CHAPTER 3

LITERATURE REVIEW

3.1 REVIEW ON IMAGE SEGMENTATION ALGORITHMS

Image segmentation is one of the primary steps in image analysis for object identification. The main aim is to recognise homogeneous regions within an image as distinct and belonging to different objects. Segmentation stage does not worry about the identity of the objects, which can be labelled later. The segmentation process can be based on finding the maximum homogeneity in grey levels within the regions identified.

There are several issues related to image segmentation that require detailed review. One of the common problems encountered in image segmentation is choosing a suitable approach for isolating different objects from the background. The segmentation does not perform well if the grey levels of different objects are quite similar. Image enhancement techniques seek to improve the visual appearance of an image, which emphasize the salient features of the original image and simplify the task of image segmentation. The type of operator chosen has a direct impact on the quality of the resultant image. It is expected that an ideal operator will enhance the boundary differences between the objects and their background making the image segmentation task easier. Issues related to segmentation involve choosing good segmentation algorithms, measuring their performance, and understanding their impact on the scene analysis system.
Numerous low-level processing techniques have been developed for image segmentation. These techniques include thresholding, edge detection, and region growing. These methods are often used in combination with other approaches, such as neural networks and Bayesian inference. A common approach is to use a combination of region growing and thresholding. Other approaches include clustering and graph-based methods. For instance, Forrer et al. (1988) developed three segmentation algorithms to eliminate large amounts of clear wood image data.
variance of the Normal density function to segment defect regions from clear wood regions. The basis of this method assumes that the majority of the image useful on segmenting outdoor colour images. This is based on constructing O’Holleran (1975) has proposed a thresholding technique that is very

3.2.1 Histogram Thresholding

determine the best color channel to use in segmentation. Features are represented by color would provide a basis from which to perform the color channel selected. Prior knowledge of how wood performance of the color channel selected. It is not discussed if a change in species or resolution will affect the features are represented by a particular color channel. Image processing is based on previously used methods for pattern recognition and segmentation is based on previously used methods for pattern recognition and the color channel used for the segmentation of features. Information used for data for classification has been used to compare the performance of from regions of interest, detect regions from clear wood, and the reduction of wood features are represented by color would provide a basis from which to

While these methods describe the segmentation of backgrounds and determination of the colour data, all developments represent defects. Histogram will contain clear wood and half by imposing an estimated variance of the Normal density function to segment defect regions. The basis of this method assumes that the majority of the image
until no separate peaks are found in any of the histograms. The criteria to separate peaks was based on the ratio of peak maximum to peak minimum to be greater than or equal to two. Textured areas were separated from uniform regions by using a Sobel operator marking regions that contain large edge activity (more than 25 edge pixels in a 9x9 pixel window). These were called ‘busy’ areas. The original area was not to segment inside these busy areas based on thresholding but sometimes it is necessary to do so as in the case of segmenting skylines.

Conners et al (1983) utilized histogram thresholding based on gray level histograms generated from an image collected by a colour line scan camera for segmentation. Difficulties were encountered in separating decay and stain from clear wood areas based on black and white histogram information. Later, intensity histograms from all the colour channels were used to improve performance.

The thresholds are selected based on valley and inflection points of the gray level histograms from the entire input image. Once thresholds are found, regions are formed and marked. All pixels having the same gray level value between thresholds of the histogram are marked the same. Connected component labeling is then performed where regions are connected and labeled if unique. Very small regions are eliminated as noise and then properties are extracted from the regions for the high level processing.

Klinkhachorn et al (1993) also used a method of histogram thresholding based on multiple threshold segmentation of black and white gray-level histograms. Thresholding of the entire image is used to separate background, clearwood, and defect regions. The multiple thresholds are obtained with neural networks trained with known inputs using backpropagation. Kothari and Huber (1991) and Lampinen and Smolander (1996) also use neural network based thresholding in a similar manner.
Cheriet et al (1998) have presented a general recursive approach for image segmentation by extending Otsu’s (1978) method. This approach has been implemented in the area of document images, specifically for segmenting bank cheques. This approach segments the brightest homogeneous object from a given image at each recursion, leaving the darkest homogeneous object. This method is developed without any constraints on the number of objects in the digital image. The method is based on discriminant analysis. The thresholding operation is regarded as the partitioning of pixels of an image into two classes: object and background. For each iteration, the histogram of the image is drawn and the largest peak is separated from the rest of the image. The process is continued till there are no more peaks left in the histogram. Similar to Otsu’s method, the proposed method only performs well on the images with two classes. Using the method iteratively one can segment more than two objects. This technique also facilitates the process of information extraction from document images and preserves the topological properties of the extracted information that is used for further recognition. In this study, a set of human visual criterion was defined, where each criterion was given a weight. The system was trained on 220 bank cheques and 505 different cheques were used for testing the performance. Results of the performance analysis indicate that the percentage recognition rate varied between 93%-100% for different sets of test samples. The authors concluded that the method gives good results when the target object is the darkest object in a given image. However, when the target object is not the darkest, the method fails to segment properly.

In a number of applications, histogram thresholding is not possible simply because the histogram may be unimodel. In some cases the images may be of such quality that any pre-processing may not improve the contrast between objects sufficiently and hence one may not achieve two or more peaks in the histogram for selecting thresholds for segmentation. Unimodel
distributions are typically obtained when the image consists of mostly of a large background area with small, but significant regions. This often happens in medical imaging applications. Similarly in aerial scenes with many different objects, the histogram may only have one peak because of the vast range of intensities for each object and an overlap between these. Bhanu and Faugeras (1982) have proposed a gradient relaxation algorithm for solving the above problem and compare it to non-linear probabilistic relaxation algorithm proposed by Rosenfeld (1976). The process is based on assigning each pixel an initial probability of occurrence and then using gradient relaxation based on compatibility function that takes into account the relationship between a pixel and its eight neighbours. The relaxation process is iterative where pixel values are changed so that the histogram is no longer unimodal and thresholds can be easily detected. Compared to Rosenfeld’s method, the authors claim that this technique provides a better control by defining parameters that are under user control. The user can therefore control the degree of smoothing at each iteration as well as the initial assignment of probabilities. An extension of the gradient relaxation approach for image segmentation is presented in Bhanu and Parvin (1987). This new segmentation technique is based on recursive split and merges and has a number of advantages. The method avoids any heuristics and arbitrary measures for partitioning the image. Hence it does not require a robust merging step in order to remove boundary artefacts. In addition, the method does not require detailed peak location and selection procedure.

Li et al (1997) have proposed the use of two dimensional histograms of an image which is more useful for finding thresholds for segmentation rather than just using grey level information in one dimension. In 2D histograms, the information on point pixels as well as the local grey level average of their neighbourhood is used. The authors show that the application of Fisher linear discriminant to the histogram results in an optimal
projection where the data clusters are better defined and hence easier to separate by choosing appropriate thresholds. The experimental results show that the proposed technique has the same computational demands as of one dimensional techniques but gives better segmentation results.

3.2.2 Edge Based Segmentation

Ahuja et al (1980) describe how pixel neighbourhood elements can be used for image segmentation. For each, its neighbours are first identified in a window of fixed size. A vector of these neighbours as individual grey values or vector of average grey levels in windows of size 1x1, 3x3 and 5x5 is determined, which used both vector representations. The aim is to identify a weight matrix that multiplied with these vectors will yield a discriminant value that allows the classification of pixel in one of the two classes. Two different sets of experiments were performed on Forward Looking Infra Red (FLIR) images. The first experiment is performed on 10 images. In the first experiment, two choices of feature vectors are chosen for every pixel. The images were of size 64x64 pixels containing a single tank each. The segmentation was done using the super-slice algorithm. The features were averaged over different sized windows (3x3 pixels and 5x5 pixels) and classification was performed using Fisher’s linear decision rule. In the second experiment, a separate set of 25 images was chosen with light objects on dark background. Ten images were used for training and 12 points were chosen from the background and 12 from the interior of the object for each image. All of these samples were then used to train the classifiers. Testing was done on the remaining set of 15 images with some added noise. The authors stated that, for the noisy data, using the pixel grey level as the only feature, the results obtained were better than those obtained by pixel grey level properties. The authors proposed that the error rate for noisy images seems to depend primarily on number of features used. The authors concluded that the results
suggest that the grey-levels of the pixel and its neighbours are a good set of features to use in pixel classification.

