CHAPTER 2

LITERATURE SURVEY

2.1. Introduction

Object recognition is a computationally expensive process. Efficient algorithms are essential at all stages of the recognition process including feature extraction, invariant computation, and matching objects in target images with those in the model-base. Desirable properties of a model include good representation of objects, fast and efficient learning algorithms for the classification of the objects [50]. Literature possesses wider studies on the field of object recognition. Many practical object recognition systems are appearance-based or model-based. Here in this research, we have reviewed some of the techniques available in the literature for the recognition of objects in digital images.

Colour histograms are an early view-based approach to object recognition proposed by Swain and Ballard [24]. Given an example image of an object, the percentage of pixel values are measured in different colour bins. Then when a new object is presented the system can compare its colour histogram with that of previously seen example
objects to classify it. The main drawback of this approach is its sensitivity to lighting conditions, since an object can have a different colour when it is illuminated [48]. Not every object can be described by only its colour property. Schiele and Crowley [63] extended the idea of histograms to include other local image features such as orientation and gradient magnitude, forming multidimensional histograms. Although these methods could perform better than the previous method and are robust to changes in rotation, position and deformation, they cannot cope with recognition in a cluttered scene. Schneiderman and Kanade [64] proposed a method for object categorisation in natural scenes, computing histograms of wavelet transform coefficients over localised parts of an object.

The other models generally used for pattern recognition includes Statistical model, Structural model, Template matching model, Neural network based model, Feature based model, etc [77].

2.2. Statistical Model

In Statistical model each pattern is described in terms of features. Features are chosen in such a way that different patterns occupy non-overlapping feature space. The decision boundary is determined after analyzing the probability distribution of a pattern belonging to a class.
Here patterns are pre-processed to make them suitable for training [71]. System learns to classify patterns and adapts itself for unknown Patterns. Test patterns are applied to check the ability of the system to recognize patterns. Feature measurement is done during testing and then presented to the learned system for performing classification.

The parametric classification schemes are used when conditional probability density distribution is known, and otherwise non parametric classification scheme need to be used. Various decision rules are available to determine the decision boundary such as Bayes Decision Rule, Optimal Bayes Decision Rule, Maximum Likelihood Rule, etc. In case of noisy patterns the choice of statistical model is a good solution as feature spaces are partitioned, system becomes noise insensitive. Depending on the learning method used statistical technique can be categorized as: Discriminant Analysis and Principal Component Analysis.

Discriminate Analysis is a supervised technique with the approach for dimensionality reduction. Here a linear combination of features and, a discriminant function for each pattern class is defined to perform the classification function [3], [79]. The Discriminant Analysis methods that are used depending on the application and system requirement are Linear Discriminant Analysis (LDA), Null-LDA (N-LDA), Fisher Discriminant
Analysis (FDA), Two Dimensional Linear Discriminant Analysis (2D-LDA), Two Dimensional Fisher Discriminant Analysis (2D-FDA) [53].

Principal Component Analysis (PCA) is a multi-element unsupervised technique with the approach for dimensionality reduction [75]. In PCA Eigen vectors with largest Eigen values are computed to form the feature space. Kernel PCA is a solution for nonlinear feature extraction [59], [58]. Other nonlinear feature extraction techniques are Multidimensional scaling (MDS) and Kohonen feature Map [40]. Discriminant Analysis is more efficient as compared to PCA in terms of accuracy and time.

2.3. Structural Model

In structural approach of pattern recognition a collection of complex patterns are described by a number of sub-patterns. This model is concerned with structure and attempts to recognize a pattern from its general form. The structural description of patterns in terms of pattern primitives is collectively termed as pattern description language. Selection of type of grammar for pattern description depends on the primitives and the grammar’s descriptive power and analysis efficiency [7]. Number of languages have been suggested for the description of patterns such as chromosome images, 2D-mathematics, chemical
structures, spoken words, English characters and fingerprint patterns. High dimensional patterns need high dimensional grammars such as web grammars, tree grammars, graph grammars and shape grammars for efficient description [6]. Structural model is used for textured images, shape analysis of contours and image interpretation where patterns have a definite structure [9].

