ANATOMY ON PATTERN RECOGNITION

MAYANK PARASHER

Department of Computer science, YMCA University of Science & technology, Sector-6,
Faridabad, Haryana 121006, India
mayankparasher@gmail.com

SHRUTI SHARMA

Department of Computer science, YMCA University of Science & technology, Sector-6,
Faridabad, Haryana 121006, India
shruti_mattu@yahoo.co.in

A.K SHARMA

Department of Computer science, YMCA University of Science & technology, Sector-6,
Faridabad, Haryana 121006, India
ashokkalia123@gmail.com

J. P. GUPTA

Vice Chancellor, JIIT, NOIDA,
UP, India
jpgupta@jiit.in

Abstract

Pattern Recognition is the science of recognizing patterns by machines. This is very wide research area as of today, because every new research tries to make machine as intelligent as human for recognizing patterns. Pattern recognition is an active research and an important trait of 'artificial intelligence'. This review paper introduces pattern recognition, its fundamental definitions, and provides understanding of related research work. This paper presents different types of algorithms, their limitations & applications of pattern recognition.

Keywords**:** Pattern recognition; PCA; Neural Network; Supervised learning; unsupervised learning.

1. Introduction

In Artificial Intelligence Science, Machine Perception is defined as the ability of machine to process visual input data and deduce the aspects of world. Having capability of analyzing visual input data from input device like camera and having capability of making decision on input visual data, gives birth to "*PATTERN RECOGNITION SYSTEM*" on machines, which is the biggest trait of Human beings.

Human ability to recognize patterns is remarkable. A Human can recognize face out of thousand faces, irrespective of varying illumination intensity, varying facial rotation, varying facial expressions, varying facial biometrical changes, varying facial scaling and even with occluded face images.[2][3]

Pattern Recognition is still an ongoing wide research study, which tries to make machine as intelligent as human being for recognizing patterns. Before understanding the system and related algorithms, let's first understand the basic definitions.

1.1. *Pattern*

A pattern is defined by the common denominator among the multiple instances of an entity. For example, commonality in all fingerprint images defines the fingerprint pattern; thus, a pattern could be a fingerprint image, a handwritten cursive word, a human face, a speech signal, a bar code, or a web page on the Internet.[2]

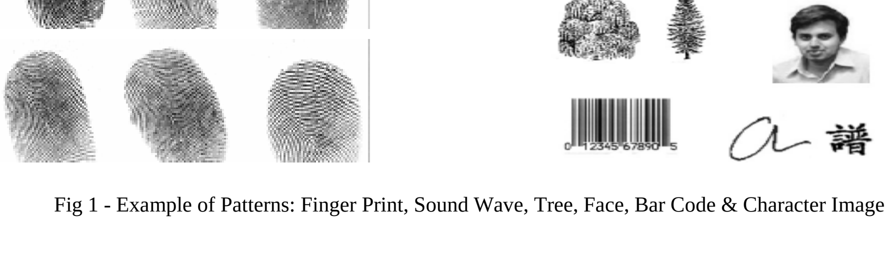

Fig 1 - Example of Patterns: Finger Print, Sound Wave, Tree, Face, Bar Code & Character Image

1.2. *Pattern Class*

Often, individual patterns may be grouped into a category based on their common properties; the resultant group is also a pattern and is often called a pattern class.

1.3. *Pattern Recognition*

Pattern recognition is the science for observing (sensing) the environment, learning to distinguish patterns of interest (e.g., animals) from their background (e.g., sky, trees, ground), and making sound decisions about the patterns or pattern classes (e.g., a dog, a mammal, an animal).

2. Pattern Recognition System

Pattern recognition (PR) is the most important trait of cognitive ability, be it of humans or animals. The ability to recognize patterns is central to intelligent behavior. We receive signals from environment through our sensory organs which are processed by the brain to generate suitable responses. The whole process involves extraction of information from the sensory signals, processing it using the information stored in the brain to reach a decision that induces some action. All these information we work with are represented as patterns. We recognize voices, known faces, scenes, written letters and a multitude of other objects in our everyday life.[5]

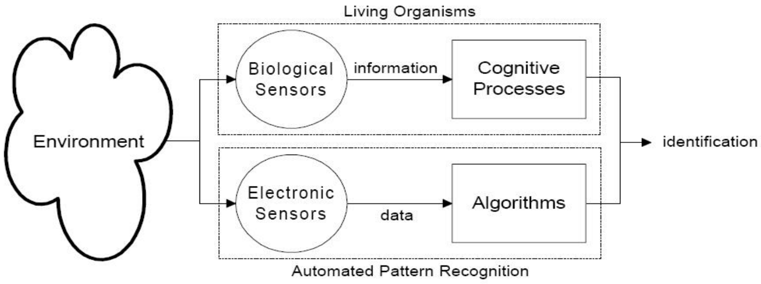

Fig. 2 Pattern Recognition for Human & Machines

Figure 2, gives overview of pattern recognition process of human & machines.
Pattern Recognition System consists of three basic units.

