2. REVIEW OF LITERATURE

The ability to detect a person’s unconstrained hand in a natural video sequence has applications in sign language recognition and human computer interaction. Since sign language is gesticulated fluently and interactively like other spoken languages, a sign language recognizer must be able to recognize continuous sign vocabularies in real-time. A system has been build for Indian Sign Language in which gestures are usually understood as hand and body movement which can pass information from one to another. Since the concentration of research work is on hand gestures, the term ‘gesture’ is always referred to the hand gesture in this thesis.

2.1 Related Works

The use of gestures as means to convey information is an important part of human communication. The automatic recognition of gestures enriches human-computer interaction by offering a natural and intuitive method of data input. Automated systems for aiding those with impaired hearing have recently been one of the major areas of research. In sign language, hand gesture is one of the typical methods of non-verbal communication for human beings and various gestures are used to express the own intentions in everyday life. Study has been made to formulate an initial understanding and to identify the next step of functional development. Hand detection is a fundamental step in many practical applications as gesture recognition, video surveillance, and multimodal machine interface and so on. For the flexibility and useful features, vision based technique has been proposed for gesture data collection. This chapter will present research progress and findings on techniques and algorithms for hand detection as it will be used as an input for gesture recognition process for Indian Sign Language.

2.1.1 Vision-based sign language recognition

Most attempts to detect hands from video have restrictions on the environment. For examples, skin colour is surprisingly uniform, so colour-based hand detection is
possible. However, this by itself is not a reliable modality. Hands have to be distinguished from other skin-coloured objects and these are cases of sufficient lighting conditions, such as coloured light or grey-level images. Motion flow information is another modality that can fill this gap under certain conditions, but example for nonstationary cameras this approach becomes increasingly difficult and less reliable. Statistical information about hand locations is effective when used as a prior probability, but it requires application-specific training.


Feng-Sheng Chen, et al., (2003) introduced a hand gesture recognition system to recognize ‘dynamic gesture’ of which a gesture was performed singly in complex background using 2D video input in which the system tracks the moving hand and analyzes the hand-shape variation and motion information as the input to the HMM-based recognition system. They come out with the experimental result of 4-state HMM that has proved to generate the best performance for modelling the gesture. Each input image sequence is pre-processed by hand region extraction process for contour information and coding has been implemented either for contour information only or combination of both contour and motion information. The extracted information is converted to vector sequences and then quantized into symbol sequences for both training and recognition processes. Around 1200 image sequences are collected for 20 different gestures, each kind of gesture with 60 sequences in average, in training phase and other 1200 sequences are collected for test. For method (1), recognition rate of using training data for testing is 97%, and the recognition rate of using testing data is 90.5%, meanwhile for method (2), recognition rate of using training data for testing is 98.5% and while the recognition rate of using testing data rises to 93.5%.
Eng-Jon Ong and Richard Bowden (2004) presented a novel, unsupervised approach to train an efficient and robust detector which is applicable for only detecting the presence of human hands within an image and also classifying the hand shape. Their approach is to detect the location of the hands used as a boosted cascade of classifiers to detect shape alone in grey scale image. Accordingly database of hand images was clustered into sets of similar looking hands by applying the k-mediod clustering algorithm using a distance metric based on shape context. A tree of boosted hand detectors was then formed, consisting of two layers, the top layer for general hand detection, whilst branches in the specialized second layer in classifying the sets of hand shapes resulting from the unsupervised clustering method. When tested the detector with an unseen database of 2509 images, the success rate is 99.8% and with shape classifier the success rate is 97.4%.

Kolsch, M. and Turk, M. (2004) did a study on view-specific hand posture detection with an object recognition method recently proposed by Viola and James. First, they demonstrated the suitability of the integral image approach to the task of detecting hand appearances. Then, the qualitative measure was presented that amounts to an a priori estimate of ‘detectability’, alleviating the need for compute intensive training. Finally, parameters of the detection methods were optimized, achieving significant speed and accuracy improvements. They suggested that most convex appearances with internal grey-level variations are better suited for the purpose of detection with rectangle feature-classification method.

Henrik Jonsson (2008) has carried out investigation to determine the feasibility of creating a system to automatically recognise Swedish sign language based on visual information. The aim of tracking gestures was not realised during the project, as it was focused on segmenting hands in difficult lighting conditions. This involved studying and describing the theoretical backgrounds of the visual processing area and implementing a number of methods in software. Under certain assumptions skin colour models can be constructed automatically and used to extract the shape of a hand with good accuracy.
Tirthankar Dasgupta, et al., (2008) did an initial work on multilingual multimedia Indian Sign Language dictionary tool. “They presented a cross platform multilingual multimedia Indian Sign Language (ISL) dictionary building tool.ISL is a linguistically under-investigated language with no source of well documented electronic data and also research on ISL linguistics gets hindered due to lack of ISL knowledge and unavailability of any educational tools. The proposed system can be used to associate signs corresponding to a given text. The current system facilitates the phonological annotation of Indian signs in the form of HamNoSys structure. The generated HamNoSys string can be given as input to an avatar module to produce an animated sign representation”.

2.1.2 Existing system

Sushmita Mitra and Tinku Acharya (2007) have done a survey on gesture recognition. “Gesture recognition pertains to recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head, and/or body. It is of utmost importance in designing an intelligent and efficient human–computer interface. The applications of gesture recognition are manifold, ranging from sign language through medical rehabilitation to virtual reality. In this paper, they provided a survey on gesture recognition with particular emphasis on hand gestures and facial expressions. Applications involving hidden Markov models, particle filtering and condensation, finite-state machines, optical flow, skin color, and connectionist models are discussed in detail. Existing challenges and future research possibilities are also highlighted in their survey”.

Maryam Pahlevanzadeh, et. al., (2007) have worked on Sign Language Recognition using image processing algorithms. The interpretation has been done on Taiwanese Sign Language, which is one of the sign languages used by the majority of the deaf community. “The process involves two layer classifications. At first, coarse classification is done according to detection of hand motion and tracking the hand location and second classification is based on key frame selection and hand shape recognition of key frames. Motion history image and Fourier descriptor are used for
motion direction recognition and key frame selection respectively. Generic cosine descriptor (GCD) has been proposed for feature extraction of hand postures. GCD is invariant to scale, translation and rotation of hand shapes. The system has been tested using 15 different hand gestures of 10 persons. The experimental results show that the system can achieve 100% recognition rate for test persons”.