2.4. Template Matching Model

Template matching is simplest and most primitive among all pattern recognition models. It is used to determine the similarity between patterns. The pattern to be recognized is matched with the stored templates while assuming that template can be gone through rotational or scalar changes. Correlation function is used as recognition function and is optimized depending on the available training set. The efficiency of this model depends on the stored templates. The shortcoming of this approach is that, it does not work efficiently in the presence of distorted patterns.

2.5. Neural Network Based Model

The most commonly used family of neural networks for pattern classification is the feed forward networks like Multi Layer Perceptrons
(MLP) and Radial Bias Function (RBF) networks [46]. The ability of multilayer neural networks to learn from large training sets makes them obvious for object recognition. The type of network used depends on the requirement of the application. Feed Forward Back-propagation Neural Network (FFBP-NN) is used to implement non-linear differentiable functions. General Regression Neural Network (GRNN) is a highly parallel structure which performs efficiently on noisy data than Back-propagation. FFBP Neural Network does not work accurately if available data is large enough. On the other hand in GRNN, as the size of data increases, the error approaches towards zero [34]. Increasing the number of hidden layers up to a certain extent enhances the performance of the neural networks. The number of neurons must be large enough to adequately represent the problem domain and at the same time small enough to permit generalization [20].

Rowley, Baluja and Kanade [56] presented a neural network-based face detection system where a fully connected neural network examines sub-windows of an image and decides whether each sub-window contains a face. Bootstrapping of multiple networks is used to improve detection performance over a single network, incorporating false detections into the training set as training progresses. Convolution neural networks are a kind of multilayer network, trained with the back-
propagation algorithm. They are designed to recognize visual patterns directly from images with minimal pre processing. They can recognise patterns with extreme variability, and with robustness to distortions and geometric transformations. Discriminative models have been proven to outperform their generative counter parts in terms of their ability to discriminate between different categories of objects.

2.6. Fuzzy Based Model

Kittler [61] states the intimate relation between theory of fuzzy sets and theory of Pattern Recognition and classification rests on the fact that most real world classes are fuzzy in nature. Semantic techniques are used when fuzzy partitions of data sets are to be produced. Then a similarity measure based on weighted distance is used to obtain similarity degree between the fuzzy description of unknown shape and reference shape.

2.7. Feature-based Model

The central idea of feature-based object recognition algorithms lies in finding interest points that are invariant to change due to scale, illumination and affine transformation. The Scale-Invariant Feature
Transform (SIFT) descriptor, proposed by Lowe, is one of the most widely used feature representation schemes for vision applications [69]. Objects can be indexed and recognized using the key points in images. Numerous applications have been developed using the SIFT descriptors, including object retrieval and object category discovery. SIFT-based methods perform better for objects with sufficient number of key points that can be extracted.

The simplest approach to classifying an image is to treat it as a collection of regions. This approach was inspired by the document retrieval community, who can determine the subject or theme of a document by analysing its keywords, without regard to their ordering in a document. Such models are called 'bag of words' models, since a document is represented as a distribution of a fixed vocabulary of words. In the visual domain, building blocks of objects, even without any geometric information, can give a lot of meaning to visual objects, just as keywords can give a sense of the type of document. For each object, it is necessary to generate a dictionary of these building blocks. This is done by analysing an image and counting how many of each building block is in the image. The histogram of this dictionary becomes the representation of the image. Popular algorithms which implement the
bag of words model are naive Bayes classifiers and hierarchical bayesian models.

### 2.8. Part-based Models

The part-based model represents an object as a combination of its appearance and the geometric relationship between its parts. The concept of pictorial structure models were proposed in the early seventies by Fischler and Elschlager [5]. Their basic idea was to represent an object by a collection of parts in a deformable arrangement with links between them. These models were suitable for the descriptions of an object’s appearance and for generic recognition tasks. Agarwal et al. [68] presented an approach to detect objects in grey images based on their part-based representation. Images are represented using parts from the vocabulary, together with spatial relations observed between them. The representation, which encodes information on both shape and appearance was first introduced by Burl [52] and developed further by Weber [65] and Fergus [74]. An object is represented as a set of regions, with information on both appearance and geometric information.