- ✓ **Data Perception** - "Visual input data is perceived from the environment and get pre-processed"
- ✓ **Feature Extraction** - "The process of extracting relevant feature information from processed data"
- ✓ **Classification / Recognition** - "Process of classify/recognize feature information among training data"

In pattern recognition system, first the visual data is captured from the environment using input device like camera / microphone etc. The input data is then being pre processed, so that data being able to read & analyze by computer as per the algorithm. This process can be termed as 'collecting and processing' data from environment.

Then extracting relevant features from the processed data. These relevant features collectively form entity of object to be recognized. Then a classifier is deployed to classify or recognize the input data. This classifier acts upon the descriptor of features extracted.[6][7]

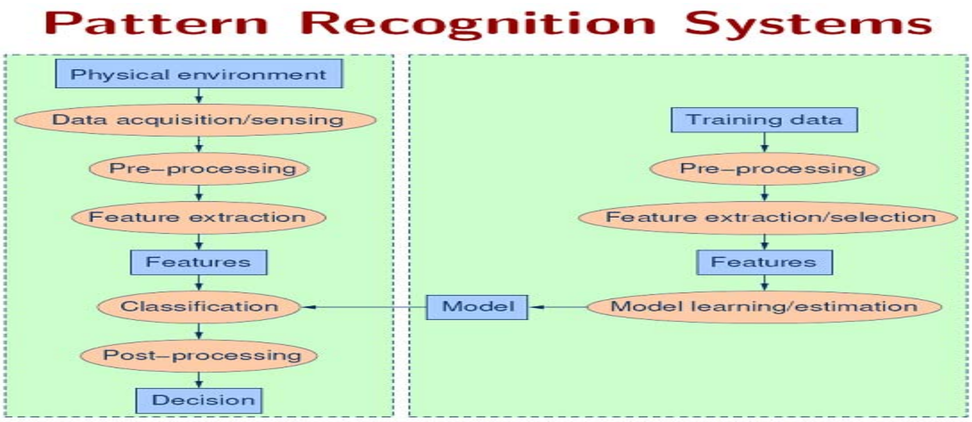

Fig. 3 Pattern recognition System

3. Pattern Recognition Methodologies

There are four major methodologies in pattern recognition; these are statistical approach, syntactic approach, template matching & neural network.[3][4]

3.1. *Statistical Approach*

Typically, statistical pattern recognition systems are based on statistics and probabilities. In these systems, features are converted to numbers which are placed into a vector to represent the pattern. This approach is most intensively used in practice because it is the simplest to handle. In this approach, patterns to be classified are represented by a set of features defining a specific multidimensional vector: by doing so, each pattern is represented by a point in the multidimensional features space. To compare patterns, this approach uses measures by observing distances between points in this statistical space.

3.2. *Syntactic Approach*

Also called structural pattern recognition systems, these systems are based on the relation between features. In this approach, patterns are represented by structures which can take into account more complex relations between features than numerical feature vectors used in statistical PRSs. Patterns are described in hierarchical structure composed of sub-structures composed themselves of smaller sub-structures. The shape is represented with a set of predefined primitives called the codebook and the primitives are called code words. For example, given the code words on the left of figure, the shape on the right of the figure can be represented as the following string S, when starting from the pointed codeword on the figure:

S = d b a b c b a b d b a b c b a b (1).

The system parses the set of extracted features using a kind of predefined grammar. If the whole features extracted from a pattern can be parsed to the grammar then the system has recognized the pattern. Unfortunately, grammar-based syntactic pattern recognition is generally very difficult to handle.

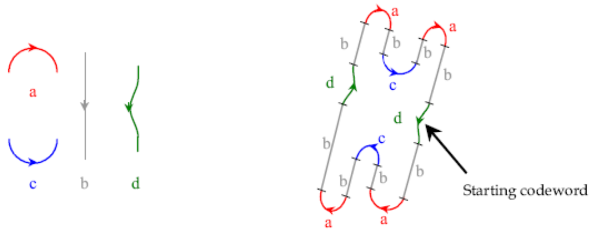

Fig. 4 Example of syntactic description feature

3.3. *Template Matching*

Template matching approach is widely used in image processing to localize and identify shapes in an image. In this approach, one looks for parts in an image which match a template (or model). In visual pattern recognition, one compares the template function to the input image by maximizing the spatial cross-correlation or by minimizing a distance: that provides the matching rate.

