



Review Paper

## A Survey on use of Neuro-Cognitive and Probabilistic Paradigms in Pattern Recognition

Khan Y.D.<sup>1,2</sup>, Ahmad F.<sup>2</sup> and Khan S.A.<sup>1</sup>

<sup>1</sup>Department of Computer Science, Abdul Wali Khan University, Mardan, PAKISTAN

<sup>2</sup>Faculty of Information Technology, University of Central Punjab, Lahore, 54500, PAKISTAN

Available online at: [www.isca.in](http://www.isca.in)

Received 30<sup>th</sup> November 2012, revised 27<sup>th</sup> December 2012, accepted 24<sup>th</sup> February 2013

### Abstract

The state of any system is defined by data of some sort. Data by itself is meaningless unless it is transformed into information by putting it into a meaningful context. Images are also defined by data accumulated by quantification of color intensities of pixels. Either the values of the pixels form patterns or the arrangement of pixels form patterns. Various statistical and probabilistic models have been employed by the researchers to extract interesting information from data. Also the pattern recognition problems are divided into two levels termed as generic and specific. In this piece of work we identify problems from both these domains of pattern recognition. This comprehensive survey not only provides a tutorial on the probabilistic and neuro-cognitive techniques used for this purpose but it also explores monumental issues in the domain.

**Keywords:** Probabilistic models, image processing, pattern recognition.

### Introduction

A cognitive process aims to quantify characteristics and attributes of a complex structure with a measurable state. Quantification of any state yields data. To interact with such a structure one needs to manipulate data defined by its state. In the digital world such quantification of data is of finite nature. Hence such data is subjected to various analyses to extract information. Data by itself is meaningless but the information embedded within the data is useful. Information is either extracted from data itself or from the patterns formed by data. Retrieval of information from any type of data requires determination of patterns and trends formed by it<sup>1</sup>. This extraction of information from data is performed by human mind whenever it perceives an object, action or event. Patterns in fact form a method of disbursement of data where occurrence of each pattern connotes an interpretation. The ability to identify these patterns is a powerful tool; it allows reasoning with samples which may previously be unknown. Every day we see new faces but still we identify them as humans, we hear new voices but still we are able to identify its language, we see newly made objects but still we identify their type.

Pattern Recognition in the field of computer science that deals with identification of these obscured patterns in a given set of data. In Image Processing images are used as source of data. Based on the identified patterns a generalization or classification of the image is produced. Researchers over the years have found methods for identifying patterns within the image and formed models to interpret them<sup>2</sup>. Many such applications are found such as biometrics, content based image retrieval, situation

recognition and character recognition. All these applications rely on images as their primary source of data. For any mathematical, stochastic or statistical model to work on it the data can be of finite or even infinite nature<sup>1,3,4</sup>. In the digital world data is certainly of finite nature. A digital image contains numerous pixels which in fact are quantification of various points within the image and are finite in number. Not only the color intensities within the pixels hold information but also the arrangement of pixels in different images provides some information. The color intensities are discrete quantification of the color in real life. The accuracy of this quantification is bounded not just by the precision level of its units but also by the nature of the hardware in use.

A biometric application uses the image of some human biological feature as the data source for identification of individuals established by recognition of peculiar pattern within an acquired image. Various accurate and feasible methods have been designed for this purpose using inimitable features of a person such as fingerprints, facial features, sutures, ears and iris patterns. Such biometric techniques have gained acceptance and popularity for its accuracy and precision. One such technique gaining popularity is iris recognition. Human iris can be uniquely identified by virtue of its textural characteristics. Iris texture is formed by radial and longitudinal muscles which dilate or constrict the pupil as a stimulus to light changes. A reasonably vivid image of iris shows rings, pustules, undulation and stripes forming a peculiar pattern. The essence of iris related identification technique lies in the recognition of these peculiar patterns<sup>5</sup>.

