Indian J.Sci.Res. 17(2): 190-196, 2018
ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

A STUDY OF FEATURE SELECTION IN COMPRESSED MEDICAL IMAGES USING NEURAL NETWORKS

[1]Shaik.Jakeer Hussain, [2] Dr. R. Kiran Kumar
[1,2] Department of Computer Science and Engineering, Krishna University, Machilipatnam, A.P.

Abstract - Content based image retrieval (CBIR) or content based visual information retrieval (CBVIR) has been one of research areas in the field of computer science and engineering from the two decades. In medical field, X-Rays, Medical Resonance Image (MRI), ECG and CT Scan are produce digital images for diagnosis and treatment to the patients. These digital images required more storage space and high bandwidth. Compression is a process to reduce the total number of bits of an image. The quality of the compressed images evaluated some factors like Compression Ratio (CR), Root Mean Square Error(RMSE) and Normalized Mean Square Error (NMSE) and Peak Signal-To-Noise Ratio(PSNR). Image compression techniques applied on images, to minimise the storage and bandwidth. CBIR technique used to extract images which are visually like a specified target image. CBIR technique based on features like shape, texture and colour. Each feature is represented by one or more feature descriptors.Texture is another important property of images, powerful descriptor that helps in the retrieval process. Gabor filter used for texture analysis. Gabor filter provides the best pattern retrieval accuracy. Edge detection reduces the amount of data and filters out useless information. Sobel edge detector extract the edge features from the compressed medical images. The feature selection is important to speed up learning and to improve concept quality.Fisher score is one of the most widely used supervised feature selection methods. Shape and texture features are extracted and best features are selected from fisher technique.

keywords— Compressiontechniques, Classification Techniques, Haar wavelet transform, Sobel edge detection, Gabor Filter, Neural Networks, Image Compression, Fisher Score

I. Introduction

Image Processing has two objectives. Create more images to suitable for people and computers can recognise the images and understand as per user requirement CBIR has three features like color, texture, shape to classify the image and retrieval. Feature extraction approach in Medical Resonance Image (MRI) is important to perform diagnostic image analysis.Digital images occupied more space to store and high bandwidth to transmit them. Image compression is a technique to shrink the images and less bandwidth.Haar functions are simplest wavelet function, extracts the image features. It is matrix based and wavelet analysis can be used many areas in image processing. Texture analysis can make significant contribution is the area of content based image retrieval in large image databases. Gabor texture feature is provided the best overall retrieval accuracy. Feature selection searches and extracts a subset of relevant features, to effective of classification accuracy is optimised. Fisher Score is mixed integer programming, can be reformulated as a quadratically constrained linear programming (QCLP). It is used for feature selection.

II. Literature Overview

Yang et al., [20] presented a boosting framework for distance metric learning that aims to preserve both visual and semantic similarities. Syam et al., [21] proposed a CBIR that using Medical images for retrieval and the feature extraction is used along with color, shape and texture feature extraction to extract the query image from the database medical images. Kumar &Kumaraswamy [22] proposed to implement a novel feature selection mechanism using Discrete Sine Transforms (DST) with Information Gain for feature reduction. Inverse fisher criterion [24]was augmented by adding a constraint in PCA procedure so that singularity phenomenon will not occur. Nithya&Menaka [23] proposed Rough set theory (RST) based feature selection technique for removing irrelevant features and producing high accuracy for post processing data.Peter et al., [24] developed a feature selection approach by projecting the image into sub-space based on Fisher's Linear Discriminant.

III. Image Compression

Image compression applied to digital images to reduce the overall size of image to cost for storage and transmission. Image compression may be lossy and losslesscompression. Lossy compression achieves a high compression ratio 50:1 or higher, shows some acceptable degradation. In this compression, cannot completely recover the original data. Lossless compression can completely recover the original data, reduces the compression ratio is 2:1. In medical applications, lossless compression has most important requirement because, facilities accurate diagnosis due to no degradation on the original data. The measurements of lossless compression are compression ratio and bit ratio.

[1]**Corresponding Author**

Compression Ratio (CR) is the ratio of the number of nonzero elements in original matrix to the number of nonzero elements in updated transformed matrix [1].

The measurements of lossy compressions are root mean square error(RMSE), normalized mean square error (NMSE) and peak signal-to-noise ratio(PSNR).

$$RMSE = \sqrt{\frac{1}{N*M}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}[g(i,j)-g'(i,j)]^2}$$

$$NMSE = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}[g(i,j)-g'(i,j)]^2}{\left[\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}[g(i,j)]^2\right]}$$

The PSNR for gray scale (8 bits per pixel) is

$$PSNR = 20 \times log_{10}\left(\frac{255}{RMSE}\right)$$

Where$g(i,j)$ represents the original image and $g'(i,j)$ represents the reconstructed image for decompression.

The PSNR is not satisfy the quantified of picture quality. so quality measure is MOS and PQS has been applied. The measure uses the five grades like 5 for excellent, 4 for Good, 3 for Slightly Annoying, 2 for Annoying and 1 for Very Annoying.

The MOS can be calculated by

$$MOS = \sum_{i=1}^{5} i.p(i)$$

Where

$i$is the grate and $p(i)$ is the grade probability.

PQS can be calculated by

$$PQS = b_0 + \sum_{i=1}^{5} b_i Z_i$$

IV. Haar Wavelet Transform

Haar functions are used since 1910. They were introduced by Hungarian mathematician Alfred Haar [2].The wavelet methods are strongly connected with classical basis of the Haar functions. The HWT scaling and dilation of a basic wavelet can generate the basis Haar functions.