Prager (1980) has proposed a set of algorithms to perform segmentation of natural scenes through boundary analysis. The goal of the algorithm is to locate the boundaries of an object correctly in a scene. First, pre-processing of the images is done to clean up the raw data by smoothing and noise-removal. Second, the edge representation is generated. Differentiation is done to find the edge-strength at each point in the image. Suppression is then done to remove multiple edges formed by spatial differentiation of boundaries. Third, the edges are joined into line segments and features are computed. The features include: length, contrast, frequency, mean, variance and location of each line segment. Fourth, post-processing is done to remove unwanted line segments and to build confidence for each of the remaining segments. The output of the system is a set of line segments with a list of attributes, such as length and confidence. The authors concluded that advantages of this approach are twofold. Firstly, the effects of each stage are well defined. Secondly, any stage can be omitted or replaced by a more sophisticated variant if desired.

Perkins (1980) has used an edge based technique for image segmentation. It is acknowledged that edge based segmentation has not been very successful because of small gaps that allow merging of dissimilar regions. In order to avoid these problems, the paper proposes an expansion-contraction technique in which edge regions are expanded to close gaps and then contracted after the separate regions have been labelled. The size of expansion is controlled such that small regions are not engulfed by this process. The process involves the use of Sobel filter for producing edge strengths and directions at every point. The edges are thinned and the result is automatically thresholded leaving only ridges. The ridges separate regions of
different intensity but there may be small gaps. Segmentation is performed by expanding active edge regions, labelling the segmented uniform intensity regions, and then contracting edge regions. The results are shown for landscape pictures and for pictures of electronic circuits.

Chan et al (1998) have developed a new adaptive thresholding algorithm for image segmentation using variation theory. The method is a heuristic algorithm, which consists of seven steps. First, image smoothing is done using an average filter. Grey-level gradient magnitude is then derived from it. A thresholding and thinning algorithm is applied to the gradient magnitude to find the object boundary points. The image is then sampled at the boundary points as the local thresholds. The threshold surface is interpolated by the sampled local thresholds. Then the image is segmented by the threshold surface. Noise is then removed from the segmented image by a variational method. The method is performed iteratively to segment an image. The iteration process is stopped using an error or iteration time threshold. The authors presented the results on an image with 16 objects and simulated background and also on some handwriting images. The results were found to be more satisfactory when there are a large number of objects in the image as they yield a better thresholding surface. The method is demonstrated successfully on segmenting images for an OCR application.

3.2.3 Tree/Graph based Approaches

Cho and Meer (1997) have proposed a new approach for segmentation, which is derived from the consensus of a set of different segmentation outputs on one input image. Instead of statistics characterising the spatial structure of the local neighbourhood of a pixel, for every pair of adjacent pixels their collected statistics are used for determining local homogeneity. Several initial segmentations are derived from the same input image by changing the probabilistic component of the hierarchical
Region Adjacency Graph (RAG) pyramid based technique. From the ensemble of these initial segmentations, for every adjacent pixel pair a co-occurrence probability is derived, which captures global information (about the image) at the local level (pixel level). The final segmentation of the input image is obtained by processing the co-occurrence probability field with the same RAG pyramid technique. The pixel pairs with high co-occurrence probability are then grouped together based on the consensus about local homogeneity. This technique can also be used to extract the high confidence homogeneous regions from the co-occurrence probability field. Bayesian networks were then used to extract features from images. The features extracted were variance of the width of the region, ratio of average width to length and the average grey level. Then, post-processing of over-segmented images is done based upon a priori information about the sought features. The RAG of the final segmentation provides the spatial relationship between regions and can be used for further interactive analysis of the image. This segmentation method is completely unsupervised. Experiments were performed on an aerial image, and images of boat, pentagon, and house.

Yeung et al (1998) have proposed the technique of segmentation of video by clustering and graph analysis. The method is also extended to the Scene Transition Graph (STG) representation for the analysis of temporal structures extracted from a video. First, a video is taken and then similar shots are recognised to reduce the amount of information to be processed. Then, dissimilarity between two shots is defined based upon the dissimilarity indices for all image pairs in the two shots. Then automatic segmentation of scenes and story units is done. The images are segmented based on the presentation of the featured video sequence along the timeline and the visual contents. The video shots are then clustered together based upon visual similarities of image contents and the temporal dynamics shown in the visual contents within individual shots (termed as time-constrained clustering). Then, a STG is
drawn and analysed. A STG is then used to represent compactly the structures of shots and the temporal flow of the story of videos. Some new attributes are defined from STG representation for the further analysis. It is assumed that two consecutive scenes in a video presentation do not share significant similarities in visual qualities. Cluster labels are given to shots in different scenes. The experiment was performed on a sequence of shots to extract story units. The video was successfully decomposed into a hierarchy of story units, each of which consisted of clusters of similar shots. The methodology also provided the mean of non-linear access to a featured program which in turn facilitated browsing of video contents.

3.2.4 Region Growing

A range of image segmentation algorithms are based on region growing. Some relevant studies are reviewed that have used region growing algorithms. Region growing algorithms take one or more pixels, called seeds, and grow the regions around them based upon a certain homogeneity criteria. If the adjoining pixels are similar to the seed, they are merged with them within a single region. The process continues until all the pixels in the image are assigned to one or more regions.

Chang and Li (1994) have proposed a region-growing framework for image segmentation. This process is guided by regional feature analysis and no parameter tuning or a priori knowledge about the image is required. The algorithm is known as Fast Adaptive Segmentation (FAS) algorithm. The image is first divided into many small primitive regions that are assumed to be homogeneous. These primitive regions are then merged to form larger regions until no more merges are possible. Two regions are merged if they pass the homogeneity test and also if the value of the edge connecting them is weak. The focus of this study is on investigating how different merge criteria affect the quality of segmentation and the processing time. The experiments
designed to evaluate the merge criteria are based on four important aspects of segmentation output: region mergeability, boundary accuracy, merge rejections, and number of iterations required. In these experiments, 300 images of size 50x50 pixels were used. Each image had two equal sized regions that share a simple straight 50-unit boundary. The data was randomly generated for the two regions from a pair of normal distributions having different means and variances. The best results using the fast merge method gives the correct classification rate of 86% (with less than 5 regions in the image). The authors concluded that the algorithm automatically computes segmentation thresholds based on local feature analysis. The algorithm is robust and produces high quality segmentation on a wide range of textured and grey scale images. This framework can also be easily adapted to different image applications by substituting the suitable features. The main limitation of this algorithm is however the limited applicability of the adaptive homogeneity tests on very small regions and order dependency of its segmentation results.

For region growing, seeds can be automatically or manually selected. Their automated selection can be based on finding pixels that are of interest, e.g. the brightest pixel in an infra-red image can serve as a seed pixel. They can also be determined from the peaks found in an image histogram. On the other hand, seeds can also be selected manually for every object present in the image. Adams and Bischof (1994) studied the effectiveness of seeded region growing approach for image segmentation of greyscale images where the seeds are manually selected. The method is employed to segment an image into different regions using a set of seeds. Each seeded region is a connected component comprising of one or more points and is represented by a set S. The set of immediate neighbours bordering the pixel is calculated. The neighbours are then examined and if they intersect any region from set S, then a measure (difference between a pixel and the intersected region) are
computed. If the neighbours intersect more than one region, then the set is taken as that region for which difference measure is maximum. The new state of regions for the set then constitutes input to the next iteration. This process continues until the entire image pixels have been assimilated into regions. Hence, for each iteration the pixel that is most similar to a region that it borders is appended to that region. The SRG algorithm is inherently dependent on the order of processing image pixels. One implication of this algorithm is that raster order processing and anti-raster order processing do not lead to the same tessellation. The algorithm was applied on images with different types of objects in them. The authors concluded that the method has the advantage that it is fairly robust, quick, and parameter free except for its dependency on the order of pixel processing.