Content Based Image Retrieval forms another interesting field within image processing. The need of efficient searching and indexing methods for visually interesting data has been significantly increasing due to the effortless availability of multimedia devices. This has thrived need for devising algorithms which are efficient and are adequately designed for parameterization of search criterion as per requirements. A simple solution proposed to this problem is to preserve multimedia documents and images in conjunction with foot notes describing the image<sup>6</sup>. However, this technique may often fail to give desired results as the foot note may not be able to describe all characteristics of the given image. Also the user may not be able to describe accurately the contents of an image in form of text, to be used as search criteria. A better approach for retrieval of resembling images from a large image base is to employ image processing techniques to search for possible matches based on the contents of the given image. Therefore, algorithms need to be developed that extract features of an image in the form of parameters and use it as a search criterion. Moreover certain probabilistic models are adapted to find any possible similar patterns between the images being compared. A number of approaches have been proposed in the literature over time to analyze and match images on the basis of its semantic contents. One class of such approaches is stochastic or probabilistic in nature which generally extracts certain local or global semantically significant information from the image by applying various filters. Such type of information describes features within the image and is represented mathematically in form of vectors and matrices. Moreover, statistical and/or stochastic models are defined to homologize the extracted information from an image<sup>7-12</sup>.

## Classification of Pattern Recognition Problems

The literature survey discussed in the last section reveals two different problems both of which are related to pattern recognition. A closer look into the problems establishes a categorical difference between them. Although generally both the problems pertain to the field of pattern recognition but in biometrics a peculiar biological feature is searched while in CBIR a pattern for presence of some arbitrary object is searched. Simply we may say that CBIR problems pertain to detection of an object within the image whereas in biometrics the presence of a certain object exists by principle and its association to a class is established. For example a CBIR application will inform about the presence of an object (a person for example) in an image but will not tell us who that person is while if a biometric application is provided with an image of a person it will tell us who that person is. It is noticed that a CBIR application will provide a higher level classification while a biometric application will provide lower level classification. Based on these general aspects Pattern Recognition problems are categorized into two categories: i. Generic, ii. Specific.

A generic pattern recognition algorithm shall be empowered to identify objects within an image. Any further classification, if required, about the type of object will be the task of a Specific

Pattern Recognition algorithm. In this text we shall discuss CBIR as a general and iris recognition as a specific pattern recognition problem.

**Iris Recognition:** Researchers, over the years, have developed various techniques for iris recognition making use of mathematical and statistical models like wavelet transforms, SVM, Gabor filtering, Laplacian pyramid and other probabilistic methods. In some literature the author presents a technique using a bank of Gabor filters, using local and global iris characteristics, forming a fixed length feature vector<sup>6</sup>. Iris matching is established based on the weighted Euclidean distance between the two iris images being compared. In another article a technique is devised using discrete cosine transform. Iris coding is based on the differences of discrete cosine transform coefficients of overlapped angular patches from normalized iris images<sup>13</sup>. Certain researchers have employed various statistical models for the purpose. A non-parametric statistical model namely neural networks (NN) has also been used for pattern matching and data compression<sup>5,14</sup>. Image processing technique using specially designed kernels for iris recognition are used to capture local characteristics so to produce discriminating texture features<sup>15</sup>. Several sorts of transformations also prove helpful in extracting useful features from an iris image. This feature vector is further used to form a classification approach for identifying a person based on his iris image<sup>16</sup>. In another ground breaking work the iris recognition principle is based on the failure of statistical independence test on iris phase structure encoded by multi-scale quadrature wavelets. The combinatorial complexity of this phase information across different persons generates discriminating entropy enabling most probable decision about a person's identity<sup>17</sup>.

**Content Based Image Retrieval:** Many probabilistic and statistical solutions to the CBIR problem have been proposed by numerous researchers. A localized probabilistic approach gathers feature vectors from interest points which usually require manual intervention. Statistical evaluation of these vectors or their components yields some inferences<sup>9</sup>. A technique has also been developed using Support Vector Machine (SVM) which assigned each image a label employing fuzzy logic after analyzing the contents of the image. SVM used this information during its training which enabled it to retrieve images on the basis of the content of a given image<sup>10</sup>. Certain mechanisms, based on relevance feedback collected initially from users, built interesting systems. The model based on this paradigm is trained using this feedback information. The feature identification and correlation capability is enhanced iteratively<sup>18,19</sup>. Learning based mathematical models, such as neural networks (NN) are used for analysis and matching of images. Several authors have proposed such methods, various examples can be found in text where several sample images are fed into a neural network to acquire some results<sup>20-23</sup>. Moreover, fuzzy logic is incorporated into the model to provide numerous levels of output<sup>24</sup>.