The advantage of a part-based model is that it is computationally easier to deal with a relatively small number of parts than the entire set of pixels. However, a major drawback of the part-based approach is that
the model doesn’t quite express the object in its entirety since it focuses on small parts of the image. It is difficult to classify an image based on this representation. Part-based models often fail to compete with geometry-free bag of features models because the latter technique makes better use of the available image information. Prior knowledge can be used in a natural way to tackle this problem. Fei-Fei et al. [73] proposed a Bayesian framework to use priors such as (1) how parts should come together, and (2) how much parts can vary in order to still be valid, which are derived from the previously learned categories in order to speed up the learning of a new class. Fei-Fei et al. showed that in a set of experiments on four categories, as little as one to three training examples were sufficient to learn a new category. Once a model has been trained, recognition is then performed by first detecting regions, and then evaluating these regions using the model parameters estimated during the training stage.

2.9. Nearest Neighbour Methods

In the late seventies, Rosch et al. [8] suggested that object categories are defined by prototype similarity, as opposed to feature lists. The k nearest neighbours (K-NN) classifier needs a labelled training data set \( \{X,Y\} = \{(x_i,y_i), \ldots, (x_n,y_n)\} \) which consists of \( d \) dimensional feature
vectors $x_i$ and their class labels $y_i$. In order to classify a new feature vector $x$ for $k=1$, it locates the closest element $x_i$ in $X$ and assigns the label $y_i$ to $x$.

The advantage of a nearest neighbour framework is that scaling to a large number of categories does not require adding more features, because the distance function need only be defined for similar enough objects. Training can proceed with relatively few examples by building intra-class variation into the distance function because the object categories are defined on similarity [18]. The drawback of nearest neighbour methods is that they suffer from the problem of high variance in the case of limited sampling.

### 2.10. Support Vector Machines

Support Vector Machines (SVM) has been suggested as a new technique for pattern recognition. For binary classification, a linear SVM tries to find an N-dimensional hyper plane which optimally separates the two classes. The closest points to the hyper plane are called the support vectors. If the input data are not linearly separable a non-linear transformation can be applied which maps the data points of the input space into a high dimensional feature space. To carry out object recognition, the hierarchical SVMs learning technique is offered as an
alternative by Katarina Mele and Jasna Maver. With the help of the hierarchical structures the crisis of cluttered background can be prevented. Hierarchical SVMs allow us to identify the sides of the object. A representation tree with more levels is necessary for the complex 3D objects whereas symmetrical objects do not require hierarchical representation [55]. In [28], Vapnik et al. showed that for optical character recognition, SVMs can demonstrate better generalization ability than other learning algorithms.

2.11. Hybrid Model

Primitive approaches to design a Pattern Recognition system which aims at utilizing a best individual classifier have some drawbacks [57]. To solve hybrid problems Statistical and Structural models can be combined together. Fu [9] gave the concept of attributed grammars which unifies statistical and structural pattern recognition approach. One can also use a set of individual classifiers and combiner to make the final decision which improves the performance of the system. Multiple classifiers can also be used in several ways to enhance the system performance. Each classifier can be trained in a different region of feature space or in other way, each classifier can provide probability estimate and decision can be made upon analyzing individual results.
2.12. Shape Representation and Description Techniques

Shape is an important visual feature and it is one of the basic features used to describe the content of the image [76]. Shape representation and description techniques can be generally classified into two classes of methods: contour-based methods and region-based methods. This classification is based on whether shape features are extracted from the contour only or are extracted from the whole shape region [66]. These classes are further divided into structural and global approaches based on whether the shape is represented as a whole or represented by segments or sections called primitives. These approaches can be further distinguished into space domain and transform domain, based on whether the shape features are derived from the spatial domain or from the transformed domain.

