ISSN:0975-9646

A Review on Feature Extraction for Indian and American Sign Language

Neelam K. Gilorkar, Manisha M. Ingle

Department of Electronics & Telecommunication,
Government College of Engineering, Amravati, India

***Abstract*- In the field of human computer interaction (HCI) Sign Language have been the emphasis of significant research in real time. Such systems are advance and they are meant to substitute interpreters. Recently several techniques have been advanced in this area with the improvement of image processing and artificial intelligence techniques. Indian Sign Language(ISL) are double handed and hence it is more intricate compare to single handed American Sign Language(ASL). Many researchers use only single ASL or ISL sign for creating their database. In this paper recent research and development of sign language are reviewed based on manual communication and body language. Sign language recognition system typically elaborate three steps preprocessing, feature extraction and classification. Classification methods used for recognition are Neural Network(NN), Support Vector Machine(SVM), Hidden Markov Models(HMM), Scale Invariant Feature Transform (SIFT),etc.**

***Keywords*-Feature extraction, Hidden Markov Model(HMM) , Neural Network (NN), Support Vector Machine(SVM), Scale Invariant Feature Transform.**

1. Introduction

Sign language is the medium of communication language generally used by deaf-dump community. It uses gestures instead of sound patterns to convey meaning. Various gestures are composed by movements and orientations of hand, body or facial expressions and lip-patterns for communicating information or emotions. Sign language is not universal and just like spoken language, it has its own regional dialects. American Sign Language (ASL), British Sign Language (BSL), Indian Sign Language (ISL) etc. are some of the common sign languages in the world. Across the world million's of people are deaf. They find it difficult to communicate with the normal people as the hearing or normal people are unaware of sign language. There arises the need for sign language interpreters who can interpret sign language to spoken language and vice versa. But, the availability of such interpreters is limited ,expensive and does not work throughout the life period of a deaf person. This resulted in the development of automatic sign language recognition systems which could automatically translate the signs into corresponding text or voice without the help of sign language interpreters. Such systems can help in the development of deaf community through human computer interaction and can bridge the gap between deaf people and normal people in the society.

By deaf community in ISL is expressed by both hand gestures and in ASL is expressed by single hand gesture as illustrated in figure 1. It consists of both word level gestures and finger spelling. Fingerspelling is used to form words with letter by letter coding. Letter by letter signing can be used to express words for which no sign exists, the words for which the signer does not know the gestures or to emphasis or clarify a particular word. So the recognition of the fingerspelling has a key importance in sign language recognition. This paper present a method of automatic recognition system of static gestures in Indian and American sign languages alphabet and numbers. The sign considered for recognition are 26 letters alphabets and the 0-9 numbers. Some ASL and ISL gestures are shown in figure 1.

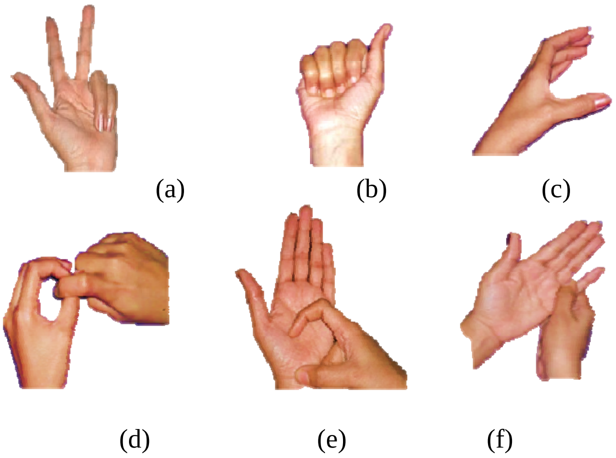

Figure 1. ASL signs for (a)3, (b)A, (c)C and ISL signs for (d)Q, (e)R, (f) S

2. Different Approaches For Sign Language Recognition

Many researchers have endorsed various techniques for sign language recognition systems. Mainly, these systems are alienated into two approaches, namely the glove-based and the vision-based approach[1]. The first category requires user to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation during processing. The Glove based approaches utilize sensor devices for digitizing hand and finger movements into multi-parametric data. The additional sensors facilitate detection of hand configuration and movement. However, the sensor devices are quite costly and wearing gloves or trackers is sore and enlarges the "time-to interface," or setup time. However, the data glove and its attached wires are still inconvenient and awkward for users to wear.