Below is the table 2.1 constructed for exploring the past work in sign language recognition system.

Table 2.1: Table showing existing gesture recognition systems found during research

<table>
<thead>
<tr>
<th>Paper</th>
<th>Primary method of recognition</th>
<th>Number of gestures recognised</th>
<th>Background To gesture Images</th>
<th>Additional Markers Required</th>
<th>Number of training images</th>
<th>Accuracy</th>
<th>Frame rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Bauer&amp;Hienz,2000]</td>
<td>Hidden Markov Models</td>
<td>97</td>
<td>General</td>
<td>Multi – colored gloves</td>
<td>7-hours signing</td>
<td>91.7%</td>
<td>-</td>
</tr>
<tr>
<td>[Starner,Weaver &amp; Pentland]</td>
<td>Hidden Markov Models</td>
<td>40</td>
<td>General</td>
<td>No</td>
<td>400 Training sentences</td>
<td>97.6%</td>
<td>10</td>
</tr>
<tr>
<td>[Bowden &amp; Sarhadi,2000]</td>
<td>Linear approximation to non-linear point distribution</td>
<td>26</td>
<td>Blue screen</td>
<td>No</td>
<td>7441 images</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[Davis &amp; Shah,1994 ]</td>
<td>Finite state machine model matching</td>
<td>7</td>
<td>Static</td>
<td>Markers on glove</td>
<td>10 Sequence s of 200 Frames</td>
<td>98%</td>
<td>10</td>
</tr>
</tbody>
</table>

Chetan, A. Burande et al., (2010) has extensively worked in advanced recognition techniques for human computer interaction. “Hand gestures are an important modality for Human Computer Interaction (HCI). Compared to many existing interfaces, hand gestures have the advantages of being easy to use, natural, and intuitive. Successful applications of hand gesture recognition include computer games control, human-robot interaction, and sign language recognition, to name a few. Vision-based recognition systems can give computers the capability of understanding and responding to hand gestures. The aim of this technique is real time vision system for its application within visual interaction environments through hand gesture recognition, using general-purpose hardware and low cost sensors, like a
simple personal computer and USB web cam, so that any user can make use of it in
his office or home. The basis of the approach is a fast segmentation process to obtain
the moving hand from the whole image, which will be able to deal with a large
number of hand shapes against different backgrounds and lighting conditions, and a
recognition process that identifies the hand posture from the temporal sequence of
segmented hands. Three stages of the proposed algorithm are based on a new hand
tracking technique to recognize the actual beginning of a gesture using a Kalman
filtering process, hidden Markov models and graph matching. Processing time is
important in working with large databases”.

Helen Cooper, et al., (2011) have explored intensive research in Sign Language
Recognition. “They have covered the key aspects of Sign-Language Recognition
(SLR), starting with a brief introduction to the motivations and requirements, followed
by a précis of sign linguistics and their impact on the field. The types of data available
and the relative merits are explored allowing examination of the features which can be
extracted. Classifying the manual aspects of sign (similar to gestures) it is then
discussed from a tracking and non-tracking viewpoint before summarising some of the
approaches to the non-manual aspects of sign languages. Methods for combining the
sign classification results into full SLR are given showing the progression towards
speech recognition techniques and further adaptations required for the sign specific
case. Finally the current frontiers are discussed and the recent research presented. This
covers the task of continuous sign recognition, the work towards true signer
independence, how to effectively combine the different modalities of sign, making use
of the current linguistic research and adapting to more noisy data sets”.

Shujjat Khan, et al., (2011) has done a preliminary investigation on sign
language analysis and recognition. “In their review, they have analysed the basic
components of sign language and examined several techniques which are helpful to
design a large vocabulary recognition system for a sign language. The main focus of
this research is to highlight the significance of unaddressed issues, their associated
challenges and possible solutions over a wide technology spectrum”. 

26
2.2 Preprocessing

Parker, J. et al., (1983) have compared interpolating methods for image resampling. When resampling an image to a new set of coordinates, there is often a noticeable loss in image quality. To preserve image quality, the interpolating function used for the resampling should be an ideal low-pass filter. To determine which limited extent convolving functions would provide the best interpolation, five functions are compared in their work. The best response is obtained with the high-resolution cubic spline functions. “The location of the resampled points with respect to the initial co-ordinate system has a dramatic effect on the response of the sampled interpolating function and the data are exactly reproduced when the points are aligned”. The response has the most smoothing effect when the resampled points are equidistant from the original co-ordinate points. “Thus, at the expense of more computing time, image quality has been improved by cubic spline function when compared to NN, Linear or cubic B-spline functions”.

Ruikang Yang, et al., (1995) have performed optimal weighted median filtering under structural constraints. A new expression for the output moments of weighted median filtered data is derived. The noise attenuation capability of a weighted median filter can now be assessed using the L-vector and AI-vector parameters in the new expression. The second major contribution of the work is the development of a new optimality theory for weighted median filters. This theory is based on the new expression for the output moments, and combines the noise attenuation and some structural constraints on the filter’s behavior. In certain special cases, the optimal weighted median filter can be obtained by merely solving a set of linear inequalities. In some cases this leads to closed form of solutions for optimal weighted median filters. Some applications of the theory developed in this paper, in 1-D signal processing are also discussed.

Bax Smith (2002) in his EN9821 Design Assignment explored about the various interpolation techniques for image processing applications. Image processing
applications of interpolation include image magnification or reduction, subpixel image registration, to correct spatial distortions, and image decompression, as well as others. Of the many image interpolation techniques available, nearest neighbour, bilinear and cubic convolution are the most common, and the author has experimented the same. In his implementation of Bicubic Convolution Interpolation he has arrived at the best results in terms of image quality, but took the greatest amount of processing time. Finally, Keys implementation has been shown to be equivalent to that of Matlab.

Thomas Heseltine, et al., (2002) have done evaluation of image pre-processing techniques for eigenface based face recognition. “They have presented a range of image processing techniques as potential pre-processing steps, which attempt to improve the performance of the eigenface method of face recognition. Verification tests are carried out by applying thresholds to gather false acceptance rate (FAR) and false rejection rate (FRR) which results from a data set comprised of images that present typical difficulties when attempting recognition, such as strong variations in lighting direction and intensity, partially covered faces and changes in facial expression. Results are compared using the equal error rate (EER), which is the error rate when FAR is equal to FRR. The work determines the most successful methods of image processing used with eigenface based face recognition, in application areas such as security, surveillance, data compression and archive searching”.

