Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Comparison and Classification of Pictures by SOM and Metaclassifiers

[1]**M.Priyadharshini**, II M.Tech (cse), School of computing,Bharath University, India.
[2] **Dr.A.Kumaravel,Dean**,School of computing,Bharath University, India.

***Abstract:*** ***This paper presents a technique in classifying the pictures into variety of categories or clusters desired by suggests that of Self Organizing Map (SOM) Artificial Neural Network methodology. variety of 250 color pictures to be classified as antecedently done SOMe process, such as RGB to grayscale color conversion, color bar graph, and so classifying by the SOM Feature vector choice in this paper can use 2 strategies, specifically by PCA (Principal part Analysis) and LSA (Latent linguistics Analysis) during which every of those strategies would have taken the characteristic vector of 50, 100, and150 from 256 initial feature vector into the method of color bar graph. Then the choice are processed into the SOM network to be classified into 5 categories victimization a learning rate of 0.5 and calculated accuracy that is classified by the meta classifier. Classification of a number of the check results showed that the best share of accuracy obtained once victimization PCA and the choice of a hundred feature vector that's capable half of 1 mile, compared to once victimization LSA choice that 74% therefore it will be terminated that the strategy fits the PCA feature choice strategies area unit applied in conjunction with SOM And has an accuracy rate higher than the LSA feature choice strategies.***

***Keywords: Color bar graph, Feature choice, LSA, PCA, SOM, Meta classifier***

## I. Introduction:

Image classification technology is presently being wide developed. The SOM (Self Organizing Map) Neural Network or normally referred to as a Kohonen Neural Network system is one of the unsupervised learning model which will use the units by the similarity of a selected pattern to the realm within the same class. Compare the images by using the algorithms Pca, Lsa, that giving the result of the images which is processed in the retrieval of images, then we classify the images by using Som and the final result will be classified by Meta classifier then the total images will be classified by their categories. The image meant in this paper is a assortment of natural pictures (total 250 images) that contain standard objects like mountains, rivers, cars, animals, and flowers which is able to then be sorted into completely different categories by suggests that of SOM artificial neural network strategies. In the presence of this classification technique we first describe the feature vector classification then how the images are converted to histogram. Next the algorithm PCA, and LSA which is performing the image retrieval then we classify the images by SOM, and Meta Classifier to get the final result.

**1.FEATURE VECTOR ILLUSTRATION:**

The image can be sorted into clusters antecedently needed 2 previous process, the color conversion and bar graph to provide a feature vector of a collection of image.

1.1COLOR CONVERSION:

Color conversion is referred to in this paper is the conversion of RGB color pictures (24 bits) into grayscale (8 bits), thus that the color model can be less complicated with every component gray level between zero to 255. The conversion formula is:

$$RGB = \frac{R+G+B}{3}$$

Where R is that the red price of component, G is that the inexperienced price and B is that the blue price of component pictures.

**1.2 HISTOGRAM:**

Bar graph is called histogram either histogram or color hiatogram is one of the techniques of applied math options that will be used to take the feature vector of an information set, images, video or text. The generated feature vector of color bar graph are a chance price of h i , n i is that the range of components of i color intensity that seems within the m image divided by the entire n image pixel.

$$h_i = \frac{n_i}{n} \qquad i = 0 \ldots 255$$

In this method feature vector matrix obtained from the entire image can be sized (i x m), that can then be carried through the characteristic dimension reduction strategies of feature choice on consequent method.

## 2 FEATURE VECTOR SELECTION:

Feature vector choice technique during this paper focuses on 2 strategies specifically PCA (Principal part Analysis) and LSA (Latent linguistics Analysis), that is a preprocessing stage of the classifying method. the aim of this feature vector choice is to cut back the K dimensional matrix obtained from the characteristic options of the previous vector to < K dimensions while not reducing the necessary info on that.

## 3 PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA methodology could be a world feature choice algorithmic program that initial planned by Hotteling (1933) as SOM way to scale back the dimension of a area that is drawn in statistics of variables (x i , i = 1,2…. n) that reciprocally related to with one another. throughout its development, PCA algorithmic program (also referred to as the Hotteling transformation) is used to cut back noise and extract options or essential characteristics of knowledge before the classification method. Election of world feature choice technique in this case as a result of the pictures are classified to own comparatively low frequency (low level), that the PCA methodology is sort of appropriate to be applied.