Moreover, the cost of the data glove is often too expensive for regular users. The shortcoming of this approach is that the user has to wear the sensor hardware along with the glove during the operation of the system. On the other hand, vision-based methods are more natural and useful for real-time applications. Vision based approach uses image processing algorithms to detect and track hand signs as well as facial expressions of the user. This approach is easier to the user since there is no need to wear any extra hardware. However, there are accuracy problems related to image processing algorithms and these problems are yet to be modified.

Also,vision based systems need application-specific image processing algorithms, programming, and machine learning. The necessity for all these to work in a variety of environments causes several issues because these systems require user and camera independent and invariant against the background and lighting changes to attain real-time performance. From the perspective of the features used to represent the hand, vision based hand tracking and gesture recognition algorithms are classified into two categories:

- 3D hand model-based approach
- Appearance based approach

2.1 3D Hand Model Based Approach

Many methodologies used 3D kinematic hand model with significant degrees of freedom (DOF), and calculate the hand parameters by matching the input frames and the appearance projected by the 3D hand model. 3D model based methods make use of 3D information of key elements of the body parts. Using this information, several important parameters, like palm position, joint angles etc., can be obtained. This approach uses volumetric or skeletal models, or a combination of the two. In computer animation industry and for computer vision purposes, volumetric approach is better suited. This approach is very computational intensive and also, systems for live analysis are still to be developed. However, 3D hand model method required a huge database with the entire characteristic shapes under several views due to deformable objects with many DOFs.

2.2 Appearance Based Approach

Appearance-based systems use images or videos as inputs. They directly understand from these videos/images. They don't use a spatial representation of the body. The parameters are derived directly from the images or videos using a template database. In this method image features are extract to model the visual appearance of the hand and compare these features with the extracted features from the video frames as our approach. They have real time performance because of the easier 2-D image features that are used[2].

The prime issues in hand gesture recognition are: (i) hand localization, (ii) scale and rotational invariance and (iii) viewpoint and person/user independence. Earlier works expected the gestures to be performed in a uniform background. This required a simple thresholding technique to obtain the hand silhouette. For a non-uniform background, skin color detection is the most popular and the general method for hand localization . The skin color cues are combined with the motion cues and the edge information for improving the efficiency of hand detection. Segmented hand images are usually normalized for the size, orientation and illumination variations. The features can be extracted directly from the intensity images or the binary silhouettes or the contours.

A general block diagram of sign language recognition system is shown in Figure 1. The recognition process is alienated into two phases- training and testing. In the training phase, the classifier has to be trained using the training dataset. The database can be either created by the researcher himself or an available database can be used. An external webcam, digital camera or inbuilt webcam in the laptops can be used to capture the training images. Most of the sign language recognition systems classify signs according to hand gestures only or in other words, facial expressions are omitted. The important steps involved in training phase are creation of database, preprocessing, feature extraction and training the classifier. The testing phase contains video/image acquisition (input can be videos or images), preprocessing, feature extraction and classification.

