ISBN: 978-81-927230-0-9

Pattern Recognition Methods in Image Processing

Tripti Dongare-Mahabare

Department of Computer Science
Pratibha College of Commerce & Computer Studies,
Chinchwad Pune
tru.neel@gmail.com

Abstract - Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant detail. Image processing and recognition is carried out on the actual image transformation and so as to achieve the aim of identification. Because of the characteristic of the image information is that it is a two-dimensional space, so the amount of information it contains is very large. Neural network image recognition technology is the modern computer technology, image processing, artificial intelligence, pattern recognition theory developed a new kind of image recognition technology. Before the image recognition need to use digital image processing techniques for image preprocessing and feature extraction. With the theory of artificial intelligence research and the development of computer technology, the application of neural network in image pattern recognition research is increasingly active. Extensive research and development has taken place over the last 20 years in the areas of pattern recognition and image processing. Areas to which these disciplines have been applied include business (e.g., character recognition), medicine (diagnosis, abnormality detection), automation (robot vision), military intelligence, communications (data compression, speech recognition),and many others. This paper presents a very brief survey of developments in basic pattern recognition and image processing techniques.

Keywords: Image Processing, Image segmentation, Pattern recognition, Object Recognition, Neural Network

I. INTRODUCTION

Pattern recognition, a field concerned with machine recognition of meaningful regularities in noisy or complex environments". The word pattern is derived from the same root as the word patron and, in its original use, means something which is set up as a perfect example to be imitated. Thus pattern recognition means the identification of the ideal which a given object was made after. pattern recognition is a search for structure in data. Pattern recognition (PR) is the science that concerns the description or classification (recognition) of measurements. DURING the past twenty years, there has been a considerable growth of interest in problems of pattern recognition and image processing. Although pattern recognition and image processing have developed as two separate disciplines, they are very closely related. "pattern recognition consists of two parts: feature selection and classifier design."

The area of image processing consists not only of coding, filtering, enhancement, and restoration, but also analysis and recognition of images. On the other hand, the area of pattern recognition includes not only feature extraction and classification, but also preprocessing and description of patterns. It is true that image processing appears to consider only two-dimensional pictorial patterns and pattern recognition deals with one-dimensional, two-dimensional, and three-dimensional patterns in general. However, in many cases, information about one dimensional and three-dimensional patterns is easily expressed as two-dimensional pictures, so that they are actually treated as pictorial patterns.

In order to provide an effective and efficient description of patterns, preprocessing is often required to-remove noise and redundancy in the measurements. Then a set of characteristic measurements, which could be numerical and/or non numerical, and relations among these measurements, are extracted for the representation of patterns. Classification and/or description of the patterns with respect to a specific goal is performed on the basis of the representation. In order to determine a good set of characteristic measurements and their relations for the representation of patterns so good recognition performance can be expected, a careful analysis of the patterns under study is necessary. Knowledge about the statistical and structural characteristics of patterns should be fully utilized. From this point of view, the study of pattern recognition includes both the analysis of pattern characteristics and the design of recognition systems. The many different mathematical techniques used to solve pattern recognition problems may be grouped into two general approaches. They are:

1) The decision-theoretic (or discriminant) approach and
2) The syntactic (or structural) approach.

In the decision-theoretic approach, a set of characteristic measurements, called features, are extracted from the patterns. Each pattern is represented by a feature vector, and the recognition of each pattern is usually made by partitioning the feature space. On the other hand, in the syntactic approach, each pattern is expressed as a composition of its components, called subpatterns or pattern primitives. This approach draws an analogy between the structure of patterns and the syntax of a language. The recognition of each pattern is usually made by parsing the pattern structure according to a given set of syntax rules. In

ISBN: 978-81-927230-0-9

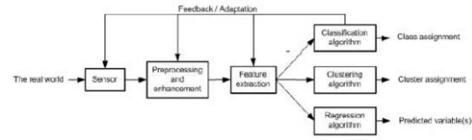
some applications, both of these approaches may be used. For example, in a problem dealing with complex patterns, the decision-theoretic approach is usually effective in the recognition of pattern primitives, and the syntactic approach is then used for the recognition of subpatterns and of the pattern itself.

A. **Decision- Theoretic Methods:** A block diagram of a decision-theoretic pattern recognition system is shown in Fig. 1. The upper half of the diagram represents the recognition part, and the lower half the analysis part. The process of preprocessing is usually treated in the area of signal and image processing. Our discussions are limited to the feature extraction and selection, and the classification and learning. Several more extensive surveys on this subject have also appeared recently .