## Technical Framework

Statistical Models enable us to represent any complex finite or infinite structure in a discrete and finite representation. It formalizes the relation of random variables with one or more other random variables in a mathematical form. The defined relationship is probabilistic in nature as the defined relationship is stochastic and not deterministic. A statistical model is mathematically represented as a two tuple  $(B, P)$  where  $B$  is the observed data and  $P$  is the set of probability functions. It is based on the assumption that one of the functions within  $P$  will produce the observed data in  $B$ . In other words a statistical model  $P$  is a set of probability distribution function or probability density function. They form two major categories: i. Parametric Statistical Models, ii. Non-Parametric Statistical Models.

**Parametric Statistical Models:** A statistical model which is parametric in nature will contain probability functions described by a unique multidimensional parameter<sup>24,25</sup> formally

$$P = \{P_\theta : \theta \in \Theta\} \quad (1)$$

where  $\theta$  is a parameter and  $\Theta$  is a feasible region of parameters and also  $\Theta \subseteq \mathbf{R}^d$  where  $\mathbf{R}^d$  is a d-dimensional Euclidean space.

Some of the most commonly used parametric models are:

The normal distributions is parameterized by  $\theta = (\mu, \sigma)$ , where  $\mu$  and  $\sigma$  are location and scale parameters respectively. The distribution function is given as

$$P = \{f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} | \mu \in \mathbf{R}, \sigma > 0\} \quad (2)$$

The Poisson distribution is parameterized by a single parameter  $\lambda$  and is given as

$$P = \{p_\lambda(j) = \frac{\lambda^j}{j!} e^{-\lambda}, j = 0, 1, 2, 3, \dots | \lambda > 0\} \quad (3)$$

where  $p_\lambda$  forms the probability mass function.

The Weibull distribution has three parameters  $\theta = (\lambda, \beta, \mu)$ . The distribution function is given as

$$P = \{\varrho(x) = \frac{\beta}{\lambda} \left(\frac{x-\mu}{\lambda}\right)^{\beta-1} \exp\left(-\left(\frac{x-\mu}{\lambda}\right)^\beta\right) \mathbf{1}_{\{x>\mu\}} | \lambda > 0, \beta > 0, \mu \in \mathbf{R}\} \quad (4)$$

**Non-Parametric Statistical Models:** The term non-parametric has varied impact. A non-parametric statistical model may mean a number of things. A model is non-parametric if it is distribution free. It means that the model essentially does not require that the data is derived from a certain probability distribution. Also a non-parametric model does not essentially

imply that the structure of the model is fixed. The model may grow in size as the complexity of data increases. More formally the non-parametric model is a paradigm containing the set of probability distributions with infinite dimensional parameters<sup>26</sup>. Some typical examples of such models are: i. Kernel density estimation. It is refined technique used for density estimates. ii. A histogram is a non-parametric estimate of a probability distribution, iii. Data envelopment analysis provides efficiency coefficient. This technique does not make use of any distribution but still is used for refined analysis of data using the coefficients. iv. Neural Networks. A feed forward artificial neural network is a non-parametric statistical model used to extract non-linear relation in data.

## The Neural Networks

The neural network is a computational or mathematical model. It is based on the biological neural network used by humans to process information. It forms a network of several interconnected neurons. These neurons pass information to all the neurons in the adjacent layer. It processes data using connectionist computational approach. Typically they are used to solve non-linear problems. They are adaptive in nature they adapt themselves to patterns of interest hidden within data. They are used to find complex patterns or relationship that may exist between the input data and the required output<sup>27</sup>. Neural networks are often used as a non-linear statistical tool.