Muhammad Sajjad, et al., (2007) have done work on image magnification using adaptive interpolation by pixel level data-dependent geometrical shapes. “One common difficulty addressed in this work is the magnification techniques of preserving details, i.e. edges and at the same time smoothing the data for not introducing the spurious artefacts. A definitive solution to this is still an open issue. In this work, an image magnification using adaptive interpolation by pixel level data-dependent geometrical shapes is proposed that tries to take into account information about the edges (sharp luminance variations) and smoothness of the image. It calculates threshold, classifies interpolation region in the form of geometrical shapes and then assigns suitable values inside interpolation region to the undefined pixels
while preserving the sharp luminance variations and smoothness simultaneously. The results of proposed technique have been compared qualitatively and quantitatively with five other techniques. The qualitative results show that the proposed method beats completely the Nearest Neighbouring (NN), bilinear(BL) and bicubic(BC) interpolations. The quantitative results are competitive and consistent with NN, BL, BC and others”.

Chi Chang-Yanab, et al., (2008) have done a study on methods of noise reduction in a stripped image. Through their analysis they have found out by image spectrum that its difference can help us to choose different methods to do noise reduction while the information of the image is reduced to be the least. They have also illustrated some methods of noise reduction and taken one test image as an example. Since this image is affected by notable striping, the noise reduction methods of this stripped image are being mainly studied. “The Gray Value Substitution and Wavelet Transformation are satisfactory in stripped noise reduction. Then, MSR and PSNR are calculated to evaluate the processed image. Results suggest that the methods used are suitable in processing this noise”.

Mariusz Leszczyński (2010) has worked on image preprocessing for illumination invariant face verification. “Performance of the face verification system depends on many conditions. One of the most problematic conditions is varying illumination condition. They have compared 14 normalization algorithms based on histogram normalization, illumination properties and the human perception theory using 3 verification methods. The results obtained from the experiments shows that the illumination preprocessing methods significantly improve the verification rate and it is a very important step in face verification system”.

Mythili, C. and V. Kavitha (2011) have done good study on various types of noise in images especially in digital color image. “Noise can occur during image capture, transmission, etc. Noise removal is an important task in image processing. In general, the results of the noise removal have a strong influence on the quality of the
image processing technique. Several techniques for noise removal are well established in color image processing. The nature of the noise removal problem depends on the type of the noise corrupting the image. In the field of image noise reduction several linear and non linear filtering methods have been proposed. Linear filters are not able to effectively eliminate impulse noise as they have a tendency to blur the edges of an image. On the other hand non linear filters are suited for dealing with impulse noise. Several non linear filters based on classical and fuzzy techniques have emerged in the past few years, for example most classical filters that remove and simultaneously blur the edges, while fuzzy filters have the ability to combine edge preservation and smoothing. Compared to other non linear techniques, fuzzy filters are able to represent knowledge in a comprehensible way. In this work results are presented with different filtering techniques and comparison are made with these techniques”.

Geoffrine Judith.M.C and Kumarasabapathy, N. (2011) have done a study on analysis of impulse noise reduction filters. “In their work, a new decision based on median filtering algorithm is presented for the removal of impulse noise from digital images. They have replaced impulse noise corrupted pixel by the median of the pixel scanned in four directions. The signal restoration scheme of this filter adapts to the varied impulse noise ratios while determining an appropriate signal restorer from a reliable neighbourhood. The experimental results of this filter applied on various images corrupted with almost all ratios of impulse noise favour the filter in terms of objectivity and subjectivity than many other prominent impulse noise filters”.

2.3 Edge detection

Philip L. Worthington (2002) has proposed enhanced canny edge detection using curvature consistency. “Edges are often considered as primary image artifacts for extraction by low-level processing techniques, and the starting point for many computer vision techniques. As a result, reliable edge detection has long been a research goal. This author has described initial investigations into recovering reliable edges using curvature models. Canny’s edge detector has been enhanced by adjusting the gradient finding the zero crossings in those directions”.

30
Ravikiran, J. et al., (2009) have worked on finger detection for Sign Language Recognition. They have introduced an efficient and fast algorithm for identification of the number of fingers opened in a gesture representing an alphabet of the American Sign Language. Finger Detection based on the concept of boundary tracing and finger tip detection is accomplished. The system does not require the hand to be perfectly aligned to the camera or use any special markers or input gloves on the hand. The Canny algorithm implemented uses an optimal edge detector based on a set of criteria which include finding the most edges by minimizing the error rate, marking edges as closely as possible to the actual edges to maximize localization, and marking edges only once when a single edge exists for minimal response.

Wenshuo Gao, et al., (2010) have proposed an improved sobel edge detection. Combining sobel edge detection operator and soft-threshold wavelet de-noising is done for edge detection on images which include white Gaussian noises. “In recent years, a lot of edge detection methods have been proposed. The commonly used methods which combine mean de-noising and Sobel operator or median filtering and can not remove salt and pepper noise very well. Here, the soft-threshold wavelet is to remove noise, then use Sobel edge detection operator to do edge detection on the image. This method is mainly used on the images which includes white Gaussian noises using the pictures obtained by the experiment compared to the traditional edge detection methods. The method proposed in this paper has more obvious effect on edge detection”.

2.4 Segmentation

Mie Sato et al., (2000) have experimented gradient magnitude based region growing algorithm for accurate segmentation. “They express that the undesirable partial-volume –effect which lies on the boundary between a high intensity region and a low intensity region, makes unerring boundary determination a difficult task. A new approach to segmentation is required for removing the adverse effect on the boundary, which is unwanted especially from the point of view of volume rendering. A gradient is useful for enhancing the boundary because it emphasizes on the difference among
voxel values”. By analysing the gradient magnitude, the sufficient contrast can be seen which must be presented on the boundary region and the contrast is used to increase the accuracy of segmentation method. Experiments are done on the boundary region segmentation.