2.12.1. Spectral Transform

Spectral descriptors overcome the problem of noise sensitivity and boundary variations by analyzing shape in spectral domain. Spectral descriptors include Fourier descriptor (FD) and wavelet descriptor (WD) [45], they are derived from spectral transforms on 1-D shape signatures. One of the most widely used shape description methods is FD [10]. Conventional FD methods only deal with closed curve, however, Lin et
al. and Mitchell et al. used FD to describe partial shapes [13]. The Fourier invariants which describe the rotational symmetry of shapes [4] was introduced by Granlund. Rauber proposed a UNL FD (named after Universidade Nova de Lisboa, Portugal) used to describe disjoint or articulated contour shape [37]. The UNL FD is acquired by applying 2-D Fourier transform on the UNL transformed shape image. Richard and Hemami introduced a complex distance measurement, called the true distance measurement, for measuring the similarity between two set of FDs. This distance measurement is the weighting sum of the variance of magnitude ratios and the variance of phase difference between two sets of Fourier coefficients. Since the true distance measurement requires two Fourier transforms for each matching, it requires 15 times more computation than a normal distance measurement.

Recently, several researchers have proposed the use of WD for shape description [57]. Although WD has the advantage over FD in that it is of multi-resolution in both spatial space and spectral space, the increase of spatial resolution will certainly sacrifice frequency resolution [31]. WD suffers the same drawback in primitive determination as that in the structural approach. The advantages of FD over many other shape descriptors are (i) simple to compute (ii) each descriptor has special physical meaning (iii) simple to do normalization, making shape
matching a simple task (iv) captures both global and local features. With sufficient features for selection, FD overcomes the weak discrimination ability of those simple global descriptors.

**2.12.2. Rajan Transform**

Rajan Transform introduced in the year 1997 by Rajan [50] is a coding morphism by which a number sequence (integer, rational, real or complex) of length equal to any power of two is transformed into a highly correlated number sequence of the same length. It is a homomorphism that maps a set consisting of a number sequence, its graphical inverse and their cyclic and dyadic permutations, to a set consisting of a unique number sequence ensuring the invariance property under such permutations. Given a number sequence \( x(n) \) of length \( N \), which is a power of 2, first it is divided into the first half and the second half each consisting of \( (N/2) \) points so that the following hold good.

\[
\begin{align*}
g(i) &= x(i)+x(i+N/2) ; \quad 0 \leq j \leq N/2 ; \quad 0 \leq i \leq N/2 \ldots (1) \\
h(j) &= |x(i)-x(i-N/2)| ; \quad 0 \leq j \leq N/2 ; \quad 0 \leq i \leq N/2 \ldots (2)
\end{align*}
\]

Now each \( (N/2) \) point segment is further divided into two halves each consisting of \( N/4 \) points so that the following hold good.

\[
\begin{align*}
g1(k) &= g(i)+g(j+(N/4)) ; \quad 0 \leq k \leq N/4 ; \quad 0 \leq j \leq N/4 \ldots (3) \\
g2(k) &= |g(i)-g(i-(N/4))| ; \quad 0 \leq k \leq N/4 ; \quad (N/4) \leq j \leq N/2 \ldots (4) \\
h1(k) &= h(j)+h(j+(N/4)) ; \quad 0 \leq k \leq N/4 ; \quad 0 \leq j \leq N/4 \ldots (5) \\
h2(k) &= |h(j)-h(j-N/4)| ; \quad 0 \leq k \leq N/4 ; \quad 0 \leq j \leq N/4 \ldots (6)
\end{align*}
\]
This process is continued till no more division is possible. The total number of stages thus turns out to be $\log_2 N$.

RT is applicable to any number sequence and it induces an isomorphism in a class of sequence, that is, it maps a domain set consisting of the cyclic and dyadic permutations of a sequence on to a range set consisting of sequence of the form $X(k)E(r)$ where $X(k)$ denotes the permutation invariant RT and $E(r)$ an encryption code corresponding to an element in the domain set.