## Mixture Models

Mixture models are another interesting statistical tool. They are used when the data is of hierarchical nature. It is a probabilistic model which is used to signify the presence of groups within the population. These models may not essentially require the identification of the groups to which a certain observation belongs. A specific mixture model will have correspondence with a distribution which is assumed to represent the distribution of observations within the overall population. The most interesting fact about mixture models is that it is used to form statistical inferences about the attributes of a group within the population, using the given observation, without requiring identification of the groups within the population individually<sup>28,29</sup>. In general let there be  $m$  observations on the  $k^{th}$  subject given as  $X_k = \{X_{k,1}, X_{k,2}, \dots, X_{k,m}\}$  for  $k = 1, 2, \dots, n$ . Now in accordance with some distributions  $D$  the mean and variance vectors ( $\mu$  and  $\Sigma$  respectively) are computed using the acquired observations. Let  $x_1, x_2, \dots, x_n$  be the realized values for each  $X_i$ . Assuming that the data has the inclination towards incongruous groups, the overall population is diversified into  $p$  different pockets. Also if  $p$  is fixed then each  $X_i$  has p-component mixture density given as :

$$f_p(y_k; \rho) = \sum_{j=1}^p \lambda_j g_j(y_k; \theta_j), \quad (5)$$

where  $\lambda_j$  represents the weight of the  $j^{th}$  component within the mixture model. Also  $g_j$  is the known density function pertaining to the distribution  $D$ , it is parameterized with  $\theta_j$  where  $\theta_j \in \Theta_j \subseteq \mathbb{R}^{qj}$  and  $\Theta_j$  is the d-dimensional space for the d-dimensional vector  $\theta_j$ . Hierarchically the mixture density function is parameterized by  $\rho$  which belongs to the parameter space for all the possible unknown parameters within the mixture<sup>30,31</sup>.

## Statistical Moments

A probability density function describes the dispersement of data within the given structure. The parameters of the distribution function represent the signature of the patterns embedded in the structure. Moments are the characteristics of a probability density function that describes its property like kurtosis and skewedness. A probability density function is also formed from images using the color intensities of pixels. Image moments describe the properties of such distribution along its axes. In particular image moments form the weighted average of pixels in an image. The moments are typically chosen to depict a certain interesting property of the image. Moments extraction from segmented image only produce meaningful results. Properties of an image such as centroid, area and orientation are quantified by this process. Another dividend of image moments is that they bring together the local and global geometric details of a gray scale image. An image in real world is modeled using a Cartesian distribution function  $f(x, y)$  in its analog form. This function is used to provides moments of the order of  $(p+q)$  over the image plane  $P$  and is generalized as

$$M_{pq} = \iint_P \psi_{pq}(x, y) f(x, y) dx dy; p, q = 0, 1, 2, \dots, \infty \quad (6)$$

where  $\psi_{pq}$  is the basis function. The eq(6) yields a weighted average over the plane  $P$ . The basis function is designed such that it depicts some invariant features of the image. Furthermore the properties of the basis function are passed onto moments. A digital image is of discrete nature thus the image plane is divided into pixels each having a discrete intensity level. The eq. (7) is adopted for the digital image as

$$M_{pq} = \sum_x \sum_y \psi_{pq}(x, y) I(x, y); p, q = 0, 1, 2, \dots, \infty$$

where  $I(x, y)$  is the intensity of a pixel in the digital image at the  $x^{th}$  and  $y^{th}$  column<sup>32</sup>.

Various moments have been derived by different researcher to help extraction of patterns from data. Some of these moments are: i. Hu Invariant moments (More Commonly known as Image Moments), ii. Hahn moments, iii. Complex Zenrike moments.

## Statistical Learning

Statistical learning pertains to development of statistical models which are used as a basis of intelligent algorithms. These algorithms along with the statistical/mathematical model have the capability to imitate behaviors in response to some input data. The input data acquired contains obscured patterns for which the model acquires cognitive capabilities over a process of learning. The focus of statistical learning lies in devising automated methods for recognizing hidden patterns within data and performs decisions based on the recognized patterns. It is a probabilistic technique it does not use the entire input space for training rather it only uses few observed samples for the training process<sup>5</sup>. It requires that the learner must generalize the possible output cases with respect to the input. Various such learning algorithms have been developed by researchers. These algorithms are categorized into two categories: i. Unsupervised Learning, ii. Supervised Learning