Stoyan Donchev. (2000) has performed adaptive threshold-gradient method for segmentation of areas and objects of grey scale images. “The segmentation of an image, i.e., the separation of the object from its background is one of the most important procedures in image processing. Two basic types of segmentations exist at present realized with respect to the intensity and to the intensity gradient, and two basic types of segments areas and borders, respectively. The term “area” usually denotes topologically joined regions of the image which have comparatively homogeneous distribution of intensity, while the term “border” relates to zones where the intensity changes sharply, or in other words, zones with greater value of the intensity gradient. Borders may be situated between an object and a background as well as between different regions of the object. One of these two types of segmentation is usually applied for the purposes of image processing. That is why a new adaptive threshold-gradient method is proposed in the paper. This method treats the image as one indivisible structure containing areas and borders. The analysis of this structure gives results of the segmented image”. The application of the method is proposed for various aspects like analysis of tridimensional scenes with arbitrary location of the illuminating source, for coding of the image homogenizing areas, for analysis of printed documents with irregular background and poor quality, for reducing the number of the intensity levels and removing the information redundancy, etc.

Larrabide et al., (2005) have proposed an image segmentation method based on a discrete version of the topological derivative. “Computed tomography (CT) and magnetic resonance imaging (MRI) have been introduced 3D data sets into clinical radiology. 3D data sets provide information for analysis not available in 2D imaging and challenge the traditional 2D viewing and interpretation used in most clinical
environments. Despite the 3D format of CT and MRI, they are largely interpreted and analyzed as individual 2D image slices. One of the most important stages in medical image analysis is segmentation of objects or definition of their contours. Although improving imaging techniques (e.g., contrast agents, biological markers) should facilitate the segmentation process, medical images are relatively difficult to segment for several undesired properties like low signal-to-noise and contrast-to-noise ratios and multiple and discontinuous edges. The authors presented an image segmentation method based on a discrete version of the well established concept of topological derivative. More specifically, the topological derivative is computed for an appropriate functional associated to the image indicating the cost endowed to a specific image segmentation”.

Phung, SL. et al., (2005) have presented a research work based on skin segmentation using color pixel classification, an analysis and comparison. “This work presents a study of three important issues of the color pixel classification approach to skin segmentation: color representation, color quantization, and classification algorithm. Analysis of several representative color spaces using the Bayesian classifier with the histogram technique shows that skin segmentation based on color pixel classification is largely unaffected by the choice of the color space. However, segmentation performance degrades when only chrominance channels are used in classification. Furthermore, they say that color quantization can be as low as 64 bins per channel, although higher histogram sizes give better segmentation performance. The Bayesian classifier with the histogram technique and the multilayer perceptron classifier are found to perform better compared to other tested classifiers, including three piecewise linear classifiers, three unimodal Gaussian classifiers, and a Gaussian mixture classifier”.

Nassir Salman (2006) has put forward a work in image segmentation based on watershed and edge detection techniques. “A combination of K-means, watershed segmentation method, and Difference In Strength (DIS) map was used to perform image segmentation and edge detection tasks. An initial segmentation is obtained
based on K-means clustering technique. Starting from this, two techniques are used; the first is watershed technique with new merging procedures based on mean intensity value to segment the image regions and to detect their boundaries. The second is edge strength technique to obtain an accurate edge maps of images without using watershed method. By the methodology the authors have solved the problem of undesirable oversegmentation results produced by the watershed algorithm, when used directly with raw data images. Also, the edge maps obtained have no broken lines on entire image and the final edge detection result is one closed boundary per actual region in the image”.

Shaojun Liu and Jia Li (2006) presented a paper on automatic medical image segmentation using gradient and intensity combined level set method. This work proposes a new level set based solution for automatic medical image segmentation. “The study of the authors also shows that level set methods using image intensity or gradient information alone cannot generate satisfying segmentation on some complex organic structures, such as lung bronchia or nodules. And thus they investigated the intensity distribution of the organic structures, and proposed a calibrating mechanism to automatically weigh image intensity and gradient information in the level set speed function. The new method proposed tolerates estimation error in intensity distribution and detects object boundaries whose gradient is low. The experimental results show that the proposed method gives stable and accurate segmentation on public lung image data”.

Ivekovic,S. et al., (2008) have proposed PSO for human body pose estimation. A multi-view set of images of a person sitting at a table is acquired and the pose estimated. Reliable and efficient pose estimation from still images represents an important part of more complex algorithms, such as tracking human body pose in a video sequence, where it can be used to automatically initialise the tracker on the first frame. The body pose estimation is formulated as an analysis-by-synthesis optimisation algorithm, where a generic 3-D human body model is used to illustrate the pose and the silhouettes extracted from the images are used as constraints. A
simple test with gradient descent optimisation run from randomly selected initial positions in the search space shows that a more powerful optimisation method is required. The suitability of the Particle Swarm Optimisation (PSO) for solving this problem is investigated and compared its performance with an equivalent algorithm using Simulated Annealing (SA). The tests show that the PSO outperforms the SA in terms of accuracy and consistency of the results, as well as speed of convergence.

Mingxin Zhang et al., (2008) have done research on the image segmentation required for regions-based image processing. A novel segmentation metrics is proposed with the considerations of the implicit region sizes or image complexity. “Experimental results show that the proposed metric could keep consistent with the implicit region, sizes and numbers, while the numbers have directed relations with image segmentation complexity. It is also verified that the proposed scheme can achieve more verdict accuracy than other schemes and can achieve the accuracy of 98.5% at the requirement region size of 50 pixels”.

Liping Zheng et al., (2009) have improved image segmentation based on PSO algorithm. Image segmentation is the base of image 3D reconstruction. “Threshold segmentation is a simple and important method in grayscale image segmentation. Maximum Entropy method is a common threshold segmentation method. This method utilizes only the gray information. In order to adequately utilize spatial information of greyscale image, an improved 2D entropy segmentation method is proposed. This new method is called PSO-SDAIVE algorithm. In this new method, the computation of 2D entropy is improved. Otherwise, Particle Swarm Optimization (PSO) algorithm is used to solve maximum of improved entropy, maximum taken as the optimal image segmentation threshold. In this paper, two head CT images are segmented in experiment when compared with other segmentation methods. Experimental results show that this new method can quickly and accurately obtain segmentation threshold. Otherwise, this method has strong anti-noise capability and saves computation time”.