Feature extraction and selection: Recent developments in feature extraction and selection fall into the following two major approaches.

1)Feature space transformation: The purpose of this approach is to transform the original feature space into lower dimensional spaces for pattern representation and/or class discrimination. For pattern representation, least mean-square error and entropy criteria are often used as optimization criteria in determining the best transformation. For class discrimination, the maximization of interclass distances and/or the minimization of intraclass distances is often suggested as an optimization criterion. Both linear and nonlinear transformations have been suggested. Fourier, Walsh-Hadamard, and Haar transforms have been suggested for generating pattern features. The Karhunen-Loeve expansion and the method of principal components have been used quite often in practical applications for reducing the dimensionality of feature space. In terms of the enhancement of class separability, nonlinear transformations are in general superior to linear transformations. A good class separation in feature space will certainly result in a simple classifier structure (e.g., a linear classifier). However, the implementation of nonlinear transformations usually requires complex computations compared with that of linear transformations. Results of transformations need to be updated when new pattern samples are taken into consideration. Iterative algorithms and/or interactive procedures are often suggested for implementing nonlinear transformations. In some cases, the results of transformations based on pattern representation and class discrimination respectively are in conflict. An optimization criterion for feature space transformation should be able to reflect the true performance of the recognition system. Some recent work appears to move in this direction.

2) Information and distance measures: The main goal of feature selection is to select a subset of 1 features from a given set of N features (l<N) without significantly degrading the performance of the recognition system, that is, the probability of misrecognition, or more generally, the risk of decision. Unfortunately, a direct calculation of the probability of misrecognition is often impossible or impractical partially due to the lack of general analytic expressions which are simple enough to be treated. One approach is to find indirect criteria to serve as a guide' for feature selection.



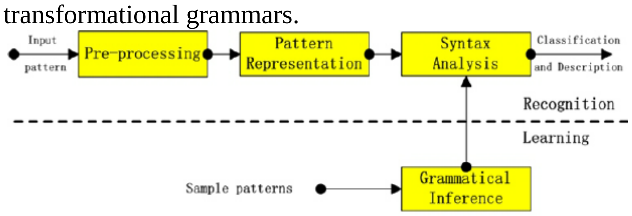
The most common approach is to define an information or (statistical) distance measure, which is related to the upper and/or lower bounds on the probability of misrecognition, for feature selection. That is, the best feature subset is selected in the sense of maximizing a prespecified information or distance measure. Recently, Kanal provided a fairly complete list of distance measures and their corresponding error bounds. Assuming that the most important characteristic of the distance measure is the resultant upper bound on the probability of misrecognition, the various measures can be arranged in increasing order of importance. For a two class recognition problem, denoting the upper bound on the probability of misrecognition by Pe, for Bhattacharyya's distance by UB, for Matusita distance by UM, for equivocation by UE, for Vajda's entropy by Uv, for Devijver's Bayesian distance by UD, for Ito's measure (for n = 0) by UI, for Kolmogorov's variational distance by UK, and for the MO-distance of Toussaint by UT, the following point-wise relations hold:

Pe = UK <=UV = UD = UI = UT <= UE <=UB = UM.

The divergence and Kullback-Leibler numbers, which are simply related to each other, are excluded from the ordering because of the lack of a known upper bound except for the case of a normal distribution where its bound is larger than UB. In terms of computational difficulty, however, the divergences and Bhattacharyya distance are easier to compute than the other distance measures. It is interesting that the best bound on the probability of misrecognition derived from the distance measures is equal to the asymptotic error of the single nearest neighbor classifier. In addition to the information and distance measures mentioned above, a generalized Kolmogorov distance, called the Ja, separability measure, was recently proposed as a feature selection criterion, and its upper and lower bounds on the probability of misrecognition derived. When a = 1, J, is equivalent to the Kolmogorov distance. For a = 2, the upper bound of the probability of misrecognition is equal to the asymptotic probability of error of the single nearest neighbor classifier.