The steps within the PCA algorithmic program specifically :

a) produce a matrix [X one , X 2 , .... X m ] that representing N2 x m knowledge matrix. X i is that the image of size N x N, wherever N2 is the total pixels of the image dimensions and m is that the range of pictures to be classified.

$$\mathrm{Y} = \frac{1}{m}\sum_{i=1}^{m} X_i$$

b) Use the following equation to calculate the average price of all pictures

$$\overline{X_i} = X_i - \mathrm{Y}$$

c) Calculated the distinction matrix :

d) Use the distinction matrix acquireed antecedently to generated the variance matrix to obtain the matrix :

$$\Sigma = \sum_{i=1}^{N} \overline{x_i}\,\overline{x_i}^{T}$$

$$\Sigma\phi_i = \lambda\phi_i$$

e) Use the matrix to judge the eigenvector

Where is orthogonal eigenvector matrix, ? is the eigenvalue square matrix with diagonal components sorted that aims to cut back the eigenvector matrix kind victimization the feature area The order of the eigenvectors with the largest eigenvalue represents the knowledge nearer or similar to the first knowledge

$$\Phi = [\phi_1 \mid \phi_2 \mid \ldots\ldots \mid \phi_n]$$

Where, $1 \le n \le N2$

f) if ϕ could be a feature vector of the sample image X, then :

$$y_n = \Phi^T \overline{X_i}$$

With feature vector y is that the n-dimensional.

## 4 LATENT SEMANTIC ANALYSIS (LSA)

LSA could be a method that was originally used in the sector of AI branch of tongue process (NLP) to analyze the knowledge as plain text or document. LSA or additionally referred to as LSI (Latent linguistics Indexing), in its application will additionally be used to method image knowledge to create a replacement matrix decomposition of the

initial matrix into 3 matrices that related to. This technique is then referred to as by the Singular price Decomposition (SVD), that is a component of the LSA

$$X = USV^T$$

If X could be a matrix of feature vector area unit sized i x m, with total of gray level i and also the total image m, then U could be a matrix of othonormal i x m, S is the m x m square matrix, and V area unit orthonormal m x m. The matrix U obtained from the search eigenvectors of the matrix operation yields X.V.S-1 . whereas the V matrix obtained from the search eigenvectors of the matrix operation yields disturbance .X.

The S square matrix can contain the eigenvalues of the matrix of Green Mountain State orthonormal eigenvectors with a sequence beginning from the largest eigenvalue to the smallest, from left to right. when the third matrix area unit obtained, then the reduction in range of r-dimensional matrix as a vector that can manufacture the U k , S k and VkT matrix. Then to get the Q

feature vector matrix with K or n-dimensional, use the formula :

$$Q = Q^T.U_k.S_k^{-1}$$

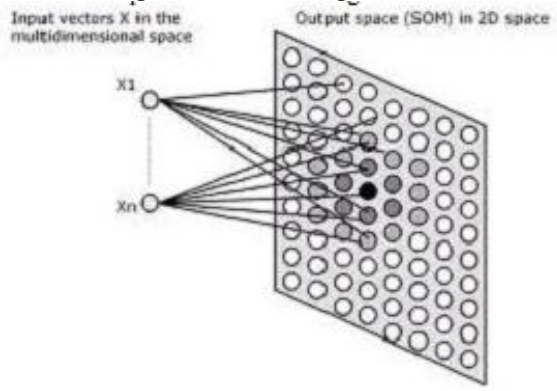
wherever $Sk^{-1}$ is a square matrix inverse, and $Q^T$ is a new feature vector transpose matrix.

## 5. SELF ORGANIZING MAP (SOM)

Artificial Neural Network (ANN) is outlined as info process system that has characteristics resembling human neural tissue. The existence of ANN provides a new technology to facilitate solve issues that need thinking of specialists and laptop based mostly routine. A few of ANN application was for classification system (clustering), association, pattern recognition, foretelling and analysis .

And in this paper, the ANN methodology that applied is SOMmethodology, that to be used to classify the image into a collection of 5 completely different categories. In the Self Organizing Map (SOM) methodology, the applied learning is AN unsupervised learning wherever the network will not utilize the category membership of sample coaching, however use the data in a very cluster of neurons to switch the native parameter.

