluttered backgrounds and different illumination conditions. ISL sentence recognition is still a very less explored area. A real-time ISL sentence recognition system which takes into consideration facial expressions and different contexts will be a great achievement. As a future work 3-D gestures and non-manual signs can be included to make the system much useful for hearing impaired people.

Acknowledgements

We would particularly like to thank Mrs. Nasriya (Teacher, Rahmaniya HSS) and students in Rahmaniya HSS special school regarding data collection. We express our special thanks to Karuna special school and National Institute of Speech and Hearing (NISH) for their valuable suggestions and support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.jksuci.2019.05.002.

References

Adithya, V., Vinod, P., Gopalakrishnan, U., 2013. Artificial neural network based method for indian sign language recognition. In: Information & Communication Technologies (ICT), 2013 IEEE Conference on. IEEE, pp. 1080–1085.

Ashok Kumar Sahoo, K.K.R., 2014. Vision based indian sign language character recognition. J. Theor. Appl. Inf. Technol. 67 (3), 770–780.

Bhuyan, M., Ghosh, D., Bora, P., 2005. Co-articulation detection in hand gestures. In: TENCON 2005 2005 IEEE Region 10, IEEE, pp. 1–4.

Bhuyan, M., Ghosh, D., Bora, P., 2006. Continuous hand gesture segmentation and co-articulation detection. In: Computer Vision, Graphics and Image Processing. Springer, pp. 564–575.

Bhuyan, M., Bora, P., Ghosh, D., 2008. Trajectory guided recognition of hand gestures having only global motions. World Acad. Sci. Eng. Technol. 21, 753–764.

Bhuyan, M.K., Kumar, D.A., MacDorman, K.F., Iwahori, Y., 2014. A novel set of features for continuous hand gesture recognition. J. Multimodal User Interfaces 8 (4), 333–343.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2. 27:1-27:27, software available athttp://www.csie.ntu.edu.tw/cjlin/libsvm.

Chen, Q., 2008. Real-time vision-based hand tracking and gesture recognition. University of Ottawa, p. 1–94.

Chen, W.-C., 2012. Real-time palm tracking and hand gesture estimation based on fore-arm contour. Department of Computer Science and Information Engineering (Master Dissertation).

Cheng, H., Yang, L., Liu, Z., 2016. Survey on 3d hand gesture recognition. IEEE Trans. Circuits Syst. Video Technol. 26 (9), 1659–1673.

Deora, D., Bajaj, N., 2012. Indian sign language recognition, First International Conference on Emerging Technology Trends in Electronics, Communications and Networking, pp. 1–5.

Dixit, K., Jalal, A.S., 2013. Automatic indian sign language recognition system. In: Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, pp. 883–887.

Dong, C., Leu, M.C., Yin, Z., 2015. American sign language alphabet recognition using microsoft kinect. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 44–52.

Elakkiya, R., Selvamani, K., 2018. Enhanced dynamic programming approach for subunit modelling to handle segmentation and recognition ambiguities in sign language. J. Parallel Distrib. Comput. 117, 246–255.

Geetha, M., Aswathi, P.V., 2013. Dynamic gesture recognition of indian sign language considering local motion of hand using spatial location of key maximum curvature points. In: 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 86–91.

Geetha, M., Manjusha, U.C., 2012. A vision based recognition of indian sign language alphabets and numerals using b-spline approximation. Int. J. Comput. Sci. Eng. 4 (03), 406–415.

Ghotkar, A.S., Kharate, D.G.K., 2014. Study of vision based hand gesture recognition using indian sign language. International Conference on Smart Sensing and Intelligent Systems 7 (1), 96–115.

Hussain, I., Talukdar, A.K., Sarma, K.K., 2014. Hand gesture recognition system with real-time palm tracking. In: India Conference (INDICON), 2014 Annual IEEE. IEEE, pp. 1–6.

Ibrahim, N.B., Selim, M.M., Zayed, H.H., 2018. An automatic arabic sign language recognition system (ArSLRs). J. King Saud Univ.-Comput. Inf. Sci.

Jalal, D.A.S., 2015. Automatic recognition of dynamic isolated sign in video for indian sign language, 1–49.

Java-ML, 2019. Java Machine Learning Library,http://java-ml.sourceforge.net/, [Online].

Kalpana Sharma, G.J., Dutta, M., 2014. Analysis of shape and orientation recognition capability of complex zernike moments for signed gestures, pp. 730–735.

Kapuscinski, T., Oszust, M., Wysocki, M., Warchol, D., 2015. Recognition of hand gestures observed by depth cameras. Int. J. Adv. Rob. Syst. 12 (4), 36.

Kharate, G.K., Ghotkar, A.S., 2016. Vision based multi-feature hand gesture recognition for indian sign language manual signs. Int. J. Smart Sens. Intell. Syst. 9 (1).