Mehnert and Jackway (1997) have improved the above seeded region-growing algorithm by making it independent of the pixel order of processing and making it more parallel. Their study presents a novel technique for Improved Seeded Region Growing (ISRG). ISRG algorithm retains the advantages of Seeded Region Growing (SRG) such as fast execution, robust segmentation and no parameters to tune. The algorithm is also pixel order independent. If more than one pixel in the neighbourhood have same minimum similarity measure value, then all of them are processed in parallel. No pixel can be labelled and no region can be updated until all other pixels with the same priority have been examined. If a pixel cannot be labelled, because it is equally likely belong to two or more adjacent regions, then it is labelled as 'tied' and takes no part in the region growing process. After all of the pixels in the image have been labelled, the pixels labelled 'tied' are independently re-examined to see whether or not the ties can be resolved. To resolve the ties an additional assignment criterion is imposed, such as assigning a tied pixel to the largest neighbouring region or to the neighbouring region with the largest mean. This is a post-processing step. The
algorithm in this study was tested on the image of man made objects such as a car, an aeroplane, and buildings. The authors concluded that ISRG algorithm produces consistent segmentation because it is not dependent on the order of pixel processing. Parallel processing ensures that the pixels with the same priority are processed in the same manner simultaneously.

Basu (1987) has developed general sets of semantics for region detection to describe a number of image models using them. The semantic set is established empirically based on simple and intuitive properties of a region. It can be extended to include new semantics without altering the conceptual and computational framework of the method. In case of region detection, each pixel of a digitised image is chosen as a primitive. An ideal region is obtained from a group of pixels of the given image by approximating the grey-level values of these pixels with a linear or quadratic approximation scheme. A set of attributes is calculated for each pixel of the group belonging to the given image as well as for the each pixel of the ideal region. In particular, four attributes: contrast, total variation, global average grey-level value, and representative grey-level value are used. Distance between the group of pixels and the ideal region is obtained as a numerical measure of dissimilarity between them. Pixel assignment to regions is decided according to the distance measure. The authors concluded that there is a noticeable improvement in these results when semantic information is added. There is no optimal set of attributes that can be used for all of the images. Any image has to be tested with a number of different attribute value combinations in order to obtain the best possible segmentation results. This scheme is found suitable for the class of problems in which fine structural detail of an image is not needed. The author has demonstrated better results as a consequence of using semantic information than other known segmentation methods on natural objects.
Beaulieu and Goldberg (1987) have proposed a hierarchical stepwise optimisation algorithm for region merging which is based on stepwise optimisation and produces a hierarchical decomposition of the image. The algorithm starts with an initial image partition into a number of regions. At each iteration, two segments are merged provided they minimise a criteria of merging a segment to another. In this stepwise optimisation, the algorithm searches the whole image context before merging two segments and finds the optimal pair of segments. This means that the most similar segments are merged first. The algorithm gradually merges the segments and produces a sequence of partitions. The sequence of partitions reflects the hierarchical structure of the image. This is also termed as agglomerative clustering. The authors concluded that the results obtained from the final grouping of the algorithm are far better than the conventional non-hierarchical methods such as k-means clustering. The advantage of this algorithm over non-hierarchical methods is that there is no need to specify the seed points.

In some studies, edge or gradient information has been used in combination with region growing for image segmentation. Gambotto (1993) proposed an algorithm that combines the region growing and edge detection methods for segmenting the images. The method is iterative and uses both of these approaches in parallel. The algorithm starts with an initial set of seeds located inside the true boundary of the region. The pixels that are adjacent to the region are iteratively merged with it if they satisfy a similarity criterion. A second criterion uses the average gradient over the region boundary to stop the growth. The last stage refines the segmentation. The analysis is based on cooperation between the region growing algorithm and the contour detection algorithm. Since, the growing process is performed by adding segments to a region, some pixels which belong to the next region and to the previous region, may be misclassified. A nearest neighbour rule is then used to locally reclassify these pixels. The algorithm was tested on a number of x-ray
images. The authors concluded that the method uses both the statistical model of the region and the average gradient computed over its boundary. The system behaves similar to the snake class of algorithms but it can also be used to segment closed regions having unknown and complex shapes.

Hojjatoleslami and Kittler (1995) have proposed a new region growing approach for image segmentation which uses gradient information to specify the boundary of a region. The method has the capability of finding the boundary of a relatively bright/dark region in a textured background. The method relies on a measure of contrast of the region which represents the variation of the region grey-level as a function of its evolving boundary during segmentation. This helps to identify the best external boundary of the region. The application of a reverse test using a gradient measure then yields the highest gradient boundary for the region being grown. The unique feature of the approach is that in each step at most one candidate pixel will exhibit the required properties to join the region. The growing process is directional so that the pixels join the grown region according to a ranking list and the discontinuity measurements are tested pixel by pixel. The authors have performed a number of experiments on synthetic and real images. The authors concluded that the results are more reliable and consistent than conventional thresholding methods. The algorithm is also insensitive to a reasonable amount of noise. The main advantage of the algorithm is that no a priori knowledge is needed about the regions.

Lu and Xu (1995) have proposed a region growing technique for texture segmentation, in which a two-dimensional autoregressive model is used for texture representation. In this technique an artificial neural-network is adopted to implement the parameter estimation process for the autoregressive model and to compute the local texture properties of regions during the segmentation process. The segmentation procedure consists of two
stages, namely the initial stage and the refining stage. At the initial stage, the image is partitioned into disjoint blocks or regions in the form of windows of a fixed size. The block is considered as a part of the internal area if it has the same texture class as its four neighbouring blocks in the horizontal and vertical direction. Otherwise, it is considered as undetermined. At the refining stage, all initial blocks are extended in parallel. Each undetermined block is divided into sub-blocks with equal size. Each sub-block is then compared with its neighbouring blocks by using the local information obtained by the neural network. The process terminates when each pixel in the image is assigned to one of the initial regions. The algorithm was tested on 38 different textures from the Brodatz album. The algorithm provided accurate segmentation with less computational time.

He and Chen (2000) have proposed a resonance algorithm for image segmentation that is not much different from seeded region growing algorithm. It is stressed that the resonance based segmentation algorithm is much more robust to illumination changes than conventional algorithms that work directly on grey scale images. The basic idea of resonance is based on the distribution of energy from a source to other elements in the system. This can be illustrated by considering each pixel in the image as a mass connected to a platform by a spring. The image contains several such mass-spring pairs that can be considered as immersed in water. When an external force acts on one such pair, there is a certain amount of displacement. If the external force matches the natural frequency of the mass, then the mass will resonate and the vibrations will spread to other mass-spring pairs that will oscillate as a result. The effect will only, however, last within a circle of certain radius and become weaker as we move away from the source. It is proposed that a seed based segmentation algorithm based on the above theory should be used only on feature images rather than original grey level images. The experiments are performed on Brodatz texture images of size 512x512 pixels. The grey level
average is extracted within windows of size 8x8 and a feature image is generated. Since averaging is illumination dependent, it is expected that the true ability of the algorithm can be experimentally determined using such a feature. The results show better performance of the resonance algorithm compared to fuzzy c-means and histogram analysis segmentation methods.

Singh and Al-Mansoori (2000) have compared region growing and gradient based techniques for detecting regions of interest in digital mammograms. These regions of interest form the basis of applying shape and texture techniques for detecting cancerous masses. The study also proposes a two stage method where gradient based techniques are applied followed by region growing method which yields less number of regions for analysis. First, the quality of image is improved by histogram equalization and fuzzy enhancement techniques. Their utility is then compared using three quantitative measures. After the enhancement step, the images are then subjected to region growing or gradient operations (masking) for segmentation purpose. The segmented image is then analysed for estimating the regions of interest and the results are compared against previously known diagnosis of the radiologist. The authors concluded that the region growing segmentation gives lesser number of regions for the analysis as compared to gradient based techniques without compromising on quality.