The SOM system is adaptively classify samples (X image data) into categories determined by choosing the winning neurons area unit competitive and also the weights area unit changed.




SOM Artificial Neural Network.

The algorithmic program on the SOM neural network as follows :

a. If the feature vector matrix of size k x m (k is the amount of feature vector dimensions, and m is the range of data), the low-level formatting :

- The range of the specified j category or cluster
- the amount of part i of the feature matrix (K is that the row of the matrix)
- The range of vector X m,i = quantity of data(matrix column)
- The initial weights W terrorist group were willy-nilly with interval zero to one
- The initial learning rate a(0)
- The range of iteration (e epoch)

b. Execute the initial iteration till the total iteration (epoch)

c. Calculate the vector image to start out from one to m :

$$D(j) = \sum_i (W_{ji} - X_{m,i})^2$$

For all of j

- Then confirm the minimum price of D(j)

$$W_{ji}^{new} = W_{ji} + \alpha( X_{mi} - W_{ji} )$$

- Make changes to the j weight with the minimum of D(j)

d. Modify the educational rate for consequent iteration which t begin from the primary iteration

$$\alpha(t+1) = 0{,}5\ \alpha\ (t)$$

e. check the termination condition Iteration is stopped if the distinction between W terrorist group and W terrorist group the previous iteration solely a very little or a modification in weights simply terribly tiny changes, then the iteration has reached convergence in order that it is stopped.

f. Use a weight of W terrorist group that has been convergence to grouping feature vector for every image, by scheming the gap vector with best weights.

g. Divide the image (X m ) into classes :

If D(1)<D(2)<D(3)<D(4)<D(5), then the images included in class 1
If D(2)<D(1)<D(3)<D(4)<D(5), then the images included in class 2
If D(3)<D(2)<D(1)<D(4)<D(5), then the images included in class 3

If D(4)<D(3)<D(2)<D(1)<D(5), then the images included in class 4
If D(5)<D(4)<D(3)<D(2)<D(1), then the images included in class

## 6. EXPRIMENTAL SETUP AND RESULT:

In this paper, the implementation stage of image classification system with SOM and feature selection method can be viewed via the following flowchart :

Figure 4 shows that some processing needs to be done previously until the images can be classified into five classes, which in this study consists of five classes namely flower class, animal class, car class, river class, and mountain classes. The image used in this study were as many as 250 images by dividing each image of the 50 pieces in each class.