Kim, K., Kim, S.-K., Choi, H.-I., 2015. Depth based sign language recognition system using SVM. Int. J. Multimedia Ubiquitous Eng. 10 (2), 75–86.

Kishore, P., Kumar, D.A., Sastry, A.C.S., Kumar, E.K., 2018. Motionlets matching with adaptive kernels for 3d indian sign language recognition. IEEE Sens. J.

Kong, W., Ranganath, S., 2014. Towards subject independent continuous sign language recognition: a segment and merge approach. Pattern Recogn. 47 (3), 1294–1308.

Kumar, A., Thankachan, K., Dominic, M.M., 2016. Sign language recognition. In: Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on. IEEE, pp. 422–428.

Kumar, P., Gauba, H., Roy, P.P., Dogra, D.P., 2017. Coupled hmm-based multi-sensor data fusion for sign language recognition. Pattern Recogn. Lett. 86, 1–8.
Kumar, D.A., Sastry, A., Kishore, P., Kumar, E.K., 2018. Indian sign language recognition using graph matching on 3d motion captured signs. Multimedia Tools Appl., 1–29
Li, H., Greenspan, M., 2007. Segmentation and recognition of continuous gestures. In: Image Processing, 2007. ICIP 2007. IEEE International Conference on, vol. 1. IEEE, pp. I–365.
Lilha, H., Shivmurthy, D., 2011. Evaluation of features for automated transcription of dual-handed sign language alphabets. In: International Conference on Image Processing, pp. 1–5.
Nandy, A., Prasad, J.S., Mondal, S., Chakraborty, P., Nandi, G.C., 2010. Recognition of isolated indian sign language gesture in real time. In: Inf. Process. Manage. Springer, pp. 102–107.
Nandy, A., Mondal, S., Prasad, J.S., Chakraborty, P., Nandi, G., 2010. Recognizing & interpreting indian sign language gesture for human robot interaction. In: 2010 international conference on computer and communication technology (ICCCT). IEEE, pp. 712–717.
Nanivadekar, P.A., Kulkarni, V., 2014. Indian sign language recognition: Database creation, hand tracking and segmentation. In: Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on. IEEE, pp. 358–363.
Neha, P.A.V.D., Tavari, V., Chatur, D.P., 2014. Chatur, Hand gesture recognition of indian sign language to aid physically impaired people, pp. 60–66.
Otiniano-Rodríguez, K., Cámara-Chávez, G., Menotti, D., 2012. Hu and zernike moments for sign language recognition. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), The Steering Committee of The World Congress in Computer Science. Computer Engineering and Applied Computing (WorldComp), p. 1.
Patil, S.B., Sinha, G., 2017. Distinctive feature extraction for indian sign language (ISL) gesture using scale invariant feature transform (SIFT). J. Inst. Eng. (India): Series B 98 (1), 19–26.
Prisacariu, V.A., Reid, I., 2011. Robust 3d hand tracking for human computer interaction. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, pp. 368–375.
Rao, G.A., Syamala, K., Kishore, P., Sastry, A., 2018. Deep convolutional neural networks for sign language recognition. In: 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES). IEEE, pp. 194–197.
Rekha, J., Bhattacharya, J., Majumder, S., 2011. Shape, texture and local movement hand gesture features for indian sign language recognition. In: Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on. IEEE, pp. 30–35.
Sharma, M., Pal, R., Sahoo, A.K., 2014. Indian sign language recognition using neural networks and knn classifiers. ARPN J. Eng. Appl. Sci. 8.
Suarez, J., Murphy, R.R., 2012. Hand gesture recognition with depth images: A review. In: RO-MAN, 2012 IEEE. IEEE, pp. 411–417.
Subhash, A.S.J., Agrawal, Chand, Bhatnagar, C., 2014. Redundancy removal for isolated gesture in indian sign language and recognition using multi-class support vector machine. Int. J. Comput. Vision Rob. 4 (1), 23–38.
Tripathi, K., Baranwal, N., Nandi, G.C., 2015. Continuous dynamic indian sign language gesture recognition with invariant backgrounds. In: Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE, pp. 2211–2216.
Wang, R.Y., Popović, J., 2009. Real-time hand-tracking with a color glove. In: ACM transactions on graphics (TOG), vol. 28. ACM, p. 63.
Yang, H.D., Sclaroff, S., Lee, S.W., 2009. Sign language spotting with a threshold model based on conditional random fields. IEEE Trans. Pattern Anal. Mach. Intell. 31 (7), 1264–1277.
Yun, L., Peng, Z., 2009. An automatic hand gesture recognition system based on viola-jones method and SVMS. Second International Workshop on Computer Science and Engineering, 72–77.
Zaki, M.M., Shaheen, S.I., 2011. Sign language recognition using a combination of new vision based features. Pattern Recogn. Lett. 32 (4), 572–577.

Please cite this article as: P. K. Athira, C. J. Sruthi and A. Lijiya, A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2019.05.002