Classification and learning: Most developments in pattern recognition involve classification and learning. When the conditional probability density functions of the feature vectors for each class (which we may call the class density functions) are known or can be accurately estimated, the Bayes classification rule that minimizes the average risk or the probability of misrecognition can be derived. When the class density functions are unknown, nonparametric classification schemes need to be used. In practice, when a large number of pattern samples is available, class density functions can be estimated or learned from the samples, and then an optimal classification rule can be obtained. If the parametric form of each class density function is known, only parameters need to be learned from the pattern samples. When the number of available pattern samples is small, the performance of density and parameter estimations is poor. Nonparametric classification schemes usually suggest a direct learning of the classification rule from pattern samples, for example, the learning of parameters of a decision boundary. Depending upon whether or not the correct classiflcation of the available pattern samples is known, the process of learning can be classified into supervised learning (or learning with a teacher) and nonsupervised learning (or learning without a teacher). Bayesian estimation and stochastic approximation and the potential function method have been suggested for the learning of class density functions or a decision boundary. When the learning is nonsupervised, a mixture density function can be formed from all the individual class density functions and a priori class probabilities. Nonsupervised learning of the parameters of each class density function can be treated as a supervised learning of parameters of the mixture density function from the unclassified pattern samples followed by a decomposition procedure. Under certain conditions, the decomposition can be accomplished and the estimates of the parameters of each class recovered. A related topic which has received an increasing amount of attention recently is learning with finite memory. When the a priori information is sufficient, the classifier may be able to make decisions with good performance. In this case, the learning process could be carried out using the classifier's own decisions; that is, the unclassified pattern samples are now classified by the classifier itself. This type of nonsupervised learning is called decisiondirected learning. When the classification of the pattern samples is incompletely known, learning with an imperfect teacher and learning with a probabilistic teacher have recently been proposed . An appropriate combination of supervised and nonsupervised modes of learning could result in a system of lower cost than those using a single learning mode. Classification based on clustering analysis has been regarded as a practically attractive approach, particularly in a nonsupervised situation with the number of classes not precisely known. Various similarity and (deterministic) distance measures have been suggested as criteria for clustering pattern samples in the feature space. Both hierarchical and nonhierarchical strategies are proposed for the clustering process. Often, some of the clustering parameters, such as the similarity measure and threshold, criteria for merging and/or splitting clusters, etc., need to be selected heuristically or through an interactive technique. It should be interesting to relate directly the distance measures for feature selection to those for clustering analysis. Recently, clustering algorithms using adaptive distance were proposed. The similarity measure used in the clustering process varies according to the structure of the clusters already observed. Mode estimation, least mean-square optimization, graph theory and combinatorial optimization have been used as a possible theoretical basis for clustering analysis. Nevertheless, clustering analysis, at its present state-of-the-art, still appears to be an experiment-oriented "art."