3.2.5 Clustering

Image segmentation can be performed effectively by clustering image pixels. Cluster analysis allows the partitioning of data into meaningful subgroups and it can be applied for image segmentation or classification purposes. Clustering analysis either requires the user to provide the seeds for the regions to be segmented or uses non-parametric methods for finding the salient regions without the need for seed points. Clustering is commonly used in a range of applications such as image segmentation and unsupervised
learning (Jain and Dubes 1998). A number of issues related to clustering are worth studying including how many clusters are the best and how to determine the validity of clusters. In a number of segmentation techniques, such as fuzzy c-means clustering, the numbers of clusters present in the image have to be specified in advance.

Kurita (1991) has developed an efficient clustering algorithm for region growing. This algorithm is a typical example of hierarchical clustering. The algorithm starts with an initial partition of a given image into N segments and sequentially reduces the number of segments by merging the best pairs of segments among all possible pairs in terms of a given criterion. The merging process is repeated until the required number of segments is obtained. The algorithm uses sorted linked lists to maintain neighbouring relations of segments and also a heap structure to store dissimilarities of all possible pairs of segments. A set of experiments was performed on a monochrome image and some range images of size 128x128 pixels. The number of regions turned out to be around 100 and the computational time was about 6 seconds. In case of range images each planar region is correctly classified into one segment but the curved regions were classified into several patches. The authors concluded that the algorithm makes the stepwise optimisation approach for region growing.

Frigui and Krishnapuram (1999) have addressed three major issues associated with conventional partition clustering. The issues addressed are sensitivity to initialisation, difficulty in determining the number of clusters and sensitivity to noise and outliers. Their proposed method Robust Competitive Agglomeration (RCA) starts with a large number of clusters to reduce the sensitivity to initialisation and determines the actual number of clusters by a process of competitive agglomeration. This method combines the advantages of hierarchical and partition clustering techniques. To
overcome the sensitivity to outliers, concepts from robust statistics are incorporated. Overlapping clusters are handled with the use of fuzzy membership. To handle doubtful regions, and to reduce the sensitivity to initialisation, the algorithm uses finite rejection. The agglomerative property makes it relatively insensitive to initialisation and local minima effects. The algorithm was tested on noisy data sets with synthetic ellipsoidal clusters and linear clusters, and a number of real range images. The results proved that RCA can provide robust estimates of the prototype parameters even when the clusters vary significantly in size and shape, and the data is noise contaminated.

Pauwels et al (1999) have investigated non-parametric clustering for image segmentation. Authors have proposed a robust and versatile method that is able to handle unbalanced and highly irregular clusters using intermediate level processing. First, a density image is constructed by convolving the image data with a Gaussian density kernel. The nearest neighbours for each pixel are determined in a neighbourhood. The neighbourhood considered is one percent of the total number of pixels in the image. After convolution, candidate clusters are identified using gradient ascent and each point is linked to the point of highest density among its neighbours. These process results in an overestimation of the number of clusters, carving up the data set into a collection of relatively small clumps centred on local maxima. A hierarchical family of derived clusters is constructed by using the data density to systematically merge neighbouring clumps. The order of merging is established by comparing the density values at neighbouring maxima with respect to density at the saddle-point (which is defined as the point of maximal density among the boundary points). The measures of cluster validity for a given cluster are then quantified by two mathematically defined indices: isolation and connectivity. Using this clustering algorithm, it becomes possible to extract image regions that are
salient and semantically meaningful. The authors have concluded that the algorithms produced a smooth version of the inverse equalisation algorithm without destroying any information in the image. The method can be used on greyscale images such as natural scenes, faces, handwritings etc. The method is analysed on a number of synthetic and real images. The regions segmented were found homogeneous and the objects were easily identified from the image background. One of the advantages of the algorithm is that it automatically determines the number of objects from the image (or the number of optimal clusters) and therefore no *a priori* knowledge about the number of objects is needed.

Ohm and Ma (1997) have proposed a feature based cluster segmentation method for image sequences. The algorithm analyses specific features from the image sequence and checks their reliability and evidence locally for images in order to build segments that are probably part of one object. The segmentation procedure is basically a clustering procedure which takes into account different features such as colour and motion. The approach is similar to vector quantisation. Different weights are applied to different features according to their reliability. The pixel-feature vector is then compared to a set of cluster-feature vectors and hence classified to a feature class to generate clusters. The set of cluster feature vectors is updated for each image in the sequence which is also used for the segmentation of the next image. The labels associated with different clusters remain identical for different images of a same scene. After a scene change, the whole process is started with a completely new set of feature vectors which are computed from the first frame after the change and hence the whole sequence is segmented. The algorithm was performed on a sequence of flower garden images. In all of the frames the segmentation and tracking was done automatically. The authors concluded that the algorithm is a hybrid combination of a block-recursive and a pixel-recursive technique and produces a dense vector
field. The algorithm also has the capability to set flexible weights for different feature components automatically.

Ng (2000) has described an extension to the conventional $k$-means algorithms by modifying the splitting rule in order to control the number of cluster members. By adding suitable constraints into the mathematical program formulation, the author developed an approach that allows the use of $k$-means paradigm to efficiently cluster data sets with a fixed number of elements in each cluster. The main objective of this algorithm is the minimisation of an objective function usually taken as a function of deviations calculated for all patterns from their respective cluster centres. This algorithm minimises the objective function through a scheme that starts with an arbitrary initial cluster membership in an iterative manner to obtain better clustering results. Distributed pruning technique is used to reduce the number of clusters in the parallel manner. The results show that as the amount of pruning increases, as the data between partitions becomes more skewed i.e., the similarity between a sample and its membership in two or more clusters increases.

The comparison of various clustering techniques and studying their behaviour is important. It is important that an optimal clustering technique is able to choose the correct number of clusters. Dubes and Jain (1976) have presented detailed comparisons on three classes of clustering techniques. A total of eight programs are compared on minimised squared error based clustering, hierarchical clustering and graph theoretic clustering. Important issues related to the use of clustering programs such as user options, computational cost, input and outputs, and comparing programs that yield different outputs have been considered. The authors have proposed that different clustering algorithms can be compared on the basis of four factors: manner in which the next clustering is chosen or updated; criterion for
creating new clusters; criterion for deleting and merging clusters; and the initialisation procedure. Approaches based on ranking clustering programs on the basis of their performance on a standard data set with performance criterion such as minimising squared error are criticised. One way of categorising such program is to cluster their performances. A set of nine criterion are defined for comparing clustering programs based on the manner in which clusters are formed, structure of data, and the sensitivity of the clustering technique to changes in data that do not essentially alter its nature. The different clustering programs are compared on the basis of these criteria using handwritten character data and a similarity matrix for these methods is shown.

The validity of clusters is important to study. Yarman-Vural and Ataman (1987) have studied several areas of clustering methodology including the definition of clusters, determination of the number of clusters, heuristic partition clustering algorithms and the effect of noise on determining accurate clusters. Cluster validity criteria including maximum likelihood information criteria and sum of squared errors are discussed. The first criteria is found to be better when the number of clusters changes. This paper also proposes a method of improving conventional clustering algorithms so that they display better performances with noise contaminated data. The ISODATA algorithm is modified to show its improved performance on up to 20% noise contamination in an artificial Gaussian mixture data set. Similar good performances are achieved on image analysis applications. Dubes (1987) has investigated the utility of Davies and Bouldin index for cluster validity and compares it with the proposed improved Hubert statistic. Results on a Monte Carlo study are reported by varying parameters such as dimensionality, number of patterns, sampling window, and number and spread of clusters. The new index is found to perform much superior as it shows more consistent performance when the above parameters are varied.
Zahid et al (1999) have proposed a new heuristic method for fuzzy cluster-validity. Its principle is based on the evaluation of fuzzy separation and fuzzy compactness of clusters. The main attempt of the study is to analyse the natural structure present in the data set. This evaluation, which measures the quality of fuzzy clustering, assigns a real number to the outputs of the used fuzzy clustering algorithm. Two kinds of outputs are considered. For each output, a combination of fuzzy separation and fuzzy compactness criterion is evaluated. Maximizing the resulting criteria results in good clustering. This maximum value allows in identifying the right number of clusters. The first function calculates a fuzzy separation and fuzzy compactness ratio by considering geometrical properties and membership degrees of the data. The second function evaluates this ratio by using only the properties of membership degree. The aim of classification phase is to generate a well-defined fuzzy c-partition that is as close as possible to the natural structure of the data. Four numerical examples were used by the authors to illustrate the use of the proposed algorithm as a validity function. The obtained results indicate that the proposed method provides better answers than other cluster-validity functions especially with overlapping fuzzy clusters.