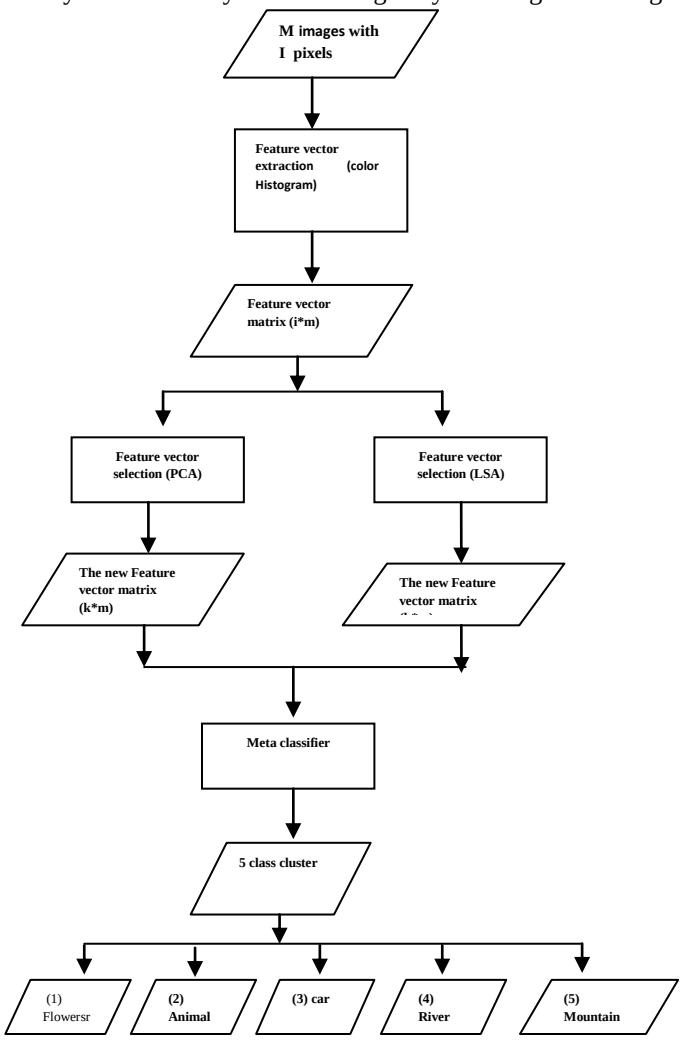



**Fig. 4 Flowchart of The Implementation**

**Fig. 5 Example of 16 Images to be Classified**

In the color histogram, feature vector of 250 images is retrieve d and will generate feature vector matrix i x m, which is 256 x 250. 256 is the number of grey level (0 – 255) from the grayscale color conversion process.

The matrix will then be reduced to the dimensions as r columns or rows through PCA and LSA feature selection algorithms. From the result of algorithm, will be gained a new feature matrix to the size k x m (k is the dimension after reduction, and m is the total images).

So if the initial dimensions of the image matrix is 256x250, then after feature selection (eg, r reduction as much as 156 vectors), the size of the matrix will be changed to 100x250.

Then from the matrix k x m will be classified into five classes by SOM neural network algorithm, with the network parameters that are used :

- The number of class (j) = 5
- The number of vector component (i) = 50 ; 100 ; 150
- The number of X vectors = 250
- The initial weights (W ji ) = 0 to 1 (random)
- The initial learning rate (a) = 0.5
- Total iteration or epoch (e) = 500

The experimental results will be seen from the large percentage of SOM accuracy in classifying the images corresponding to the class by using an application program, where the formula of accuracy is following figure shows,

$$\% = \frac{\textit{The image classification results}}{\textit{The number of image in the class should be}} \textit{ x 100\%} \quad (14)$$

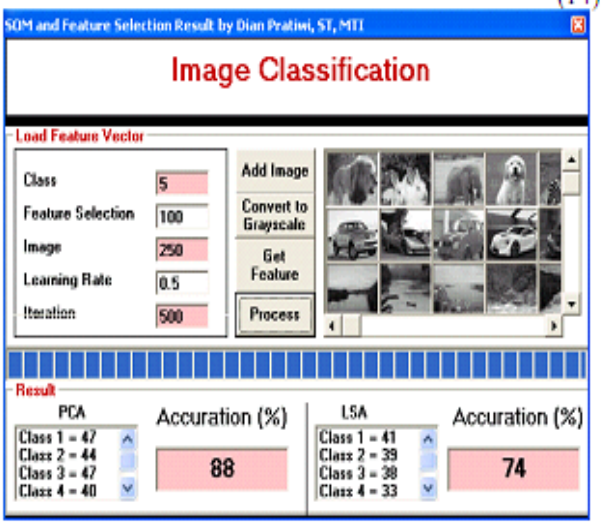



**Fig. 6 The Display of Image**

Classification Program with SOM Method and Feature Selection

From the total tests performed in this study, the results of image classification can be demonstrated by the following tables :

| Classification Result | | | | | The class should be |
|---|---|---|---|---|---|
| Flower | Animal | Car | River | Mountain | |
| 47 | 1 | 2 | 0 | 0 | Flower |
| 3 | 44 | 1 | 1 | 1 | Animal |
| 0 | 2 | 47 | 0 | 1 | Car |
| 1 | 3 | 1 | 40 | 5 | River |
| 1 | 1 | 0 | 4 | 44 | Mountain |

**Table 1: Classification Results of PCA and SOM with The 100 Feature Vectors**

| | PCA (Feature Vector) | | | LSA (Feature Vector) | | |
|---|---|---|---|---|---|---|
| | 50 | 100 | 150 | 50 | 100 | 150 |
| Flower | 42 | 47 | 43 | 40 | 41 | 46 |
| Animal | 41 | 44 | 41 | 33 | 39 | 35 |
| Car | 39 | 47 | 42 | 34 | 38 | 42 |
| River | 29 | 40 | 35 | 22 | 33 | 25 |
| Mountain | 27 | 44 | 39 | 24 | 35 | 30 |
| Σ | 178 | 222 | 200 | 153 | 186 | 178 |
| % Accuration | 71 | 88 | 80 | 61 | 74 | 71 |