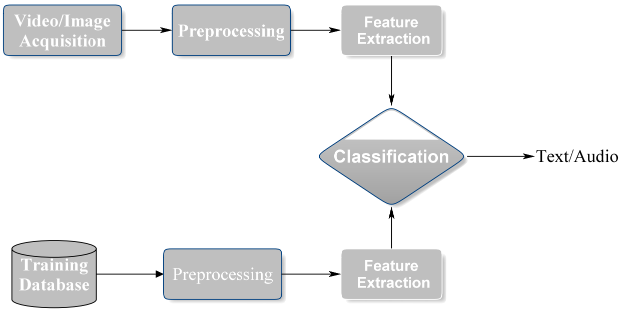

Figure 2. Generalized Block Diagram of Sign Language Recognition System

3. Data Acquisition

Data acquisition should be mostly perfect as possible for efficient hand gesture recognition. Suitable input device should be selected foe data acquisition. There are many input devices for data acquisition. Some of them are Data gloves, markers, hand images from webcam or stereo camera.

4. Preprocessing

A preprocessing step is carried out on the training images to extract the region of interest (ROI). The ROI can be hands if only hand gestures are considered or both face and hands if the facial gestures are also included. Usually the preprocessing step consists of filtering, image enhancement, image resizing, segmentation, edge detection, normalization and morphological filtering. Filtering and image enhancement can be any one of the commonly used methods. For segmentation, the algorithm that better suits the input video/images has to be selected. Thresholding, Background subtraction, skin-based and motion based segmentation, noise removal, statistical model are the commonly used segmentation techniques.

During testing phase, the test images or videos are also preprocessed to extract the region of interest.

5. FEATURE EXTRACTION

Under different condition, the performance of different feature detectors will be significantly different. The features should be efficiently and reliably extracted to find shapes and robustly irrespective of changes in illumination levels, position, orientation and size of the object in a video/image. In an image, objects are represented as summation of pixels and for recognition of an object there is need to describe the properties of these groups of pixels. The features can be obtained in different ways- wavelet decomposition , Haar wavelets, Haar-like features [6], texture features, Euclidean distance[7], scale invariant feature transform[11], Principal Component Analysis (PCA) [10], Fourier descriptors etc. The feature vector thus obtained using any one of the feature extraction methods is used for training the classifier. Thus feature extraction is the most crucial step of sign language recognition since the inputs to the classifier are the feature vectors obtained from this step.

6. CLASSIFICATION

A classifier task is to assigned a new feature vector to some predefined categories in order to recognize the sign. The category consists of a set of features obtained during the training phase using number of training images. Classification mainly concentrates on finding the best matching features vector among the set of reference features and displays the text or plays the sound. The test inputs can be images or videos. Most frequently used classifiers are Hidden Markov Models (HMM), Neural Networks (NN), Multiclass Support Vector Machines (SVM), Fuzzy systems, K Nearest Neighbor (KNN) etc. In terms of recognition rate, the performance of the classifier is measured.

6.1 Hidden Markov Models (HMM)

As a first preference, researches prefer to use Hidden Markov Model (HMM) [3] for the data containing information for dynamic hand gesture recognition. HMM is a doubly stochastic model and is appropriate for dealing with the stochastic properties in gesture recognition. The first well known application of HMM technology was speech recognition. A Hidden Markov Model is a collection of finite states connected by transitions. Each state is characterized into two sets of probabilities: a transition probability and a discrete or continuous output probability density function which gives the state, defines the condition probability of each output symbol from a finite alphabet or a continuous random vector. HMMs are employed to represent the gestures, and their parameters from the training data. Based on the most likely performance criterion, the gestures can be recognized by evaluating the trained HMMs. Another advantage of HMM is its high recognition rates. Some researches used HMM combining with other classifiers. The topology for an initial HMM can be resolute by estimating how many different states are intricate in specifying a sign.

