Abstract

Indian Sign Language (ISL) is a language of the hearing-impaired people of India. Around 10 million people use it to communicate among themselves. However, the problem becomes more pronounced when hearing impaired people want to communicate with normal people. In real world, this problem is mostly solved using a human interpreter. It is highly impossible to pair a normal human with knowledge of sign language with hearing impaired person at all times. Hence a Machine interpreter is the need of the hour which can convert sign language video data to speech or text and vice versa.

This thesis focuses on converting continuous Indian sign language data set into voice or text. The signs are from Indian sign language dataset. A video database of a continuous set of Indian sign language with 58 words having 288 samples is created with simple backgrounds with 6 different signers. The process was initially started with static sign in images and migrates to discrete sign videos and then to a single continuous sign video. The first chapter introduces sign language and gives a detailed account of the state of art research on Sign language recognition.

In the second chapter of the thesis, ISL with images is confronted to classification. Both texture and shape features are extracted from the sign images. Gabor transform and active contours methods extract features of texture and shape respectively. Artificial neural network (ANN) based classifier with backpropagation algorithm has been employed to recognize the signs from their texture and shape features. A total of 360 samples were used for testing with a recognition rate of 98.61%.

The concept of sign recognition is extended to discrete videos in the Chapter 3. Due to computational time constraint on the execution speed of ISL recognition algorithm, a simpler algorithm is developed in this chapter. Video frames are treated with morphological difference
features and canny edge features, which are fused into one using wavelet based fusion methods. This eliminates the blurring introduced by the moving signer’s hands. Elliptical Fourier descriptors (EFD) features describe the large data with few descriptors without loss of shape feature data. To further decrease the size, principal component analysis (PCA) is applied to extract the principal shape features for a video sign. ANN is trained with these features and tested for 80 different signs from ISL with a recognition rate of 92.34%. This method of approach on continuous sign videos brings down the recognition rate due to a large number of frames involved.

The focus of 4th chapter is on recognizing signs in a continuous sign language (CSL) set-up. 58 sign continuous sign video is recorded with a simple background. The problem is in identifying signs that have similar hand movement and distinguishing the signs in CSL model.

In order to identify the shape an active contour model and further tracking is introduced with optical flow model. To make the active contour extraction faster, the initial contour is chosen close to head of the signer with fixed iterations of 25. This enabled the active contour model works for 8 to 12 frames per second depending on the hand’s closeness to the head. Shape and tracking features of signs are used to train and test the classifier. ANN with backpropagation algorithm is used as classifier and obtained a recognition rate of 81.48%.

To further reduce the training time and data further, Fuzzy Inference System (FIS) is proposed as classifier in chapter 5. FIS classified the same data at 82.2%. Chapter 6 compares the performance of ANN and FIS classifiers based on training and testing data with a focus on execution time. The experimentation showed that FIS requires less data for recognition and hence training is faster. But
ANN gives accuracy during testing when trained with more data samples.

The last chapter introduces a 4-camera model for ISL. The disadvantage for data loss due to blurring, hand occlusions in videos during movement is reconstructed using 4 cameras placed at different angles to the signer. Elliptical Fourier Descriptors (EFD) for shape features are used to counteract for huge data produced from 4 cameras. EFD reduced the feature shape data, which is used to train and test the ANN. The recognition increase considerably to 95.1%, due to the presence of multiple angle images for training.

In conclusion, ISL can be transformed to machines in near future thereby removing the need for human interpreter. This thesis presents challenges and solution algorithms to improve the machine translation of Indian sign language into text or speech.