2.12.3. Chain Code Representation

Chain code introduced in 1961 by Freeman [1] describes an object by a sequence of unit-size line segments with a given orientation. In this approach, an arbitrary curve is represented by a sequence of small vectors of unit length and a limited set of possible. In the implementation, a digital boundary of an image is superimposed with a grid, the boundary points are approximated to the nearest grid point, then a sampled image is obtained. From a selected starting point, a chain code can be generated by using 4-directional or 8-directional chain code. A general chain code with N-directions is also possible. The chain code can be normalised by representing the boundary by the differences in the successive directions in the chain code instead of representing the boundary by relative directions. Chain code derived in this way is not
scale invariant. Although it is possible to scale two similar shapes into the same size, the resulted shape numbers can have a different number of digits, making it impractical to do matching between two shapes. The chain code usually has high dimensions and is sensitive to noise. It is often used as an input to a higher level analysis. Iivarinen and Visa derive a chain code histogram (CCH) for object recognition [43]. The CCH is computed as \( p(k) = nk = n \), where \( n \) is the number of chain code values \( k \) in a chain code and \( n \) is the number of links in a chain code. The CHH is translation and scale-invariant, however, it is only invariant to a rotation of 90°. Therefore, the normalized CHH (NCHH) is proposed. It is defined as \( p(k) = lknk = l \), where \( nk \) is the same as in CHH, \( l \) is the length of the direction \( k \) and \( l \) is the length of the contour. Although CHH reduces the dimensions of chain code representation, it does not solve the noise sensitivity problem.

2.12.4. Geometric Moment Invariants

The approach initiated by Hu on the use of image moment invariants for two-dimensional pattern recognition applications [2] is based on the work of the 19th century mathematicians Boole, Cayley and Sylvester, and on the theory of algebraic forms:

\[
\sum \sum \quad , \quad , \quad = 0,1,2, \ldots \quad (7)
\]
Using nonlinear combinations of the lower order moments, a set of moment invariants usually called geometric moment, which has the desirable properties of being invariant under translation, scaling and rotation, are derived. Since the values of the moment invariants are usually very small, a normalization process is needed in the implementation. Geometric moment invariants have attracted wide attention and have been used in many applications. The main problem with geometric moments is that only a few invariants derived from lower order moments is not sufficient to accurately describe shape. Higher order invariants are difficult to derive. Lu have tested geometric moment invariants on a standard shape database used by MPEG-7 and have found that geometric moment invariants perform very well on similarity transformed and affine transformed contour-based shapes. They perform poorly for arbitrarily distorted contour-based shapes. For region-based shapes they only perform satisfactorily on rotated shapes, while for scaled shapes perspective transformed shapes they perform poorly. The finding indicates that geometric moment invariants are suitable for describing simple shapes.

2.12.5. Algebraic Moment Invariants

Algebraic moment invariants have been introduced by Taubin and Cooper [25]. The algebraic moment invariants are computed from the
first m central moments and are given as the Eigen values of predefined matrices, M, whose elements are scaled factors of the central moments. Different from Hu’s geometric moment invariants, the algebraic moment invariants can be constructed up to arbitrary order and are invariant to affine transformations. However it is found that they tend to work well on objects where the distribution of the pixels and not the outline of the shape is important. On objects where the configuration of the outline is important algebraic moment invariants perform poorly.

Statistical models are the best choice for noisy patterns. For complex patterns and applications utilizing large number of pattern classes, it is beneficial to describe each pattern in terms of its components [21]. A wise decision regarding the selection of pattern grammar influences computations efficiency of recognition system. Pattern primitives and pattern grammar to be utilized depends upon the application requirements.

Contour-based approaches are more popular than region-based approaches in literature. This is because human beings are thought to discriminate shapes mainly by their contour features. Another reason is because in many of the shape applications, the shape contour is the only interest while the shape interior content is not important. As each model has its own pros and cons for complex applications it is beneficial to
append two or more recognition models at various stages of recognition process.

Our research focuses on recognition of objects by detecting the contours of images and by applying the hybrid model combining Syntactic method for the description of objects and Neural Network based classification. The entire process is discussed in the following chapters.