**Unsupervised Learning:** Unsupervised learning techniques are used to extract information from unlabeled data. The main difference between supervised and unsupervised learning is that in unsupervised learning the data is unlabeled and no errors can be assigned to each input which could help to converge towards an optimal solution. Unsupervised learning is quite similar to the statistical techniques of density estimation. Generally unsupervised learning techniques aim to compress the information within data into few parameters or coefficients<sup>31</sup>. These parameters or coefficients are further used to depict or describes the peculiar features of data. Some of the typically used unsupervised learning techniques are: i. k-means Clustering, ii. Fuzzy C-means Clustering, iii. Principal Component Analysis.

**k-means Clustering:** k-means is a clustering techniques used to identify clusters within a large given data. It aims to assign each observation to cluster based on the mean. Each observation belongs to a cluster with the nearest mean out of all the possible clusters. The k-means algorithms yields partitions of data into Voronoi cells<sup>1,3</sup>.

Using a set of multidimensional observations  $(x_1, x_2, \dots, x_n)$ , the k-means algorithm partitions the observations into  $K$  sets such that  $(K \leq n)$  generating the set  $S = S_1, S_2, \dots, S_k$  so as to minimize the following objective function

$$\arg_s \min \sum_{i=1}^K \sum_{x_j \in S_i} |x_j - \mu_i|^2 \quad (7)$$

where  $\mu_i$  is mean of all the observations in  $S_i$ .

## Fuzzy C-means (FCM)

In fuzzy clustering every observation is believed to have some degree of belonging to every cluster. Data items near the center

of a cluster will have a greater degree of belonging to the cluster. It will also have some (comparatively lesser) degree of belonging to all the rest of the cluster in the overall model. Formally the FCM algorithm aims to partition a finite collection of data  $X = x_1, x_2, \dots, x_n$  into  $c$  fuzzy clusters in terms of the given criteria. The objective of the algorithm is to form a vector  $C$  containing the centers of all the clusters and yield a matrix  $U$  illustrating the degree of belonging of each data item to the each other cluster<sup>33</sup>. The FCM algorithm minimizes the objective function

$$u_k(x) = \frac{1}{\sum_j \left( \frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{2/(m-1)}} \quad (9)$$

**Principal Component Analysis (PCA):** PCA is a mathematical transformation that converts seemingly correlated variables into uncorrelated orthogonal coefficients. This yields uncorrelated variables called principal component. This transformation is performed in such a way that the initial principal component has the highest variance, the next principal component has highest variance and is uncorrelated (orthogonal) to the subsequent components<sup>34</sup>.

**Supervised Learning:** In supervised learning a set of labeled data is used to train the model. The training data form a large set of training examples. Each example is considered to be a 2-tuple. The examples consist of the input data and the supervisory signal. The input data is generally a vector and the supervisory signal is in fact the desired output for the model. The algorithm generated an inferred function. As the model gets trained it develops the capability to predict the desired output for any arbitrary input. Formally if a model is provided the inputs  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  then it develops a function  $h: X \rightarrow Y$  where  $X$  and  $Y$  are the input and output space respectively and the function  $h$  belongs to a so called hypothesis space  $H$ . The function  $h$  is often defined using another functions  $f$  such that  $f: X \times Y \rightarrow R$  and the function  $g$  for any input yields the maximum value of  $y$  described by  $h(x) = \arg \max_y f(x, y)$ . Let the function  $f$  belong to the space  $F$ . If for example the learning model is of probabilistic nature then the function  $f$  is given as  $f(x, y) = P(x, y)$  or the function  $h$  is given as  $h(x) = P(y|x)$ <sup>1</sup>.