**Table 2: Classification Results of SOM with The LSA and PCA Method for Feature Vectors as many as 50, 100, and 150 Vectors**

In general, it can be seen from Table 2 that the SOM classification uses of both PCA and LSA feature selection showed a fairly good percentage of accuracy, with an average success percentage of 68.6% - 79.6%. But when compared to results from the use of both feature selection, the highest level of accuracy obtained when using the PCA by 88% (222 of 250 classified image is the right image) with the number of feature vector is 100 vector (shown in Table 1).While the LSA, the highest accuracy results obtained only by 74% (186 of 250 classified image is the right image) with the ame number of feature vectors (shown in Figure 6). Thus, it can be concluded that the SOM method is more suitable to be applied with PCA feature selection to classify images into classes in accordance with good results compared with LSA feature selection.

## 7. CONCLUSION

The conclusion to be drawn from the writing of this paper is:

1. The selection method of PCA and LSA feature selection precise enough to be implemented in the image classification system, because it can reduce the dimensions of the image matrix while producing a high level of accuracy which is between 61% - 88%
2. The ability of SOM classification techniques can be further increased when the number of image feature vector as many as 100 vector with the type of feature selection is PCA. This is because the number of feature vectors is a number of feature vector with the best accuracy, that is 88%, of which there remain some important information in these vectors although the dimension of characteristic matrix has been reduced.
3. With the image classification techniques in the database will be able to facilitate and speed up image retrieval system, because the image can look directly into the appropriate classes without having to search one by one from each class.

**REFERENCES :**

[1] The Use of Self Organizing Map Method and Feature Selection in Image Database Classification System 1 Dian Pratiw Department of Information Engineering, Trisakti University Jakarta, 15000, Indonesia

[ 2] Siong, A.W, Pengenalan Citra Objek Sederhana dengan Menggunakan Metode Jaringan Syaraf Tiruan SOM Prosiding Seminar Nasional I Kecerdasan Komputasional, Universitas Indonesia. 1999.

[3] H. C. Wu, S. Huang, "User Behavior Analysis in Masquerade Detection Using Principal Component Analysis". Eighth International Conference on Intelligent Systems Design and Application (ISDA), 201-206, DOI : 10.1109/ISDA.2008.243.

[4] A. Hermawan, Jaringan Syaraf Tiruan Teori dan Aplikasi. Yogyakarta : Penerbit ANDI, 2006.

[5] J.J. Siang, Jaringan Syaraf Tiruan dan Pemrogramannya menggunakan Matlab, Yogyakarta : Penerbit Andi, 2005 and [5] Jolliffe, I.T., Principal Component Analysis, 2 ed., New York : Springer-Verlag, 2002

[ 6] Journal of Electronic Imaging 13(4), 780–788 (October 2004). "Medical image classi?cation using genetic algorithm based fuzzy-logic approach Du-Yih Tsai Yongbum Lee Masaru Sekiya Masaki Ohkubo Niigata University Department of Radiological Technology School of Health Sciences, Faculty of Medicine 2-746 Asahimachi-dori Niigata City Niigata 951-8518, Japan E-mail: tsai@clg.niigata-u.ac.jp

[7] C. J. Lin, C.H. Chu, C.Y. Lee, Y.T. Huang, "2D/3D Face Recognition Using Neural Networks Based on Hybrid Taguchi Particle Swarm Optimization", Eighth International Conference on Intelligent Systems Design and Application (ISDA), 307-312, DOI : 10.1109/ISDA.2008.286.

[8] T.K. Laundauer, P.W. Foltz, D. Laham, Introduction to Latent Semantic Analysis, Discourse Process, 25, 259-284, 1998.

[9] M.E. Wall, A. Rechtsteiner, L.M. Rocha, Singular Value Decomposition and Principal Component Analysis, Portland tate University, 2003.

[10] N. Navaroli, D. Turner, A.I. Conception, R.S. Lynch, "Performance Comparison of ADRS and PCA as a Preprocessor to ANN for Data Mining", Eighth International Conference on Intelligent Systems Design and Application (ISDA), 47-52, DOI : 10./ISDA.2008.133