3.2.6 Probabilistic and Bayesian Approaches

Haddon and Boyce (1990) have used co-occurrence based approach to image segmentation making use of region and boundary information in parallel for improved performance on a sequence of images. The authors examined image segmentation by unifying region and boundary information using co-occurrence matrices. The co-occurrence matrices were used to generate the feature space. The analysis was performed in the context of an ensemble of images. Based on the location of the intensities of each pixel and its neighbours in the co-occurrence matrix, initial segmentation is done. Each
pixel is then associated with a tuple which specifies whether it belongs to a given region or if it is a boundary pixel. This tentative segmentation was then refined by relaxation labelling that ensures local consistency of pixel labelling during segmentation by minimising the entropy of local neighbourhoods (Hansen and Higgins 1997, Hummel and Zucker 1976). If a pixel does not belong to the boundary, then it is assigned to one of the regions. This classification is entirely uni-dimensional in the co-occurrence direction and contains no explicit local consistency. The consistency for regions and boundary was obtained assuming that boundaries are not wider than one pixel. The algorithm was applied to synthetic images and infrared images of natural scenes. The authors found that segmentation worked reasonably well for most of the areas in the images while the majority of real edges and boundaries were correctly located. The algorithm was applied on a series of infrared greyscale images taken by a low flying aircraft. The authors conclude that the technique is robust and performs well even in poor imaging conditions. The algorithm is less effective if the clusters in the co-occurrence space have substantial overlap due to the imposition of local consistency. Since the techniques use global information in a local context, it was possible to adapt it to varying image characteristics i.e. variation in colour and texture. The difference between the co-occurrence matrices generated by a sequence of images is negligible if the image content and conditions do not change significantly.

A further extension of the above methodology is provided by Haddon and colleagues in their later study, Haddon and Boyce (1993). The authors have suggested two intrinsically similar techniques for image segmentation and edge detection based on co-occurrence matrices. The first technique defines transforms that enhance the difference between typical and atypical image features. The second technique uses the location of region distributions in an edge co-occurrence matrix and defines the location of
corresponding boundary distributions. The techniques are analysed by labelling the matrices. The labelled matrix can be used to segment the regions of an image and to simultaneously detect prominent regions. The image segmentation and edge detection is done at the same time. The operator used to generate the edge co-occurrence matrix effectively generates a lookup position within the labelled matrix. The operator is convolved with the neighbourhood of a pixel and this pixel is then replaced by the label of the co-occurrence matrix at the lookup position. This process is repeated using the same operator for a variety of operator orientations. Correlation techniques and conjugate iteration schemes are used to determine the location and the extent of the distributions in the co-occurrence matrix. The algorithm is applied to FLIR imagery and multispectral data. The majority of components in the algorithm are performed in parallel. The results are significantly improved as compared to the previous algorithm where segmentation and edge detection were performed simultaneously.

Haddon and Boyce (1994) have used co-occurrence method to segment a sequence of FLIR images into its major regional components. These regions are subjected to texture classification. The co-occurrence matrices are temporally smoothed to ensure consistency of segmentation in the sequence. The edge co-occurrence matrices are decomposed using discrete Hermite functions because their underlying structure is Gaussian due to the Guassian noise present in the images. If this structure is removed, what is left is the texture structure of the image. Neural networks are then used to classify the images using the coefficients of the Hermite functions. Co-occurrence matrix is a multidimensional histogram, in which each element \((i k)\), is the frequency with which two events \(i\) and \(j\) co-occur with specific relationship to each other. Edge co-occurrence matrices emphasize the intensity difference between two pixels. The authors of this paper used the canny edge operator to form the matrices. A sequence of 300 FLIR images
from a low flying aircraft approaching a bridge are obtained and classified. These images are first segmented using co-occurrence matrices into their major regions and using temporally smoothed averaged matrices. Then a co-occurrence matrix is formed for every major region. The matrices are decomposed using orthogonal Hermite functions. The higher order functions describe the texture of the region and form the input vector of 121 features. Every third data sample was taken to form the validation set. Training and validation sets are disjoint. MLP networks were trained until validation error started to increase (over-train). Additionally, principal components analysis was used to extract the most descriptive of those 121 features. Very good results have been obtained with both a single layer network with 96.9% training and 98.3% validation accuracy and a two layer network with 94.8% training and 88.7% validation recognition accuracy. The major regions classified were grass, trees, sky, river reflecting trees and river reflecting sky. The major problems are the reflecting classes. PCA results were not as good as with the use of the full 121 features. The techniques described in this paper seem to be robust with noise. Overall, 93.4% of image area was correctly classified.

Haddon et al (1997) have developed an algorithm for automatic segmentation and classification of images using a co-occurrence based approach using Hermite functions. The basic approach consists of four steps: segmentation, texture analysis, initial classification and labelling and spatio-temporal classification. In the segmentation phase, the key regions are separated from each other and from the remainder of the scene using co-occurrence matrices and edge detection simultaneously. The texture classification of segmented regions is performed using discrete Hermite functions. The co-occurrence matrices are decomposed for the calculation of features of segmented regions. The result of texture analysis is a low order feature vector that describes the texture of a segmented region. The features
that contain information on discriminating between classes are selected and used as inputs to a multi-layer neural network classifier. Local consistency of interpretation is achieved in regions both spatially within an image and temporally within a sequence of images. A relaxation labelling approach is used to associate pixels belonging to particular coefficients. The compatibility coefficients take into account the current image and its predecessor in a sequence of images and results of relaxation labelling are applied to the previous image for classification. The authors evaluated the results on a sequence of FLIR images taken from a low flying aircraft and also performed snow profile analysis. The authors found that the algorithm is generally applicable to a wide range of imaging applications.

Haddon and Boyce (1998) have presented a method of ensuring consistency of interpretation in terms of the classification of segmented images, both spatially within an image and temporally through a sequence of FLIR images taken from a low-flying aircraft. It is important to maintain consistency with time in video analysis. Inconsistencies in segmentation between frames of a sequence are unacceptable in autonomous vehicle navigation applications. The image is first segmented using a co-occurrence based technique. After the image is segmented, the regions are classified based on texture. Co-occurrence matrices are also used for describing the texture of each region. The co-occurrence matrices are then decomposed using discrete Hermite functions. Features selected by using the linear discriminant analysis and principal components analysis were fed into a Multi Layer Perceptron (MLP) neural network for classification. In enforcing spatio-temporal consistency, the output values of the neural network are taken into consideration. Two or more roughly equal outputs indicate that both outputs are considered equally likely. Consistency is enforced only when there is insufficient evidence of a region class. Consistency is enforced by using relaxation labelling, however, instead of using it in its classical way, the
authors treat a region as a single entity with both spatial and temporal neighbours. Adjacent regions, both spatially and temporally, are considered as neighbours. These neighbours influence the probability that a region belongs to a particular class if the result of the neural network is insufficient. These techniques were applied to a 12 second 300 frame sequence of infrared images. Five classes, trees, grass, sky, river reflecting trees and river reflecting sky are considered. The authors find very good results after the initial classification (first frame) regions. Regions that have equal membership in all classes were left as white. After consistency enforcement, all regions are classified, and only a few misclassification errors are found.