**Back Propagation of Errors in Neural Networks:** Researchers have developed various neural networks. Each network may vary in terms of its size, it layers and its learning algorithm. A multi-layer network consists of layers of neurons. Each neuron in a layer is connected to all the neurons in the preceding layer. A back propagation network is one in which

the network converges to the expected output in a number of iterations. In each iteration the weights of the neurons are adjusted to tune the network. The change in the weight of the neuron is dependent upon the difference in actual and the expected output. A learning algorithm which determines the change in weight at each iteration. Delta rule can be used for updating the weights at each iteration. The objective of the delta rule is to minimize error in output through gradient descent<sup>35,36</sup>. The delta (i.e. change in weight) at each neuron is given as

$$\Delta_r w_{jk} = -\gamma \frac{\partial \mathcal{E}^t}{\partial w_{jk}} \quad (10)$$

where  $\mathcal{E}^t$  can be given as

$$\mathcal{E}^t = \frac{1}{2} \sum_{r=1}^N (d_{(t)r} - y_{(t)r})^2 \quad (11)$$

here note that  $\mathcal{E}^t$  is the sum of quadratic error while  $d_{(t)r}$  is the desired output and  $y_{(t)r}$  is the actual output for a certain neuron  $t$  which is fed an input  $r$ .

## Conclusion

Images are digitally described by means of pixel whose color intensities are quantified. Either the occurrence of this data or the arrangement of this data forms hidden patterns. These obscured patterns are detected by various probabilistic or statistical techniques as discussed earlier. Also we have identified types of pattern recognition problems. A generic pattern recognition algorithm will search for general patterns within an image. It will be able to identify if a certain object exist within an image or whether the two images are similar based on their contents example of which is the Content Based Image Retrieval (CBIR) systems. A specific pattern recognition algorithm will look for specific pattern within an already identified object; examples of such paradigms are biometrics applications. An iris recognition application is provided always with an image of an iris, the application in turn identifies the person to whom that iris belongs. The described models have been widely used by various researchers in order to provide solution to pattern recognition problems. The choice of a model for any such problem which may prove to be robust largely depends upon the nature of the problem. A careful and appropriate choice of model requires a good insight into the nature of the problem and an understanding of useful data that should be extracted from the image.

## References

1. Hasti T., Tibshirani R. and Friedman J., *The Elements of Statistical Learning*, Springer-Verlag, (2000)
2. Mohamadian Zahra, Image Duplication Forgery Detection using Two Robust Features, *Res. J. Recent Sci.*, 1(12), 1-6 (2012)
3. MacKay D.J.C., *Information Theory, Inference and Learning*