The strategy of this approach is: for each possible position (in the image), each possible rotation, or each other geometric transformation of the template, compare each pixel’s neighborhood to this template. After computing the matching rate for each possibility, select the largest one that exceeds a predefined threshold. It is a very expensive operation while dealing with big templates and/or large sets of images.[3][4]

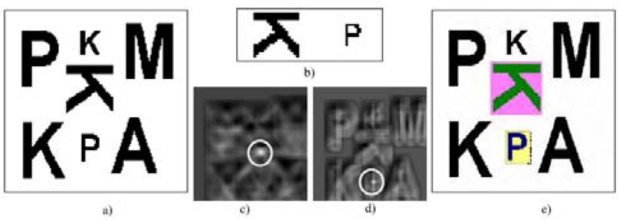

Fig. 5 Illustration of template matching method

3.4. *Neural Network*

Typically, an artificial neural network (ANN) is a self-adaptive trainable process that is able to learn to resolve complex problems based on available knowledge. A set of available data is supplied to the system so that it finds the most adapted function among an allowed class of functions that matches the input. An ANN-based system simulates how the biological brain works: it is composed of interconnected processing elements (PE) that simulate neurons. Using this interconnection (or synapse), each neuron (or PE) can pass information to another. As can be seen on figure 3, these interconnections are not necessarily binary (on or off) but they may have varying weights defined by the weight matrix W: the weight applied to a connection results from the learning process and indicates the importance of the contribution of the preceding neuron in the information being passed to the following neuron.[4]

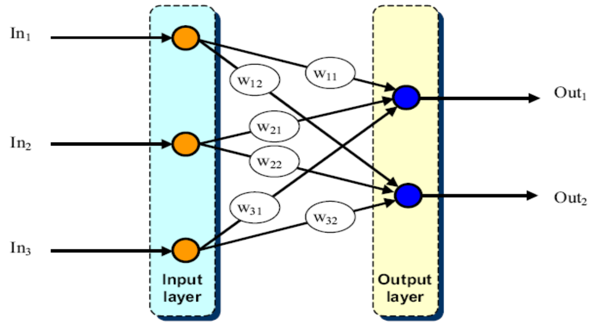

Fig. 6 Example of Neural Network

Table 1 Pattern Recognition Models

Approach	**Representation**	**Recognition function**	**Typical criterion**
Template matching	Sample, pixel, curves	Correlation, distance measures	Classification error
Statistical	Features	Discriminant function	Classification error
Syntactic or structural	Primitives	Rules, grammar	Acceptance error
Neural networks	Samples, pixels, features	Network function	Mean square error

4. Pattern Recognition Algorithms

For pattern recognition, various researches & various algorithms have been proposed. The types of pattern recognition are governed by its design cycle. As we know, it consists of basic elements like "visual perception", "feature extraction" & "classification". There are various different techniques and algorithms to implement these basic elements. So what technique is chosen for each element in design cycle tells the algorithm characteristic of pattern recognition.[2]

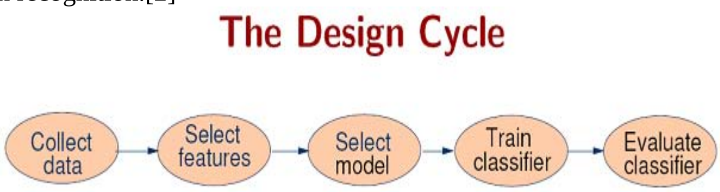

Fig. 7 The design cycle of pattern recognition

Algorithms for pattern recognition depend on the type of label output, on whether learning is supervised or unsupervised, and whether the algorithm is statistical or non-statistical in nature. Statistical algorithm can be further categorized as generative or discriminative.[5]
Before going various types of algorithms, let first understand few definitions.