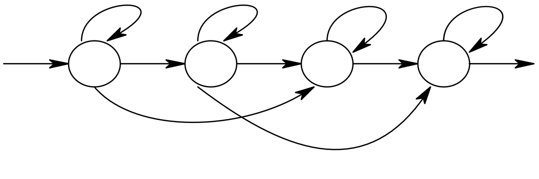

Figure 3. The four state HMM

6.2 Haar-like Technique

In [6], the Haar-like feature is applied for hand detection. This approach consider adjacent rectangular regions at a specific location in a detection window. Haar-like features concentrate more on the information within a certain area of the image rather than each single pixel. To enhance classification accuracy and attain real-time performance, the AdaBoost learning algorithm, was used which adaptively choose the best features in each step and combine them into a strong classifier. Haar-like techniques use Haar wavelet for successful gesture recognition. Each Haar-like feature comprises of two ot three connected " black" and "white" rectangles. Figure 4 shows the extended Haar-like feature set that was proposed by Lienhart and Maydt[13]. The value of a Haar-like feature is the difference between sums of the pixel values in the black and white rectangles,i.e.

$$f(x) = \sum_{black} (pixel\ value) - \sum_{white} (pixel\ value) \qquad (1)$$

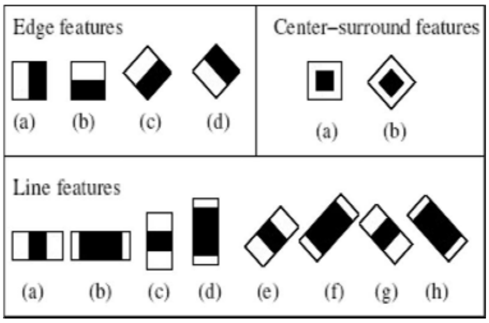

Figure 4. Extended set of Haar-like features [13]

6.3 Neural Network

Neural network model is motivated by the biological nervous system, like the brain to process information. In [7], a neural network model was used to recognize a hand gesture in an image. The model consists of a large number of highly interconnected processing electrons (artificial neurons) working in unity to solve particular problems. There are various neural network algorithms used for gesture recognition such as feed forward and back propagation algorithms. Feed forward algorithm is used to calculate the output for a specific input pattern. Back propagation algorithm is used for learning of the network.

6.4 Support Vector Machine (SVM)

The SVM [10], is a popular pattern recognition learning technique for classification. Classifiers of SVM are nearly related to neural networks. Actually, four basic concepts:-separating hyperplane,maximum margin hyperplane, soft margin and the kernel function need to be know, for optimal utilization of SVM classifier. In some cases researches only used basic SVM and in some Multiclass SVM or in a fused state with other methods.

6.5 Scale Invariant Feature Transform (SIFT)

Scale Invariant Feature Transform (SIFT**)** [9] [11], are features (key points) extracted from imges to help in reliable matching between different views of the same object, image classification and object recognition. The extracted keypoints are invariant to scale, orientation and partially invariant to illumination changes and are highly distinctive of the image. Computation of SIFT images features is executed through four phases :

- Scale-Space Local Extreme
- Keypoint Localization
- Orientation Assignment
- Keypoint Descriptor

7. Research Results

Sign language recognition is thought-provoking domain which uses computer vision and graphics, image processing , machine learning techniques and psychology [12] for effective operation. The overall research summary is presented in table 1. Table describes the type of backgrounds used, methods amended for segmentation, extracted features and the recognition techniques along with accuracies.

There are numerous restrictions in this systems. They can be framed as below .

Special hardware: A number of special hardware like range camera, data gloves[2] have been used.

Skin-like coloured objects: Objects with similar colour of human skin may be existing in the environment and may lead to confusion for recognition[4].

Change in illumination: When illumination and lightening conditions variations occured between the training datasets and inputs , the system fails to recognize accurately[7].

Difficult background: Maximum papers use uniform background [6] for recognition. But in real time uniform background is not desirable.

Orientation limitation: A sign language recognition system fails to recognize if the hand is orientated in a different angle[7][6][3].

Scaling problem: Due to different field of applications, scaling problem may arise.

Every single model has certain limitation. Each researches addressed one issue at a time not simultaneously. From the discussions , it has found that below conditions are essential for the sign recognition.