B. **Syntactic (or Structural) Methods** A block diagram of a syntactic pattern recognition system is shown in Fig. 2. Again, we divide the block diagram into the recognition part and the analysis part, where the recognition part consists of preprocessing, primitive extraction (including relations among primitives and subpatterns), and syntax (or structural) analysis, and the analysis part includes primitive selection and grammatical, (or structural) inference. In syntactic methods, a pattern is represented by a sentence in a language which is specified by a grammar. The language which provides the structural description of patterns, in terms of a set of pattern primitives and their composition relations, is sometimes called the "pattern description language." The rules governing the composition of primitives into patterns are specified by the socalled "pattern grammar." An alternative representation of the structural information of a pattern is to use a "relational graph," of which the nodes represent the subpatterns and the branches represent the relations between subpatterns. Primitive extraction and selection: Since pattern primitives are the basic components of a pattern, presumably they are easy to recognize. Unfortunately, this is not necessarily the case in some practical applications. For example, strokes are considered good primitives for script handwriting, and so are phonemes for continuous speech; however, neither strokes nor phonemes can easily be extracted by machine. The segmentation problems for script handwriting and continuous speech, respectively, are still subjects of research. An approach to waveform segmentation through functional approximation has recently been reported. Segmentation of pictorial patterns is discussed in Section III under Segmentation. There is no general solution for the primitive selection problem at this time. For line patterns or patterns described by boundaries or skeletons, line segments are often suggested as primitives. A straight line segment could be characterized by the locations of its beginning (tail) and end (head), its length, and/or slope. Similarly, a curve segment might be described in terms of its head and tail and its curvature. The information characterizing the primitives can be considered as their associated semantic information or as features used for primitive recognition. Through the structural description and the semantic specification of a pattern, the semantic information associated with its subpatterns or the pattern itself can then be determined. For pattern description in terms of regions, half-planes have been proposed as primitives. Shape and texture measurements are often used for the description of regions; see Section III under Properties. Pattern grammars: After pattern primitives are selected, the next step is the construction of a grammar (or grammars) which will generate a language (or languages) to describe the patterns under study. It is known that increased descriptive power of a language is paid for in terms of increased complexity of the syntax analysis system (recognizer or acceptor). Finite-state automata are capable of recognizing finite-state languages although the descriptive power of finite-state languages is also known to be weaker than that of context-free and context-sensitive languages. On the other hand, nonfinite, nondeterministic procedures are required, in general, to recognize languages generated by context-free and context-sensitive grammars. The selection of a particular grammar for pattern description is affected by the primitives selected, and by the tradeoff between the grammar's descriptive power and analysis efficiency. Context-free programmed grammars, which maintain the simplicity of context-free grammars but can generate context-sensitive languages, have recently been suggested for pattern description. A number of special languages have been proposed for the description of patterns such as English and Chinese characters, chromosome images, spark chamber pictures, two-dimensional mathematics, chemical structures, spoken words, and fingerprint patterns. For the purpose of effectively describing high dimensional patterns, high dimensional grammars such as web grammars, graph grammars, tree grammars, and shape grammars have been used for syntactic pattern recognition. Initial applications include fingerprint pattern recognition and the interpretation of Earth Resources Technology Satellite data. Ideally speaking, it would be nice to have a grammatical (or structural) inference machine which would infer a grammar or structural description from a given set of patterns. Unfortunately, such a machine has not been available except for some very special cases. In most cases so far, the designer constructs the grammar based on the a priori knowledge available and his experience. -In some practical applications, a certain amount of uncertainty exists in the process under study. For example, due to the presence of noise and variation in the pattern measurements, segmentation error and primitive extraction error may occur, causing ambiguities in the pattern description languages. In order to describe noisy and distorted patterns under ambiguous situations, the use of stochastic languages has been suggested.With probabilities associated with grammar rules, a stochastic grammar generates sentences with a probability distribution. The probability distribution of the sentences can be used to model the noisy situations. Other approaches for the description of noisy and distorted patterns using syntactic methods include the use of approximation and transformational grammars.



The effectiveness of these approaches remains to be developed and tested. Syntactic recognition: Conceptually, the simplest form of recognition is probably "template-matching." The sentence describing an input pattern is matched against sentences representing each prototype or reference pattern. Based on a selected "matching" or "similarity" criterion, the input pattern is classified in the same class as the prototype pattern which is the "best" to match the input. The structural information is not recovered. If a complete pattern description is required for recognition, a parsing or syntax analysis is necessary. In between the two extreme situations, there are a number of intermediate approaches. For example, a series of tests can be designed to test the occurrence or nonoccurrence of certain subpatterns (or primitives) or certain combinations of them. The result of the tests, through a table lookup, a decision tree, or a logical operation, is used for a classification decision. A parsing procedure for recognition is, in general, nondeterministic and, hence, is regarded as computationally inefficient. Efficient parsing could be achieved by using special classes of languages such as finite-state and deterministic languages for pattern description. The tradeoff here between the descriptive power of the pattern grammar and its parsing efficiency is very much like that between the feature space selected and the classifier's discrimination power in a decision-theoretic recognition system. Special parsers using sequential procedures or other heuristic means for efficiency improvement in syntactic pattern recognition have recently been constructed. Error-correcting parsers have been proposed for the recognition of noisy and distorted patterns. Different types of segmentation and primitive extraction errors (substitution, deletion and addition) are introduced into the pattern grammar. The recognition process is then based on the parsers designed according to the expanded pattern grammar. The error-correcting capability is achieved by using a minimum-distance criterion. Since the original grammar is expanded to include all possible error situations, the parser so designed is less efficient than that designed according to the original grammar. This tradeoff between error-correcting capability and parsing efficiency seems to be expected. Nevertheless, it could be a very serious drawback in practical applications. When stochastic grammars are used for pattern description, the probability information is useful in resolving ambiguous situations. For example, if a sentence is found to be generated by two different pattern grammars, the ambiguity can be resolved by comparing the generation probabilities of the sentence in the two grammars. A maximum-likelihood or Bayes decision rule based on the two generation probabilities will yield the final recognition. Besides, the probability information can also be utilized to speed up the parsing process. The use of a sequential decision procedure could result in further reducing the parsing time by slightly increasing the probability of misrecognition. Of course, when a sequential procedure is used, the parsing procedure stops most of the time before a sentence is completely scanned, and, consequently, in these cases the complete structural information on the pattern cannot be recovered. Remarks: Compared with decision-theoretic pattern recognition, syntactic pattern recognition is a newer area of research. When the patterns are complex and the number of pattern class is very large, it would be advantageous to describe each pattern in terms of its components and to consider description and classification of