Comer and Delp (1999) have proposed a new algorithm for the segmentation of textured images using a multi-resolution Bayesian approach. This algorithm uses a Multi-resolution Gaussian Auto-Regressive (MGAR) model for the pyramid representation of the observed image. This algorithm effectively uses larger neighbourhoods to perform the segmentation than the single resolution algorithms. The observed data, which is a multi-resolution representation of the observed image, is obtained using Gaussian pyramid decomposition. The nodes in a binary tree then index the data and an MGAR model is used. Value of a random variable is then predicted as a linear combination of the values of the random variables at previous nodes. A Gaussian pyramid representation of the image is generated. A multiscale Markov Random Field model is then used for the class label pyramid. The optimisation criterion is then used for the image segmentation. This criterion minimises the expected value of the number of misclassified nodes in the multiresolution lattice. Expected Maximization (EM) algorithm is finally used for parameter estimation. This method also assigns a cost to an incorrect segmentation based on the number of incorrectly classified pixels in the segmentation. The algorithm was applied to two different images where three pyramid levels were considered, i.e. the original image was examined at three
different resolutions. The algorithm performed well on both the images segmenting them into regions of homogeneous textures. The algorithm was found successful in segmenting the image better than the EM and Multiresolution Posterior Marginal techniques.

3.2.7 Neural Networks Segmentation

Campbell et al (1997) have proposed an automatic segmentation and classification method for outdoor images using neural networks. First, the images are segmented using Self-Organising Feature Maps (SOFM) based on texture and colour information of the objects. SOFMs used consisted of 64x64 nodes for best segmentation. A set of 28 features is then extracted from each region. The features include: average colour, position, size, rotation, texture (Gabor filters) and shape (using principal components). Classification is then performed using a Multi Layer Perceptron (MLP) with 28 input nodes and 11 output nodes. The training is performed on 7000 regions and testing is done on an independent set of 3000 samples. Over 80% regions were classified correctly using Learning Vector Quantisation and 91.9% regions were classified correctly using the Multi Layer Perceptron.

Papamarkos et al (2000) have developed the procedure of image segmentation using self-organising maps. The use of these neural network paradigms is considered equivalent to multithresholding where the output of the network defines a number of homogeneous clusters. One of the interesting notes from this paper is the technique used for finding the optimal number of thresholds or in other words the number of segmented regions. Considering that grey level distributions within the region are Gaussian, it is suggested that the linear combination of probability density functions for individual regions should match the overall density of the global histogram of the original image. For different number of segmented regions, we can find which result minimises the error.
3.2.8 Other Approaches

Medioni and Yasumoto (1984) have proposed using fractal dimension method for image segmentation. The authors computed the fractal dimension of textures in frequency domain by approximating the log power spectrum by a straight line. The difference of the logarithm of the expected values of the two dipole-statistics yields a single texture feature. This procedure approximates the fractal dimension, which relates to the expected statistics of the grey-level difference versus the statistics of distance vector. The main drawback of this method is that segmentation based on single feature extracted from images is not always possible.

Pentland (1984) has addressed the problem of fractal based description of natural scenes. The authors addressed the following problems: (i) Representing natural shapes such as mountains, trees, and clouds and (ii) Computing such description from image data. It has been observed that many natural objects such as leaves, snowflakes, etc. have a fractal nature. This makes it impossible to measure them accurately in an image. For example, measuring the length of a coastline, no matter what the size of the measuring tool is selected, all the curves smaller than the size of the measuring tool will be missed. Standard notions of length and area do not produce consistent measurement for many natural shapes: the basic metric properties of these shapes vary as a function of the fractal dimension. Fractal dimension, therefore, is a necessary part of any consistent description of such shapes. This means that fractals have to be used in the measurement of such shapes. The characterisation of image texture by means of a fractal surface model makes it impossible to describe image in a manner that is stable over transformation of scale and linear transforms of intensity. The authors demonstrate how natural images can be segmented using this method. It is suggested that in this manner the segmentation is more stable in scaling the
images down rather than using thresholding techniques on image intensity. Additionally, comparison with correlation techniques and co-occurrence techniques has been made using a mosaic of 8 natural textures constructed by Laws (1980). On these textures, the authors demonstrate that the fractal methodology yields better results than known segmentation algorithms.

Xu et al (1998) present a segmentation algorithm that is based on partitioning the image into arbitrarily shaped connected regions to minimise the sum of grey level variations over all partitioned regions under the constraints that each partitioned region has a specified number of pixels and that two adjacent regions have significant differences in average of grey levels. A minimum spanning tree has been used to construct these partitions and as such the segmentation problem reduces to tree partitioning problem. The effect of various noise contaminations on the segmentation results is evaluated on the basis of how the noise affects the tree representation and tree partitioning.

In some studies, topological maps have been used as a structural method for image segmentation. Brequelaire and Brun (1998) proposed a novel technique for image segmentation using topological maps and inter-pixel representation. The authors describe a data structure that allows the system to store and process regions of any size or form. It is called a model of discrete maps which helps in efficient computation of parameters required by the segmentation algorithm. There are two aspects in a segmented image representation: the geometrical aspect which describes the shape of regions, and the topological aspect which describes neighbourhoods and inclusions of regions. A topological map is defined as a partition of an oriented surface in a finite set of points. A discrete map is based on both geometric and topological levels of representation co-operating together. The discrete map data structure is then used to design segmentation algorithms. In the study, the
authors have described a recursive split and merge algorithm for segmentation using discrete data structure. In the first step, segmentation is done and in the second step refinement is done recursively to improve the quality of segmentation. The results are provided for a test image of Lena and images for various iterations are shown.

Ojala and Pietikainen (1999) have presented an unsupervised texture segmentation method using feature distributions. The proposed algorithm uses distributions of local binary patterns and pattern contrasts for measuring the similarity of adjacent image regions during the segmentation process. Texture information is measured with a method based on Local Binary Patterns and Contrast (LBP/C). The segmentation method consists of three phases: hierarchical splitting, agglomerative merging and pixel-wise classification. First, hierarchical splitting is used to divide the image into regions of roughly uniform texture. Then, agglomerative merging is performed to merge similar adjacent regions until a stopping criterion is met. A pixel-wise classification is then performed to improve the localisation. The authors use a non-parametric likelihood test as a pseudo-metric for comparing feature distributions. The authors tested the method on four texture mosaics and two greyscale images of natural scenes. The method performed very well in these experiments. The authors concluded that the method is not sensitive to the selection of parameter values and does not require any a priori knowledge about the number of textures or regions in the image. The method can also be easily generalised to utilise other texture features, multiscale information or colour features.

Yoshimura and Oe (1999) have proposed a segmentation algorithm for texture images using genetic algorithms that automatically determine the optimum number of segmentation areas. The authors divide the original image into many small rectangular regions and extract texture features from
the data using two dimensional autoregressive models, and other features such as fractal dimension, mean, and variance. The authors use three types of evolutionary segmentation methods. The first method uses Genetic Algorithms (GA), the second method uses GA's and Self Organising Maps (SOM), and the third method uses GA's and SOM considering the optimum number of segmentation areas in an image. Various experiments were performed on a set of simulated images with 3 or 4 different real textures. The authors found that the third method performed the best and the algorithms were visually accurate as compared to other conventional techniques. The first method produced good results but it is necessary to specify the number of regions beforehand. The second method selected the number of regions automatically, and the third method finds optimum number of regions with homogeneous texture. The authors concluded that the methods are effective for the segmentation of images that contain similar texture fields.

Perner (1999) has presented a methodology for a case based reasoning image segmentation system. The image segmentation is performed by looking up a case base for similar cases and using the segmentation parameters associated with the matched case. Hence, images are first matched for similarity on the basis of features such as image moments. Similarity is determined on the basis of measure proposed by Tversky (1984) for non image information and a weighted similarity measure for image data. If ideal segmentation parameters are known for reference images, by matching new images to these reference images, the same segmentation parameters can be used. The system is shown to perform well on brain CT segmentation.