35
Ravichandran, K.S. and Ananthi, B. (2009) have done experimentation on color skin segmentation using K-means cluster. “Segmentation is one of the essential problems in image processing. The objective of their work is to segment human skin area in the given color image. In this model, skin detection using cluster based technique is built to detect skin areas. The proposed model has been tested on various images and achieved high detection rate. This model was applied on different color images. Around 50 images are taken for consideration. The color images contained skin region of various parts like hand, face, leg and areas with color similar to that of skin. Also tested image contains more than one skin regions. From this, the proposed model has obtained a good result to segment skin pixels in given color images”.

Soumya Dutta and Bidyut B. Chaudhuri (2009) have suggested an unsupervised color image segmentation method in their paper work based on Homogenous Region. The authors say that Color image segmentation is important but still it is an open problem in image processing. “At first, FCM is applied to the image and the cluster centers are obtained. Quite similar to the famous TSK fuzzy control model, they have formed several rules (IF-THEN like) for pixel classification. The results obtained from the rules are plotted as a histogram. An effective histogram peak detection and valley extraction (PDVE) algorithm are applied to the histogram and thresholds are extracted from the histogram for segmentation. The method is unsupervised and no prior knowledge of number of regions to be segmented is required. The experimental results show that the proposed approach can find homogeneous areas effectively with high accuracy”.

Dietenbeck, T. et al., (2010) have created free software for the evaluation of image segmentation algorithms based on level-set. “The platform gives access to the implementation of six level-set methods that have been chosen in order to cover a wide range of data attachment terms (contour, region and localized approaches). The software also gives the possibility to compare the performance of the proposed algorithms on any kind of images. The performance is evaluated visually and through similarity measurements between a reference and the results of the segmentation”.
Machado, D.A. et al., (2010) proposed a segmentation approach based on topological derivative and level set. “They proposed a segmentation approach that applies the topological derivative as a preprocessing step. The obtained result is used for initializing a level set model in order to get the final result. First, the method uses a low-pass filter and the topological derivative to get a rough definition of the boundaries of interest. Then, morphological operators are applied to fill holes and discard artifacts. Finally, a level set model is used to improve the result giving the desired approximation”. The experiments are carried out with pipeline for cell image segmentation. The method applies prior knowledge about the scale of the objects of interest. The result shows that the technique is robust against noise and very powerful for multi-object segmentation.

Bruno Brandoli Machado and Souza (2011) have presented a paper on the comparison between two image segmentation approaches based on background subtraction and supervised learning. “Real images from two important issues, which have been studied by several computer vision research groups, were used in the experiments: namely, sign language interpretation and mouse behaviour classification. According to performance measures, such as accurate rate, Jaccard coefficient, Yule coefficient, relative area error, and misclassification error, best results were obtained by background subtraction segmentators using images with complex background, otherwise, segmentation based on support vector machines outperformed when simple background were used”.

2.5 Feature extraction

Nobuhiko Tanibata et al., (2002) have proposed a method for extraction of hand features for recognition of Japanese sign language words. For tracking the face and hands, they have initialized the face and hand regions by matching the initial pose template, to decide the range of skin color at the first frame. Tracking is done by overlapping face and hands by matching the texture template of the previous face and
hands. Six features of the face and hands for recognition by the HMM is taken. The system could recognize 65 JSL words in the experiment with real images in a complex background.

Noor Saliza Mohd Salleh et al., (2006) have developed a system for sign language to voice recognition using hand detection techniques for vision-based approach. The use of gestures as means to convey information is an important part of human communication. The automatic recognition of gestures enriches human-computer interaction by offering a natural and intuitive method of data input. Hand detection is a primary step in many practical applications as gesture recognition, video surveillance, and multimodal machine interface and so on. Vision based technique has been proposed for gesture data collection. This work presents a research progress and findings on techniques and algorithms for hand detection as it will be used as an input for gesture recognition process. The features of the processed image are now ready to be an input into the recognition phase. HMM has been proposed to be used for the recognition technique.

Morteza Zahedi et al., (2008) have explored SLR using different aspects of the signings for appearance-based sign language. Sign language includes different components of visual actions made by the signer using the hands, the face, and the torso, to convey his/her meaning. To use different aspects of signs, different groups of features are combined which have been extracted from the image frames recorded directly by a stationary camera. The features are combined in two levels by employing three techniques. At the feature level, an early feature combination can be performed by concatenating and weighting different feature groups, or by concatenating feature groups over time and using LDA to choose the most discriminant elements. At the model level, a late fusion of differently trained models can be carried out by a log-linear model combination. These three combinations of techniques are investigated in an automatic sign language recognition system and show that the recognition rate can be significantly improved.
Ulrich von Agris et al., (2008) have explored the significance of facial features for automatic Sign Language Recognition. “Although facial features are considered to be essential for humans to understand sign language, no prior research work has yet examined their significance for automatic sign language recognition or presented some evaluation results. This paper describes a vision-based recognition system that employs both manual and facial features, extracted from the same input image. For facial feature extraction, an active appearance model is applied to identify areas of interest such as the eyes and mouth region. Afterwards a numerical description of facial expression and lip outline is computed. An extensive evaluation was performed on a new sign language corpus, which contains continuous articulations of 25 native signers. The obtained results proved the importance of integrating facial expressions into the classification process. The recognition rates for isolated and continuous signing increased in signer-dependent as well as in signer independent operation mode”.

Mahmoud Elmezain et al., (2009) have experimented combined features for hand gesture recognition. Hand gesture is an active area of research in the vision community, mainly for the purpose of sign language recognition and Human Computer Interaction. “In this work, a system is proposed to recognize alphabet characters (A-Z) and numbers (0-9) in real-time from stereo color image sequences using Hidden Markov Models (HMMs)”. The system is based on three main stages: automatic segmentation and preprocessing of the hand regions, feature extraction and classification. In automatic segmentation and preprocessing stage, color and 3D depth map are used to detect hands where the hand trajectory will take place in further step using Mean-shift algorithm and Kalman filter. In the feature extraction stage, 3D combined features of location, orientation and velocity with respect to Cartesian systems are used. And then, k-means clustering is employed for HMMs codeword. In the final stage so-called classification, Baum- Welch algorithm is used to do a full train for HMMs parameters. The gesture of alphabets and numbers is recognized using Left-Right Banded model in conjunction with Viterbi algorithm. “Experimental results
demonstrate that, the system can successfully recognize hand gestures with 98.33% recognition rate”.