ISBN: 978-81-927230-0-9

patterns rather than classification only. Of course, the practical utility of the syntactic approach depends upon our ability to recognize the simple pattern primitives and their relationships represented by the composition operations. As research in both decision-theoretic and syntactic approaches is still in progress, heuristic methods are also being developed for specific purposes. New approaches proposed recently for pattern recognition include variable-value logic and relation theory. The effectiveness of these approaches still has to be tested.

II. CONCLUSION

There was an unbalanced development between theory and practice in pattern recognition. Many theoretical results, especially in connection with the decision-theoretic approach, have been published. Practical applications have been gradually emphasized during the last five years, particularly in medical and remote sensing areas. Pattern recognition is still very much an active research area. Decision-theoretic(Statistical) pattern recognition is based on statistical characterizations of patterns, assuming that the patterns are generated by a probabilistic system. In the decision-theoretic approach, we are still looking for effective and efficient feature extraction and selection techniques, particularly in nonparametric and small sample situations. The computational complexity of pattern recognition systems, in terms of time and memory, should be an interesting subject for investigation. Syntactic(Structural) pattern recognition is based on the structural interrelationships of features. In the syntactic approach, the problem of primitive extraction and selection certainly needs further attention. An appropriate selection of the pattern grammar directly affects the computational complexity or analysis efficiency of the resulting recognition system. Grammatical inference algorithms which are computationally feasible are still highly in demand. In image processing, better models are needed for both the images and their user (the human visual system). Image models should also be used more extensively in the design of optimal image segmentation and feature extraction procedures. Recently there is more scope in field of pattern recognition with fuzzy logic.A wide range of algorithms can be applied for pattern recognition, from very simple Bayesian classifiers to much more powerful neural networks. Typical applications, such as automatic handwrite recognition, automatic image recognition, and geomorphologic terrain features classification are examples form the subtopic image analysis of pattern recognition that deals with digital images as input to pattern recognition systems. Image and visual models need further development in both image processing and recognition.

REFERENCES

1. Introduction to Mathematical Techniques in Pattern Recognition, New York: Wiley, 1972.
2. "Statistical pattern recognition-Review and outlook," Southeastern Massachusetts Univ., North Dartmouth, MA, Tech. Rep. TR-EE 75-4, June 25, 1975.
3. Z. Chen and K. S. Fu, "On the connectivity of clusters," Inform. Sci., vol. 8, pp. 283-299, 1975.
4. G. C. Cheng, D. K. Pollock, R. S. Ledley, and A. Rosenfeld, Eds., Pictorial Pattern Recognition. Washington, DC: Thompson, 1968.
5. H. C. Andrews, Ed., "Special issue on digital picture processing," Comput., vol. 7, pp. 17-87, May 1974.
6. D. B. Cooper, "When should a learning machine ask for help," IEEE Trans. Inform. Theory, vol. IT-20, July 1974.
7. T. Cover, "A hierarchy of probability density function estimates," in Frontiers of Pattern Recognition, S. Wantanabe, Ed. New York: Academic, 1972.
8. "Recent books on pattern recognition," IEEE Trans. Inform. Theory, vol. IT-19, Nov. 1973.
9. T. M. Cover and T. J. Wagner, "Topics in statistical pattern recognition," in Digital Pattern Recognition, K. S. Fu, Ed. New York: Springer, 1976.
10. P. A. Devijver, "Decision-theoretic and related approaches to pattern classification," in Proc. NATO Advanced Study Institute on Pattern Recognition-Theory and Applications, Sept. 8-17, 1975.
11. K. S. Fu, Ed., "Special issue on feature extraction and selection in pattern recognition," IEEE Trans. Comput., vol. C-20, pp. 965-1120, Sept. 1971.
12. Syntactic Methods in Pattern Recognition. New York: Academic, 1974.
13. K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic, 1972.
14. Digital Image Processing3rd Ed. (DIP/3e) by Gonzalez and Woods © 2008