3.3 SEGMENTATION EVALUATION

Kitchen and Rosenfeld (1984) have discussed the various issues related to under- and over-segmentation of images. The problems are blamed on the fact that context information is not used in segmentation algorithms.
For example, a difference in brightness that is not significant in some contexts may well be significant in other backgrounds. Under-segmentation is considered as a much more serious problem as it is easier to recover true segments through a merging process after over-segmentation rather than trying to split a heterogeneous region. In some situations, segmentation evaluation is based on how well the segmented image corresponds to the ground truth segmentation in terms of pixel by pixel differences. In most scene analysis applications, these differences need to be minimised and hence the segmentation criteria are rather harsh in terms of penalising algorithms severely if they do not segment accurately. However, in some applications of segmentation, such as in medical imaging, it may be enough that the segmented region overlaps with the true region (e.g. if the segmented region of interest overlaps with the true region, e.g. a tumour, then the segmentation can be classed as successful). Hence segmentation can be considered acceptable even if it only partially detects the true borders or region.

The evaluation of image segmentation programs is an important field of study. The evaluation can be categorised as supervised or unsupervised based on whether a priori information is available or not. A number of different approaches have been adopted. Some qualitative guidelines have been developed in Haralick and Shapiro (1985). Quantitative techniques of evaluation have been used by several authors. Weszka and Rosenfeld (1978) have used a busyness measure and classification error as performance criteria. Lavine and Nazif (1982) have defined a set of parameters for unsupervised segmentation including region uniformity, region contrast, line contrast and line connectivity. Sahoo et al (1988) have used the uniformity criterion of Lavine and Nazif and a shape measure computed from the gradient values and the selected threshold value. In the case of supervised segmentation, errors are measured using a reference. The differences between the reference and segmentation output are measured to determine how well
the segmentation algorithm performs. The simplest measure for supervised segmentation is probability of error as shown in Yang et al (1995). This criterion can be used for finding optimal threshold values. This error can be decomposed into an under-merging or over-merging error. This error measure may not however give full information on the quality of segmentation. Yasnoff et al (1977) have in addition proposed the pixel distance error for measuring segmentation performance in which the positions of misclassified pixels are taken into account. Two error measures, the percentage area misclassified and pixel distance error are used.

Another approach to measure the quality of segmentation is based on the differences in the values of features computed from an ideally segmented image and an actual segmented image. These can be any object features, e.g. shape features such as eccentricity, or their texture, etc. This approach has been detailed by Zhang (1997), Zhang and Gerbrands (1992). Yang et al (1995) have used shape features including area, circularity, orientation and elongation of segmented objects for the comparison of three segmentation methods. The under-merging, over-merging and ultimate measure of accuracy criteria are evaluated for comparing the algorithms considered. Wang et al (1996) compare several approaches for texture image segmentation. The techniques used are $k$-means, fuzzy $k$-means, fuzzy Adaptive Resonance Theory (ART2) and fuzzy Kohonen Self Organising Feature Mapping (SOFM). In their tests five features including energy, entropy, correlation, inertia, and homogeneity are used. After segmenting the images, features are extracted using Spatial Grey Level Dependence Method (SGDLM). The method is developed by Wang et al (1994). This method is used because it is 3 times faster than SGLDM method developed by Haralick. The co-occurrence features are calculated in four directions of 0, 45, 90 and 135 degrees with respect to the horizontal axis. These features are selected because of: (i) strong within class variance; (ii) strong between class
separation; and (iii) low sensitivity to small sample size. Experiments were performed on 2 images with different textures of size 150x360 pixels and 100x100 pixels. The experiments showed some limitations of k-means algorithm such as its slow processing ability and instability. Clustering errors were reduced by SOFM. Another advantage of SOFM algorithm is that the size of the neighbourhood was automatically adjusted due to its self-organising nature. The numerical results showed that the fuzzy approaches are better than conventional k-means algorithms and SOFM provides the best approach for texture image segmentation.

Hoover et al (1994, 1996) have proposed a methodology for evaluating range image segmentation techniques. The authors have also created a framework for the comparison of range image segmentation techniques. Experiments were performed on the data acquired through laser range finder or a triangulation technique. Images are segmented using any popular method that is to be evaluated. The authors have developed a tool for a human operator to create ground truth segmentation. The operator assigns three different labels to the regions. The labels are shadow pixels, noise pixels and cross edge pixels. Then machine segmentation is compared to the ground truth. Every region in the ground truth and machine segmented images is mapped as one of the five possibilities: correct detection, over-segmentation, under-segmentation, missed and noise. The first three categories apply to mappings of regions between the ground truth and machine segmented images. A missed classification applies to a region in the ground truth image and noise classification applies to a region in the machine segmented image.

3.4 COMPARATIVE STUDIES AND REVIEWS

Haralick and Shapiro (1985) have stated that: “As there is no theory of clustering, there is no theory of image segmentation”. The authors categorise segmentation algorithms as: measurement space guided spatial
clustering, single linkage region growing schemes, spatial clustering schemes, single linkage region growing schemes, hybrid linkage region growing schemes, centroid linkage region growing schemes, and split and merge schemes. The paper details the basics of these schemes with examples.

Similarly, Fu and Mui (1981) have stated that: “Almost all image segmentation techniques proposed so far are *ad hoc* in nature. There are no general algorithms that will work for all images”. Their survey categorises image segmentation studies as follows: feature thresholding or clustering, edge detection techniques, and region extraction techniques. The authors recommend that one of the fruitful areas of further investigation lies in combining spatial and semantic information with edge detection and thresholding or clustering techniques to perform image segmentation.

### 3.5 REVIEW ON HAND GEOMETRY BASED SYSTEMS

Hand Geometry refers to the geometric structure of the hand that is composed of the lengths of the fingers, the width of a palm, etc. The advantages of a hand geometry system are that it is a relatively simple method that can use low resolution images and provides high efficiency with great users’ acceptance (Golfarelli et al 1997). Among various kinds of biometric identifiers, hand based biometrics has been attracting considerable attention over recent years due to its ease in accession. Hand-based authentication schemes in the literature are mostly based on geometrical features. A brief survey of reported systems for hand geometry verification can be found in Jain et al (1999). An elaborate survey on hand geometry verification is given in Duta (2009). Geometrical features of the hand constitute the bulk of the hand features adopted in most of the hand recognition systems. One advantage is that geometrical features are more or less invariant to the global positioning of the hand and to the individual planar orientations of the fingers and the palm.
Among numerous geometrical measures include lengths, widths, areas, and perimeters of the hand, fingers and the palm, Jain et al (1999) have shown that the hand geometrical features solely are not sufficient discriminative. Therefore, for more demanding applications one must revert to alternative features such as hand global shape, appearance and/or texture. Jain, Ross and Pankanti (1999) have used a peg-based imaging scheme and obtained sixteen features, which include length and width of the fingers, aspect ratio of the palm to fingers, and thickness of the hand. The prototype system they developed was tested in a verification experiment for web access over for a group of 10 people. Sanchez-Reillo et al (2000) use a similar set of geometric features, containing the widths of the four fingers measured at different latitudes, finger and palm heights, finger deviations and the angles of the inter-finger valleys with the horizontal. The twenty-five selected features are modeled with Gaussian mixture models specific to each individual. Wong and Shi (2002) have used lengths and interfinger baselines in addition to finger widths and employ the fingertip regions. Bulatov et al (2002) describe a peg-free system where 30 geometrical measures are extracted from the hand images. In addition to the widths, perimeters and areas of the fingers, they also incorporate the radii of inscribing circles of the fingers. Oden et al (2003) have used fourth degree implicit polynomial representation of the extracted finger shapes in addition to such geometric features as finger widths at various positions and the palm size. The resulting sixteen features are compared using the Mahalonobis distance.

The other approach in hand geometry verification is contour based. In the work of Jain and Duta (1999), the contour is completely determined by the black-and-white images of the hand and can be derived from it by means of simple image processing techniques. It can be modeled by features that capture more details of the shape of the hand. Lay (2000) has introduced a technique where the hand is illuminated with a parallel grating that serves
both to segment the background and enables the user to register his hand with one of the stored contours. The geometric features of the hand shape are captured by the quad tree code.