- Algorithms*, Cambridge University Press, (2003)
- 4. Fam D.F., Koh S.P., Tiong S.K. and Chong K.H., *Res.J.Recent Sci.*, **1(9)**, 74-78 (2012)
  - 5. Khan Y.D., Ahmed F. and Waqas M., Iris Recognition Using Back Propagation, *World Applied Science Journal*, **16(5)**, (2012)
  - 6. Tang H.L., Hanka R. and Ip H.H.S., Histological Image Retrieval Based on Semantic Content Analysis, *IEEE Trans. on Info. Tech. in Biomed.*, **7(1)**, 26-36 (2003)
  - 7. Sudhamani M.V. and Venugopal C.R., Multidimensional Indexing Structures for Content-Based Image Retrieval: A Survey, *Int. J. of Inno. Computing, Info. and Control*, **4(4)**, 867-881 (2008)
  - 8. Jiang W., Er G., Dai Q. and Gu J., Similarity Based Online Feature Selection in Content Based Image Retrieval, *IEEE Trans. on Image Proc.*, **15(3)**, 702-712 (2006)
  - 9. Rahmani R., Goldman S.A., Zhang H., Choletti S.R. and Fritts J.E., Localized Content-Based Image Retrieval, *IEEE Trans. on Pattern Analysis and Machine Intell.*, **30(11)**, 1902-1912 (2008)
  - 10. Wu K. and Yap K.H., Fuzzy S.V.M. for Content Based Image Retrieval: a psuedo-label support vector machine framework, *IEEE Comput. Intelli. Mag.*, **1(2)**, 10-16 (2006)
  - 11. Chun Y.D., Kim N.C. and Jang I.H., Content Based Image Retrieval Using Multiresolution Color and Texture Features, *IEEE Trans. on Multimedia*, **10(6)**, 1073-1084 (2008)
  - 12. Marakakis A., Galatsanos N., Likas A. and Stafylopatis A., Probabilistic Relevance Feedback Approach for Content-Based Image Retrieval Based on Gaussian Mixture Models, *IET Image Proc.*, **3(1)**, 10-45(2009).
  - 13. Monro D.M., Rakshit S. and Zhang D., DCT-Based Iris Recognition, *IEEE Trans.on Pat.Analysis Mach.Intell.*, **29(4)**, (2007)
  - 14. Abiyev R.H. and Altunkaya K., Personal Iris Recognition Using Neural Network, *Int. J.of Security and its App.*, **2(2)**, (2008)
  - 15. Ma L., Tan T., Wang Y. and Zhang D., Personal Identification Based on Iris Texture Analysis, *IEEE Trans.on Pat.Analysis Mach.Intell.*, **25(12)** (2003)
  - 16. Lim S., Lee K., Byeon O. and Kim T., Efficient Iris Recognition through Improvement of Feature Vector Classifier, *ETRI Journal*, **23(2)**, (2001)
  - 17. Daugman J., How Iris Recognition Work, *IEEE Transactions on Circuits and Systems for Video Technology*, **14**, (2002)
  - 18. Grigorova A., De Natalie F.. G.B., Dagli C. and Huang T.S., Content Based Image Retrieval by Feature Adaptation and Relevance Feedback, *IEEE Trans. on Multimedia*, **9(6)**, 1183-1192 (2007)
  - 19. Qiu G. and Lam K.M., Frequency Layered Color Indexing for Content Based Image Retrieval, *IEEE Trans. on Image Process*, **12(1)**, 102-103 (2003)
  - 20. Yap P.T. and Paramesran R., Content Based Image Retrieval using Legendre Chromaticity Distribution moments, *IEE Proceed.- Vision, Image and Signal Process.*, **153(1)**, 17-24 (2006)
  - 21. Patheja P.S.,Wao Akhilesh A.and Maurya Jay Prakash, An Enhanced Approach for Content Based Image Retrieval, *Res. J. Recent Sci.*, **1(ISC-2011)**, 415-418 (2012)
  - 22. Venkatesh Y.V., Raja S.K. and Kumar A.J., On the Application of a Modified Self-Organizing Neural Network to Estimate Stereo Disparity, *IEEE Trans. on Image Proc.*, **16(11)**, 2822-2829 (2007)
  - 23. Sao A.K. and Yegnanarayana B., Face Verification Using Template Matching, *IEEE Trans.on Info. Forensics and Security*, **2(3)**, 636-641 (2007)
  - 24. Lu H.C., and Tsai C.H., Image Recognition Study via the Neural Fuzzy System, *Proc. Int. Conf. Intelligent Engineering Systems INES '06*, 222-226 (2006)
  - 25. Mirza Nawazish and Saeed, Mawal Sara, *Res. J. Recent Sci.*, **1(11)**, 41-46 (2012)
  - 26. Wasserman L., *All of Nonparametric Statistics*, Springer, (2007)
  - 27. Plaut D.C., Nowlan S.J. and Hinton G.E., *Experiments on learning by back propagation*, Carnegie-Mellon University, (1986)
  - 28. Hogg R.V., McKean J.W. and Craig A.T., *Introduction to Mathematical Statistics*, Pearson Education, (2005)
  - 29. Davidson A.C., *Statistical Models*, Cambridge University Press, (2003)
  - 30. McLachlan G.J. and Peel D., *Finite Mixture Models*, John Wiley and Sons (2000)
  - 31. Lindsay B.G., *Mixture Models: Theory, Geometry and Applications*, IMS, (1995)
  - 32. Hu M., Visual pattern recognition by moment invariants, *IRE Trans. on Info. Theory*, **8(2)**, (1962)
  - 33. Bezdek J.C., Krisnapuram R. and Pal N.R., *Fuzzy models and algorithms for pattern recognition and image processing*, Springer, (2005)
  - 34. Jolliffe I.T., *Principal Component Analysis*, Series: Springer Series in Statistics, Springer (2002)
  - 35. Duda R. O., Hart P. E., Stork D. G., *Unsupervised Learning and Clustering*, Wiley, (2001)
  - 36. Rao Sathish U. and Rodrigues L.L. Raj, *Res.J.Recent Sci.*, **1(5)**, 75-82 (2012)