Table 2.2 represents the current state of the art of video based sign language recognition system. The compilation reveals that most existing recognition systems exploit manual features only. (Ulrich von Agris. et al., 2008).

Table 2.2 Selected video-based sign language recognition systems representing the current state of the art

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Features</th>
<th>Language Level</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vogler</td>
<td>1999</td>
<td>Manual</td>
<td>Sentence</td>
<td>ASL</td>
</tr>
<tr>
<td>Hienz</td>
<td>2000</td>
<td>Manual</td>
<td>Sentence</td>
<td>DGS</td>
</tr>
<tr>
<td>Yang</td>
<td>2002</td>
<td>Manual</td>
<td>Word</td>
<td>ASL</td>
</tr>
<tr>
<td>Zahedi</td>
<td>2007</td>
<td>Manual</td>
<td>Sentence</td>
<td>ASL</td>
</tr>
<tr>
<td>V.Agris</td>
<td>2007</td>
<td>Manual</td>
<td>Sentence</td>
<td>DGS</td>
</tr>
<tr>
<td>Parashar</td>
<td>2003</td>
<td>Manual/Facial</td>
<td>Sentence</td>
<td>ASL</td>
</tr>
</tbody>
</table>

Bhuyan, M.K. et al., (2011) have done specific empirical study on using geometric features for hand pose recognition. “A novel approach for hand pose recognition is done by using key geometrical features of hand. A skeletal hand model is constructed to analyze the abduction and adduction movements of the fingers and these variations are modeled by multidimensional probabilistic distributions. For recognizing hand poses, proximity measures are computed between input gestures and pre-modeled gesture patterns. The proposed algorithm is more robust to the improper hand segmentation and side movements of fingers. Experimental results show that the proposed method is very much suitable for the applications related to Human Computer Interactions (HCI)”.
Reena Rose, R and Suruliandi, A. (2011) have improved the performance of texture based face recognition systems by segmenting face region. “Textures play an important role in recognition of images. This work investigates the efficiency of performance of three texture based feature extraction methods for face recognition. The methods for comparative study are Grey Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP) and Elliptical Local Binary Template (ELBT). Experiments were conducted on a facial expression database, Japanese Female Facial Expression (JAFFE). With all facial expressions LBP with 16 vicinity pixels is found to be a better face recognition method among the tested methods. Experimental results show that classification based on segmenting face region improves recognition accuracy”.

2.6 Classification

Dymitr Ruta and Bogdan Gabrys (2000) have given an overview of classifier fusion methods. “A number of classifier fusion methods have been recently developed by opening an alternative approach leading to a potential improvement in the classification performance. With little fusion theory information, different methods are designed for different problems and producing different results. The investigation of the authors gives an overview of classifier fusion methods and attempts to identify new trends that may dominate this area of research in future. A taxonomy of fusion methods trying to bring some order into the existing “pudding of diversities” is also provided”.

Shahabi, C. Kaghazian et al., (2001) have done analysis of haptic data for Sign Language recognition. “They have focused on analysis of a kind of immersidata known as haptic data. They propose to analyze the haptic data acquired from CyberGlove to recognize different static hand signs automatically. The ultimate objective is to understand how to model and store haptic data in a database, for similar types of applications. Several techniques are proposed to analyze subtle changes in hand signs and words (a series of signs). The techniques can recognize the most
important features to distinguish between two letters and several preliminary experiments demonstrate more than 84.66% accuracy in sign recognition for a 10-sign vocabulary”.

Aleem Khalid Alvi et al., (2005) have constructed a Pakistan Sign Language Recognition using statistical template matching. “Sign language recognition has been a topic of research since the first data glove was developed. Many researchers have attempted to recognize sign language through various techniques. However none of them have ventured into the area of Pakistan Sign Language (PSL). The Boltay Haath project aims at recognizing PSL gestures using Statistical Template Matching. The primary input device is the DataGlove5 developed by 5DT. Alternative approaches use camera-based recognition which being sensitive to environmental changes are not always a good choice. This work explains the use of Statistical Template Matching for gesture recognition in Boltay Haath. The system recognizes one handed alphabet signs from PSL”.

Liang-Guo Zhang et al., (2005) have proposed hierarchical voting classification scheme for improving visual Sign Language Recognition. “As one of the important research areas of multimodal interaction, sign language recognition (SLR) has attracted increasing interest. In SLR, especially on medium or large vocabulary, it is usually difficult or impractical to collect enough training data. Hence improving the recognition on the limited training samples is a significant issue. In this paper, a simple but effective hierarchical voting classification (HVC) scheme for improving visual SLR, which makes efficient use of limited training data, is proposed. The key idea of HVC scheme is not the same as Bagging technique. Firstly, it constructs several training sets from the original training set in a combinatorial fashion to generate the corresponding continuous hidden Markov models (CHMM) ensemble. Then, it determines the ensemble output by appropriate local voting strategy. Finally, it obtains the final recognition result by the global voting. Experimental results show
that the HVC scheme outperforms the conventional single CHMM approach in terms of recognition accuracy on the limited training data”.

Benmokhtar, R. et al., (2006) have done effective research on classifier fusion for semantic indexing in video content. “Classifier combination has been investigated as a new research field to improve recognition reliability by taking into account the complementary aspects between classifiers, in particular for automatic semantic-based video content indexing and retrieval. Many combination schemes have been proposed in the literature according to the type of information provided by each classifier as well as their training and adaptation abilities. This work presents an overview of current research in classifier combination and a comparative study of a number of combination methods. A novel training technique called Weighted Ten Folding based on Ten Folding principle is proposed for combining classifier. Experiments are conducted in the framework of the TRECVID 2005 features extraction task that consists in ordering shots with respect to their relevance to a given class”.