Statistical Classification: "Statistical classification is the problem of identifying the sub population to which new observation belong"
Supervised Learning: "Supervised learning is the machine learning task of inferring a function from supervised training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier"[6]
Unsupervised Learning: "Un-supervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution."[6]

VARIOUS ALGORITHMS TYPES

4.1. *Classification Algorithms*

Classification algorithms are Statistical algorithms, having supervised learning and predicting categorical labels. Following are the few types of classification algorithms:[5][6]

4.1.1. Support Vector Machines

A Support Vector Machine (SVM) performs classification by constructing an *N*-dimensional hyper plane that optimally separates the data into two categories. A support vector machine (SVM) is a concept in computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

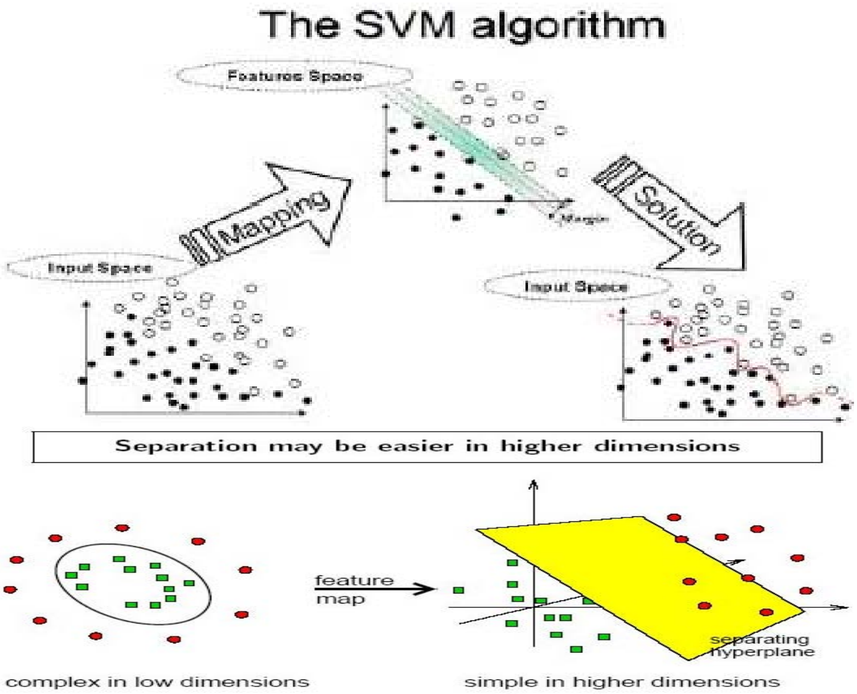

Fig. 8 Example of SVM

4.1.2. Neural Network (Multi Level Perceptron)

Neural network is an interconnected network of neurons. It works on a concept of human brain system and possesses few traits like we can train ANN for best matched solution, ANN can perform fuzzy matching and provides optimal solution. ANN acts as a classifier in pattern recognition.

4.1.3. Kernel Estimation & K-nearest neighbor

In pattern recognition, the *k*-nearest neighbor algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. *k*-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

4.1.4. Maximum Entropy Classifier

Maximum Entropy Classifier is a regression model which generalizes logistic regression by allowing more than two discrete outcomes. That is, it is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables (which may be real-valued, binary-valued, categorical-valued, etc.). Maximum entropy classifiers are commonly used as alternatives to Naive Bayes classifiers because they do not require statistical independence of the independent variables (commonly known as features) that serve as predictors.

4.1.5. Naive Bayes Classifier

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

4.2. *Clustering Algorithms*

4.2.1. KPCA (Kernel Principle Component Analysis)

Kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping. Kernel PCA has been demonstrated to be useful for novelty detection and image de-noising

4.2.2. K-means Clustering

K-means clustering is a method of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean.

4.3. *Regression Algorithms*

In statistics, regression analysis includes any techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. These are statistical algorithms predicting real valued labels but having both supervised & un-supervised learning.
Supervised: Linear Regression, Gaussian Process Regression & Neural Networks
Unsupervised: ICA (Independent Component Analysis) & PCA (Principle Component Analysis)

4.3.1. PCA (Principle Component Analysis)

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA was invented in 1901 by Karl Pearson.[1] Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigen value decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of component scores (the transformed variable values corresponding to a particular case in the data) and loadings (the weight by which each standardized original variable should be multiplied to get the component score). PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from it's (in some sense) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced. [8]

4.4. *Sequence Labeling Algorithms*

Sequence labeling is a type of pattern recognition task that involves the algorithmic assignment of a categorical label to each member of a sequence of observed values. A common example of a sequence labeling task is part of speech tagging, which seeks to assign a part of speech to each word in an input sentence or document. These types of algorithms predicting sequence of categorical labels and real valued labels. These algorithms support both supervised and un-supervised learning.
Supervised Learning for categorical labels: Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM) & Conditional Random Fields (CRF)
Unsupervised learning for categorical labels: Hidden Markov Model (HMM)
Supervised learning for real valued labels: Kalman Filters & Particle Filters