Ender Konukoglu et al (2006) have employed a hand shape-based approach for person identification and/or verification. The algorithm is based on preprocessing the acquired image, which involves segmentation and normalization for hand’s deformable shape. In this context, ‘hand normalization’ signifies the registration of fingers and of the hand to standard positions by separate rotations of the fingers as well rotation and translation of the whole hand. Yoruk et al (2006) introduced a more accurate and detailed representation of the hand using the Hausdorff distance of the hand contour, and Independent Component Analysis (ICA) features. Finally let’s note that there exist a number of patents on hand information-based personnel identification, based on either geometrical features or on hand profile.

3.6 REVIEW ON HYBRID MODEL BIOMETRIC SYSTEMS

Hybrid model biometric systems have received much attention in recent literature. A hybrid model could be adopted to make a biometric system more secure. This approach uses two or more systems or techniques to yield greater accuracy due to the statistical independence of the selected approaches. Therefore more than one identifier is used to compare the identity of the subject.

A number of works showing advantages of multimodal or hybrid biometric verification systems have been reported in literature. Brunelli et al (1980) have described a multimodal biometric system that uses the face and voice traits of an individual for identification. The authors have combined the matching scores of five different matchers operating on the voice and face features to generate a single matching score that is used for identification.
Brunelli and Falavigna (1995) have proposed personal authentication system based on acoustic and visual features, where they use a HyperBF network as the best performing fusion module. Duc et al (1997) proposed a simple averaging technique combining face and speech information. Bigun et al (1997) have developed a statistical framework based on Bayesian statistics to integrate information presented by the speech (text-dependent) and face data of a user. Kittler et al (1998) have experimented with several fusion techniques using face and voice biometrics, including sum, product, minimum, median, and maximum rules and they have found that the best combination results are obtained for a simple sum rule. Hong et al (1998) have combined face and fingerprints for person identification. Authors have consolidated multiple cues by associating different confidence measures with the individual biometric matchers and achieved a significant improvement in retrieval time as well as identification accuracy. Ben-Yacoub et al (1999) proposed several fusion approaches such as Support Vector Machines (SVM), tree classifiers and multi layer perceptrons, combining face and voice biometrics. Kumar et al (2003) combined hand geometry and palm print biometrics in a verification system.

A commercial product called BioID (Frischolz and Dieckmann 2002) uses voice, lip motion, and face features of a user to verify identity. Jain and Ross improved the performance of a multimodal biometric system by learning user-specific parameters (Jain and Ross 2002). General strategies for combining multiple classifiers have been suggested in (Ho et al 1994, Kittler et al 1998). All the approaches presented in (Ho et al 1994) attempt to reduce or re-rank a given set of classes. These techniques are thus relevant to the identification problem in which a large number of classes (identities) are present. Prabhakar and Jain (2002) have showed, in the context of a fingerprint verification system that combining multiple matchers, multiple enrollment templates, and multiple fingers of a user can significantly improve
the accuracy of a fingerprint verification system. The authors also argue that selecting matchers based on some ‘goodness’ statistic may be necessary to avoid performance degradation when combining multiple biometric modalities. There is a large amount of literature available on the various combination strategies for fusing multiple biometric modalities using the matching scores (Dieckmann et al 1997, Verlinde and Cholet 1999, Ben-Yacoub et al 1999). It is well known that independence of modalities plays a very important role in the amount of improvement when combining multiple biometric modalities. A carefully designed combination scheme that has been trained and tested on a large amount of data is expected to perform better than the best of the individual ingredient modalities.

George Chellin et al (2009) designed the multimodal biometric system that uses the iris and fingerprint traits of an individual for identification; decision fusion is designed to operate at the measurement level. Implementation showed that the system reduces the error rate and improves the performance of the system by achieving good acceptable rate during identification and authentication. Anil Jain et al (2004) proved that the recognition performance of a fingerprint system can be improved significantly by using additional user information like gender, ethnicity, and height. Zewail et al (2004) used iris colour with fingerprint and iris using a weighted average scheme and a Parzen classifier. A combination of uncorrelated modalities (e.g., fingerprint and face or two fingers of a person) is expected to result in a better improvement in performance than a combination of correlated modalities (e.g., different impressions of the same finger or different fingerprint matchers). Further, a combination of uncorrelated modalities can significantly reduce the failure to enroll rate as well as provide more security against ‘spoofing’. On the other hand, such a combination requires the users to provide multiple identity cues, which may cause inconvenience. Additionally, the cost of the system increases because of
the use of multiple sensors (e.g., when combining fingerprints and face). The convenience and cost factors remain the biggest barriers in the use of such multimodal biometrics systems in civilian applications. Anticipation for high security applications, large-scale identification systems, and negative identification applications will increasingly use multimodal biometric systems, while small-scale low-cost commercial applications will probably continue striving to improve unimodel biometric systems.

3.7 REVIEW ON FINGER KNUCKLE PRINT BASED BIOMETRIC SYSTEMS

In the past three decades, many biometric characteristics have been investigated. Recently, it has been noticed that the textures in the outer finger surface, also known as the dorsum of hand, has the potential to do personal authentication. Researchers have reported that Finger-Knuckle-Print (FKP), the image pattern of skin folds and creases in the outer finger surface, is highly unique and can serve as a distinctive biometric identifier. The advantages of employing finger knuckle images are numerous. Firstly, user acceptance of outer palm surface imaging is very high since, FKP is hard to be abraded as people hold stuffs with the inner side of the hand. In addition, unlike for finger prints, there is no stigma of criminal investigation associated with the finger knuckle surface. Second, the finger geometry can be acquired simultaneously from the same image and employed to improve the system’s performance. Moreover, people rarely leave FKP remains on the stuff surface, making the loss of private data less possible. Thus, FKP has a great potential to turn into a widely accepted biometric identifier.

Woodard and Flynn are among the first scholars who exploited the use of finger knuckle surface in biometric systems. Woodard et al (2005) used the 3D range image of the hand to calculate the curvature surface representation of the index, middle, and ring fingers for similarity
comparison. The method exploits the 3-D shape information of finger back surface but does not fully use the texture information. However, their work did not provide a practical solution in establishing an efficient system using the outer fingers. Li et al (2004) used a hierarchical classification method to study knuckle print based on location and line features. In Ribaric and Fratric (2005), Principal Component Analysis (PCA) was employed to project finger images into lower dimensional subspace. Later, Ravikanth et al (2007) proposed another approach to personal identification by using 2-D finger back surface imaging. They developed a device to capture hand-back images and then extracted the finger knuckle areas by some preprocessing steps. They applied the subspace analysis methods such as Principal Component Analysis, Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA). These methods were combined for feature extraction and person classification. The subspace analysis methods may not be able to effectively extract the distinctive line and junction features in finger back surface. These works made a good effort to validate the uniqueness of biometric features in the outer finger surface. However, they did not provide a practical solution to establish an efficient system using the outer finger surface features.

Loris and Alessandra (2009) investigated knuckle features by fusing the knuckle print pattern from the middle and ring fingers. Lei Zhang et al (2009) developed a finger knuckle based authentication for recognizing a person’s identity. A specific data acquisition device was constructed to capture the FKP images and then an efficient FKP algorithm was presented to process the acquired data in real time. The local convex direction map of the FKP image was extracted based on which a local coordinate system was established to align the images and a region of interest was cropped for feature extraction. For matching two FKPs a feature extraction scheme which combines orientation and magnitude information extracted by Gabor filtering was proposed. In their later work, the Fourier transform of the whole image
was taken as the feature and the band-limited phase-only correlation-technique was employed to calculate the similarity between two FKP images. Ajay Kumar and Ravikanth (2009) developed a completely automated system that recognizes individuals using finger-knuckle images. The system uses a machine-vision camera and automatically segments the relevant regions to extract meaningful information. The system has been designed to automatically identify the presence of rings and extract the most suitable region for finger identification.