Shanableh, T. et al., (2007) experimented, Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic Sign Language. “This work presents various spatio-temporal feature-extraction techniques with applications to online and offline recognitions of isolated Arabic Sign Language gestures. The temporal features of a video-based gesture are extracted through forward, backward, and bidirectional predictions. The prediction errors are thresholded and accumulated into one image that represents the motion of the sequence. The motion representation is then followed by spatial-domain feature extractions. As such, the temporal dependencies are eliminated and the whole video sequence is represented by few coefficients. The linear separability of the extracted features is assessed, and its suitability for both parametric and nonparametric classification techniques is elaborated upon. The proposed feature-extraction scheme was complemented by simple classification techniques, namely, K nearest neighbor (KNN) and Bayesian, i.e., likelihood ratio, classifiers. Experimental results showed classification performances ranging from 97% to 100% recognition rates. To validate the proposed
technique, a series of experiments is conducted using the classical way of classifying data with temporal dependencies, namely, Hidden Markov Models (HMMs). Experimental results revealed that the proposed feature-extraction scheme combined with simple KNN or Bayesian classification yields comparable results to the classical HMM-based scheme. Moreover, since the proposed scheme compresses the motion information of an image sequence into a single image, it allows using simple classification techniques where the temporal dimension is eliminated. This is actually advantageous for both computational and storage requirements of the classifier”.

Li Qin Fu et al. (2010) have proposed a work on classifier fusion for speech emotion recognition. “According to multidimensional emotion space model, an improved queuing voting algorithm was proposed to implement the fusion among multiple emotion classifiers for a good emotion recognition result. Firstly, three kinds of classifiers were designed based on Hidden Markov Model (HMM) and Artificial Neural Network (ANN). Then, the improved queuing voting algorithm was used to fuse them. Experimental study had been carried out using Beihang University mandarin emotion speech database and Berlin database of emotional speech respectively. The results proved that the improved queuing voting algorithm can attain better fusion effect than conventional fusion algorithm and excel any single classifier evidently”.

Mahmoud M, Zaki. and Samir, I. Shaheen (2011) have proposed a sign language recognition using a combination of new vision based features. “Sign languages are based on four components-hand shape, place of articulation, hand orientation, and movement. This paper presents a novel combination of vision based features in order to enhance the recognition of underlying signs. Three features are selected to be mapped to these four components. Kurtosis position and Principal Component Analysis, PCA and Motion Chain Code, MCC. On the basis of these features a prototype is designed, constructed and its performance is evaluated. It consists of skin color detector, connected component locator and dominant hand tracker, feature extractor and a Hidden Markov Model classifier. The input to the
system is a sign from RWTH-BOSTON-50 Database and the output is the corresponding word with a recognition error rate of 10.90%”.

Sanjay Meena (2011) has presented a thesis on hand gesture recognition technique. “This work presents a technique for a human computer interface through hand gesture recognition that is able to recognize 25 static gestures from the American Sign Language hand alphabet. The objective of this thesis is to develop an algorithm for recognition of hand gestures with reasonable accuracy. The segmentation of gray scale image of a hand gesture is performed using Otsu thresholding algorithm. Otsu algorithm treats any segmentation problem as classification problem. Total image level is divided into two classes, one is hand and the other is background. The optimal threshold value is determined by computing the ratio between class variance and total class variance. A morphological filtering method is used to effectively remove background and object noise in the segmented image. Morphological method consists of dilation, erosion, opening, and closing operation. Canny edge detection technique is used to find the boundary of hand gesture in image. Contour of a gesture is represented by a Localized Contour Sequence (L.C.S) whose samples are the perpendicular distances between the contour pixels and the chord connecting the endpoints of a window centered on the contour pixels. These extracted features are applied as input to classifier. Linear classifier discriminates the images based on dissimilarity between two images. Multi Class Support Vector Machine (MCSVM) and Least Square Support Vector Machine (LSSVM) are also implemented for the classification purpose. Experimental result shows that 94.2% recognition accuracy is achieved by using linear classifier and 98.6% recognition accuracy is achieved using Multiclass Support Vector machine classifier. Least Square Support Vector Machine (LSSVM) classifier is also used for classification purpose and shows 99.2% recognition accuracy”.

Guan-Wei Wang et al., (2012) have applied of classifier combination methods in hand gesture recognition. “Hand gesture recognition is a topic in artificial intelligence and computer vision with the goal to automatically interpret human hand
gestures via some algorithms. And hence several classifier combination techniques are employed in this work to handle this specific problem. Based on some related data, AdaBoost and rotation forest are seen to behave significantly better than all the other considered algorithms, especially a classification tree. By investigating the bias-variance decompositions of error for all the compared algorithms, the success of AdaBoost and rotation forest can be attributed to the fact that each of them simultaneously reduces the bias and variance terms of a single tree’s error to a large extent. Meanwhile, kappa-error diagrams are utilized to study the diversity-accuracy patterns of the constructed ensemble classifiers in a visual manner”.

2.7 Recognition

Peter Wray Vamplew (1996) has explored research on recognition of sign language using neural networks. “Artificial neural networks have proved to be an extremely useful approach to pattern classification tasks, but much of the research in this field has concentrated on relatively simple problems. Attempting to apply these networks to a complex real-world problem such as sign language recognition exposed a range of issues affecting this classification technique. The development of the SLARTI system inspired the creation of several new techniques related to neural networks, which have general applicability beyond this particular application. This thesis includes discussion of techniques related to issues such as input encoding, improving network generalisation, training recurrent networks and developing modular, extensible neural systems”.

Guang-Bin Huang et al., (2006) have proposed real-time learning capability of neural networks. “In some practical applications of neural networks, fast response to external events within an extremely short time is highly demanded and expected. However, the extensively used gradient-descent-based learning algorithms obviously cannot satisfy the real-time learning needs in many applications, especially for large-scale applications and/or when higher generalization performance is required. Based on Huang’s constructive network model, this work proposes a simple learning algorithm capable of real-time learning which can automatically select appropriate
values of neural quantizers and *analytically* determine the parameters (weights and bias) of the network at one time only. The performance of the proposed algorithm has been systematically investigated on a large batch of benchmark real-world regression and classification problems. The experimental results demonstrate that the algorithm can not only produce good generalization performance but also have real-time learning and prediction capability. Thus, it may provide an alternative approach for the practical applications of neural networks where real-time learning and prediction implementation are required”.

Huang, G.B. et al., (2006) have done extensive research on Extreme Learning Machine, its theory and applications. “It is clear that the learning speed of feedforward neural networks is in general far slower than required and it has been a major bottleneck in their applications in past decades. Two key reasons behind may be: (1) the slow gradient-based learning algorithms are extensively used to train neural networks, and (2) all the parameters of the networks are tuned iteratively by using such learning algorithms. Unlike these conventional implementations, they have proposed a new learning algorithm called Extreme Learning Machine (ELM) for Single-hidden Layer Feedforward Neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. In theory, this algorithm tends to provide good generalization performance at extremely fast learning speed. The experimental results based on a few artificial and real benchmark function approximation and classification problems including very large complex applications show that the new algorithm can produce good generalization performance in most cases and can learn thousands of times faster than conventional popular learning algorithms for feedforward neural networks”.