4.5. *Parsing Algorithms*

Parsing, or, more formally, syntactic analysis, is the process of analyzing a text, made of a sequence of tokens (for example, words), to determine its grammatical structure with respect to a given (more or less) formal grammar.
The algorithm supports both supervised & unsupervised learning and predicting tree structured labels.
Example: Probabilistic Context Free Grammars (PCFG)

5. Usage & Application

Pattern recognition is used in any area of science and engineering that studies the structure of observations. It is now frequently used in many applications in manufacturing industry, health care and military.[2]

Examples include:

- Optical character recognition (OCR) is becoming an integral part of document scanners, and is also used frequently in banking and postal applications. Printed characters can now be accurately recognized, and the improving performance of automatic recognition of handwritten cursive characters has diminished significantly the need of human interaction for OCR tasks.

- Automatic speech recognition is very important for user interaction with machines. Commercial systems for automatic response to flight queries, telephone directory assistance and telebanking are available. Often the systems are tuned to a specific speaker for better recognition accuracy.

- Computer vision deals with the recognition of objects as well as the identification and localization of their three-dimensional environments. This capability is required, for example, by robots to operate in dynamic or unknown environments. This can be useful from applications ranging from manufacturing to household cleaning, and even for rescue missions.

- Personal identification systems that use biometrics are very important for security applications in airports, ATMs, shops, hotels, and secure computer access. Recognition can be based on face, fingerprint, iris or voice, and can be combined with the automatic verification of signatures and PIN codes.

- Recognition of objects on earth from the sky (by satellites) or from the air (by airplanes and cruise missiles), is called remote sensing. It is important for cartography, agricultural inspection, detection of minerals and pollution, and target recognition.

- Many tests for medical diagnosis utilize pattern recognition systems, from counting blood cells and recognition of cell tissues through microscopes to the detection of tumors in magnetic resonance scans and the inspection of bones and joints in X-ray images.

- Many large databases are stored on the repositories accessible via Internet or otherwise in local computers. They may have a clear structure such as bank accounts, a weak structure such as consumer behavior, or no obvious structure such as a collection of images. Procedures for finding desired items (database retrieval) as well as to learn or discover structures in databases (data mining) are becoming more and more important. Web search engines and recommender systems are two example applications.

Example of pattern recognition application with its input pattern and pattern class.

Table 2 Examples of Pattern Recognition Applications

Problem Domain	Application	Input Pattern	Pattern Class
Bioinformatics	Sequence Analysis	DNA/Protein sequence	Known type of genes/patterns
Data mining	Searching for meaningful patterns	Points in multi dimension space	Compact and well separated clusters
Document classification	Internet search	Text document	Semantic categories
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters / words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective / non defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres e.g. action, dialogue etc
Biometric recognition	Personal identification	Face, iris & finger print	Authorized user for access control
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory enquiry with operator	Speech waveform	Spoken words

6. References

[1] Jie Liu, Jigui sun Pattern recognition: An overview, 1st edn. Wang. College of computer science & technology, china. IJCSNS, International Journal of Computer Science & Network Security. VOL.6. NO.6. June 2006.

[2] Introduction to pattern recognition; Anil k Jain, Robert P.W Duin (Oxford press)

[3] An overview of advances of pattern recognition system in computer vision; kidiyo kpalma, joseph ronsin; IETR, UMR-CNRS 6164.

[4] Vision System: Segmentation & pattern recognition; Obinata, Dutta. ISBN 987-3-902613-05-9. June 2007.

[5] Statistical pattern recognition: A Review; Robert, Jain, Mao, IEEE transaction on machine learning and pattern analysis, VOL.22. NO.1. January 2000.

[6] Introduction to pattern recognition; Web source by Wikipedia.

[7] Statistical pattern recognition: second edition; Andrew R Webb, ISBN 0-470-84513-9 John Wiley Press

[8] Agarwal, M.; Agrawal, H.; Jain, N.; Kumar, M, Face Recognition Using Principle Component Analysis, Eigenface and Neural Network Digital Object Identifier: 10.1109/ICSAP.2010.51 Publication Year: 2010 , Page(s): 310 – 314

[9] Pudil; Somol; Haindl; Introduction to statistical pattern recognition, Institution of information theory & automation. Academy of science, 18208 praha 8

[10] A Statistical Learning/Pattern Recognition book by Thomas Minka

[11] Zheng, He; Classification technique in pattern recognition; faculty of IT, University of technology, Sydney, ISBN 80-903100-8-7, WSCG 2005, January 31 – February 04, 2005.