- Robustness
- Computationally efficiency
- Scalability
- Orientation

Conclusion

In this paper, we have deliberated the static signs of ISL and ASL from images or video sequences that have been recorded under controlled conditions. Special hardware like data gloves, coloured gloves or markers is used in some systems. Certain systems use only bare hands with single background consideration. Most of the systems are user dependent and also, the user needs to wear full sleeves. Facial features are not included in majority of the systems. Improvement of user independent systems which could recognize signs from both facial and hand gestures is a major challenge in the area of sign language recognition. As sign language recognition is finding its application for nonverbal communication, normal and deaf-dump people, 3d gaming, etc. With increase in applications researchers should emphasis on the development of a well suited segmentation scheme which is capable of extracting the hand and face region from videos/images having any background. More emphasis should also be given for extracting such features which could completely distinguish each sign regardless of hand size, distance from the source, color features and lighting conditions. Lastly the necessities and shortcomings for a perfect sign language recognition system have been deliberated.

Table1. Comparison of Diverse Sign Language Recognition Systems

Method [Ref.No.]	Back-ground	Segmentation Technique	Feature Vector Re-presentation	Classifier/ Recognition	Accuracy (%)
[2]	Data Glove	Discontinuity(time variant parameter detection)	Posture, position, orientation and motion	HMM	80.4
[4]	Cluttered	Manually from Back-ground	Blob and ridge features	Particle filtering	86.5
[3]	Uniform	-	State and transition features	HMM	85
[6]	Uniform	HSV(Hue, Saturation,Value) color space	Convex hull of the contour	Haar like Technique	N/A
[7]	Uniform	RGB to Gray	Euclidean Distance	Neural Network	N/A
[10]	Cluttered	HSV color space	SIFT features	Multiclass SVM	96.25

ACKNOWLEDGMENT

I express my sense of gratitude and sincere regards to my guide M. M. Ingle I would like to thanks all the staff members of Electronics and Telecommunication Department .

REFERENCES

[1]Y. Wu, J. Lin, and T. Huang, “Capturing Natural Hand Articulation”. In IEEE International Conference on Computer Vision II, 426–432, 2001.

[2] H. Zhou and T. S. Huang, “Tracking articulated hand motion with Eigen dynamics analysis,” in Proc. of International Conference on Computer Vision, vol. 2, 2003, pp. 1102–1109.

[3]T. Starner & A. Pentland, "Real-time American sign language recognition from video using hidden markov models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375,1995.

[4] L. Bretzner, I. Laptev, & T. Lindeberg, "Hand gesture recognition using multiscale color features, hieracrchichal models and particle _ltering", in Proc. Int. Conf. Autom. Face Gesture Recog., Washington, DC.May 2002.

[5] A. Argyros & M. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse", in Proc. Workshop Comput. Human Interact., pp.4051,2006.

[6]Q.Chen, N. D. Georganas, “Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar”, *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, VOL. 57, NO. 8, AUGUST 2008.

[7]G. R. S. Murthy, R. S. Jadon, Hand Gesture Recognition using Neural Networks", IEEE Trans., Vol. 6, 2010.

[8]N. A. Ibraheem, RafiqulZaman Khan, 2012, Survey on Various Gesture Recognition Technologies and Techniques, International Journal of Computer Applications, Volume 50, No. 7.

[9]D. G. Lowe, “Distinctive image features from scale-invariant key points,” *Int. J. Comput. Vis* vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] N. H. Dardas and N. D. Georganas, “Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques,” *IEEE Transactions on Instrumentation and Measurement*, VOL.60. NO. 11, November 2011.

[11]S. Pandita1, S. P. Narote “Hand Gesture Recognition using SIFT” *International Journal of Engineering Research & Technology (IJERT)* Vol. 2 Issue 1, January- 2013.

[12]G. R. S. and Jadon, R. S. A Review of Vision Based Hand Gestures Recognition. Int. J. of Information Technology and Knowledge Management,2(2) (2009), 405-410.

[13]R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” *in Proc.IEEE Int. Conf. Image Process.*,2002, vol. 1, pp.900-903.