Nan-Ying Liang et al., (2006) have proposed a classification of mental tasks from eeg signals using Extreme Learning Machine. “A recently developed machine learning algorithm referred to as Extreme Learning Machine (ELM) is used to classify five mental tasks from different subjects using electroencephalogram (EEG) signals available from a well-known database. Performance of ELM is compared in terms of
training time and classification accuracy with a Backpropagation Neural Network (BPNN) classifier and also Support Vector Machines (SVMs). For SVMs, the comparisons have been made for both 1-against-1 and 1-against-all methods. Results show that ELM needs an order of magnitude less training time compared with SVMs and two orders of magnitude less compared with BPNN. The classification accuracy of ELM is similar to that of SVMs and BPNN. The study shows that smoothing of the classifiers’ outputs can significantly improve their classification accuracies”.

Oya aran (2008) has presented a work on vision based sign language recognition. “This work addresses the problem of vision based sign language recognition and focuses on three main tasks to design improved techniques that increase the performance of sign language recognition systems”. The experiments show that the proposed approach has a robust tracking performance during the challenging situations and is suitable for tracking long durations of signing with its ability of fast recovery. Attack the problem of the recognition of signs that include both manual (hand gestures) and non-manual (head/body gestures) components. Finally, a strategy is proposed to combine generative and discriminative models to increase the sign classification accuracy. The Fisher kernel method is applied and proposed a multi-class classification strategy for gesture and sign sequences. “The results of the experiments show that the classification power of discriminative models and the modeling power of generative models are effectively combined with a suitable multi-class strategy”.

A Sign to Voice system for hand gesture recognition is proposed by Oi Mean Foong. et al., (2008). “It is most commonly used by people who have hearing or speech problems to communicate among themselves or with normal people. Various sign language systems have been developed by manufacturers around the globe but unfortunately they are neither flexible nor cost-effective for the end users. This system prototype is able to automatically recognize sign language to help normal people to communicate more effectively with the hearing or speech impaired people. The Sign to Voice system prototype, S2V, has been developed using Feed Forward Neural
Network for two-sequence signs detection. Different sets of universal hand gestures has been captured from video camera and utilized to train the neural network for classification purpose. The experimental results have shown that neural network has achieved satisfactory result for sign-to-voice translation”.

Iwan Njoto Sandjaja and Nelson Marcos (2009) have proposed Sign Language Number Recognition system that lays down foundation for handshape recognition which addresses real and current problems in signing in the deaf community and leads to practical applications. “The input for the sign language number recognition system is 5000 Filipino Sign Language number video file with 640 x 480 pixels frame size and 15 frame/second. The color-coded gloves uses less color compared with other color-coded gloves in the existing research. The system extracts important features from the video using multi-color tracking algorithm which is faster than existing color tracking algorithm because it did not use recursive technique. Next, the system learns and recognizes the Filipino Sign Language number in training and testing phase using Hidden Markov Model. The system uses Hidden Markov Model (HMM) for training and testing phase. The feature extraction could track 92.3% of all objects. The recognizer also could recognize Filipino sign language number with 85.52% average accuracy”.

Murthy, G.R.S. and Jadon, R.S. (2009) have done hand gesture recognition using neural networks visual interpretation of gestures and it can be useful in accomplishing natural Human Computer Interactions (HCI). They have proposed a method for recognizing hand gestures and designed a system which can identify specific hand gestures and use them to convey information. At any time, a user can exhibit his/her hand doing a specific gesture in front of a web camera linked to a computer. Firstly, the hand gesture of a user is captured and stored on disk. Then those videos captured are read one by one, converted them to binary images and created 3D Euclidian Space of binary values. Supervised feed-forward neural net is used based on training and back propagation algorithm for classifying hand gestures into ten categories: hand pointing up, pointing down, pointing left, pointing right and pointing
front and number of fingers user was showing. The results achieved are up to 89% on the typical test set.

Jane J. Stephan and Sana’a Khudayer (2010) have developed a gesture recognition system for Human-Computer Interaction (HCI). “Considerable effort has been put towards developing intelligent and natural interfaces between users and computer systems. This is done by means of a variety of modes of information (visual, audio, pen, etc.) either used individually or in combination. The use of gestures as means to convey information is an important part of human communication. A new technique for static hand gesture recognition is proposed for the Human-Computer Interaction (HCI) based on shape analysis. The objective of this effort is to explore the utility of a neural network-based approach to the recognition of the hand gestures. The proposed system uses the hand contour as a geometry feature. A unique multi-layer perceptron neural network is built for the classification by using back-propagation learning algorithm. The overall model is designed to be a simple and robust gestural interface prototype for various PC applications”.

Vaishali, S. kulkarni and Lokhande,S.D., (2010) has explored appearance based recognition of American Sign Language using gesture segmentation. “The work presented is to develop a system for automatic translation of static gestures of alphabet in American Sign Language. In doing so three feature extraction methods and neural network is used to recognize signs. The system deals with images of bare hands, which allows the user to interact with the system in a natural way. An image is processed and converted to a feature vector that will be compared with the feature vectors of a training set of signs. The system does rotation, scaling of translation variant of the gesture within the image, which makes the system more flexible. The system is implemented and tested using data sets of number of samples of hand images for each signs. Three feature extraction methods are tested and the best one is suggested with results obtained from ANN. The system is able to recognize selected ASL signs with the accuracy of 92.33%”.
Lorena P. Vargas et al., (2011) have worked on Sign Language Recognition system using neural network for digital hardware implementation. “This work presents an image pattern recognition system using neural network for the identification of sign language to deaf people. The system has several stored images that show the specific symbol in this kind of language, which is employed to teach a multilayer neural network using a back propagation algorithm. Initially, the images are processed to adapt them and to improve the performance of discriminating the network, including the process of filtering, reduction and elimination noise algorithms as well as edge detection. The system is evaluated using the signs without including movement in their representation”.

2.8 Summary

This chapter has elaborately detailed the extensive research work done in ASR and the proposed methodology of ISL recognition and its execution is briefly explained in next chapter, Methodology.