Definition 1: Let $\Psi: R \to R$, the Haar Wavelet function defined by the below formula

$$\Psi(t) = \begin{cases} 1, & for\ t \in [0,1/2) \\ -1, & for\ t \in [1/2,1) \\ 0, & otherwise \end{cases}$$

Definition 2: Let $\Phi: R \to R$, the Haar scaling function is

$$\Phi(t) = \begin{cases} -1, & for\ t \in [0,1) \\ 0, & for\ \in [0,1) \end{cases}$$

The Haar has important property$V^j = V^{j-1} \oplus W^{j-1}$where $\oplus$ is orthogonality of V^j and W^j [3].

Where$W^j = span\left\{\Phi_i^j\right\}_{i=0,1,\dots 2^{j-1}}$ and $W^j = span\left\{\Psi_i^j\right\}_{i=0,1,\dots 2^{j-1}}$

A collection of linear independent functions $W^j = span\left\{\Psi_i^j\right\}_{i=0,1,\dots 2^{j-1}}$spanning $W^j$is called as wavelets. The basic functions from the space V^j are called as scaling functions.

A signal has two components for decompose, average (approximation) and detail (fluctuation)

A signal has 2^n sample values, the firstaverage (approximation) and of sub-signal is $a^1 = (a_1, a_2, a_{3,} \dots, a_{N/2})$ and $d^1 = (d_1, d_2, d_{3,} \dots, d_{N/2})$with length N.

Where N is given by $N = \frac{x_{2n-1}+x_{2n}}{\sqrt{2}}, n = 1, 2, 3 \dots . \frac{N}{2}$

The resultant matrix M has four pieces, with each piece of 2×2 (No. of rows $\times$ No. of columns). Each piece is termed as A, H, V and D.

A is approximation area, H is horizontal Area, V is vertical Area and D is Diagonal Area [3].

$$M = \begin{bmatrix} x_{11} & x_{12} & \vdots & x_{13} & x_{14} \\ x_{21} & x_{22} & \vdots & x_{23} & x_{24} \\ \dots & \dots & \dots & \dots & \dots \\ x_{31} & x_{32} & \vdots & x_{33} & x_{34} \\ x_{41} & x_{42} & \vdots & x_{43} & x_{44} \end{bmatrix}$$

From the above matrix,

$$A = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

$$H = \begin{bmatrix} x_{13} & x_{14} \\ x_{23} & x_{24} \end{bmatrix}$$

$$V = \begin{bmatrix} x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix}$$

and

$$D = \begin{bmatrix} x_{33} & x_{34} \\ x_{43} & x_{44} \end{bmatrix}$$

Feature Extraction Using Gabor Filter

Feature Extraction is most critical because the particular features made available for discrimination directly influence the efficacy of the classification task. Texture is another important property of images, powerful regional descriptor that helps in the retrieval process.

Gabor filter based methods have been widely used in computer vision, especially for texture analysis.

[4] $C(i,j)$is the co-occurrence matrix with the gray pixels values i and j at a distance d is defined as polar coordinates (d,θ) .

The co-occurrence matrix $C(i,j)$ is define as $C(i,j) =$

$$= card \left\{ \begin{matrix} \left((x_1,y_1)(x_2,y_2)\right) \in (XY) \times (XY) \\ for\ f(x_1,y_1) = i, f(x_2,y_2) = j \\ (x_2,y_2) = (x_1,y_1) + (dcos\theta, dsin\theta) \\ for\ 0 < i, i < N \end{matrix} \right\}$$

Let G is the number of gray-values in the image, then the dimension of co-occurrence matrix with size $N \times N$.

Five features extracted from the co-occurrence matrix. These features reduces space dimensionality. They are

$$Energy = \sum_i \sum_j C(i,j)^2$$

$$Inertia = \sum_i \sum_j (i - j)^2 C(i,j)^2$$

$$Correlation = \frac{\sum_i \sum_j (ij)C(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j}$$

$$Difference\ Moment = \sum_i \sum_j \frac{1}{1 + (i - j)^2} C(i,j)^2$$

Where

$$\mu_i = \sum_i i \sum_j C(i,j)$$

$$\mu_j = \sum_j j \sum_i C(i,j)$$

$\sigma_i$is defined as

$$\sigma_i = \sum_i (i - \mu_i)^2 \sum_j C(i,j)$$

$\sigma_j$is defined as

$$\sigma_j = \sum_i (j - \mu_i)^2 \sum_j C(i,j)$$

The two-dimensional Gabor Filter is defined as

$$Gab(x,y,W,\theta,\sigma_x,\sigma_y) =$$

$$\frac{1}{2\pi\sigma_x\sigma_y} e^{\left[-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right) + jW(xcos\theta + ysin\theta\right]}$$

where $= \sqrt{-1}$, σ_x and σ_y are scaling parameters,

W is radial frequency of the sinusoid and $\theta \in [0,\pi]$ specifies the orientation of Gabor filter.

An image $F(x,y)$filter with $Gab(x,y,W,\theta,\sigma_x,\sigma_y)$ is

$$FGab(x,y,W,\theta,\sigma_x,\sigma_y) =$$

$$\sum_k \sum_l F(x-k, y-l) * Gab(x,y,W,\theta,\sigma_x,\sigma_y)$$

The magnitudes of the Gabor filter represented by three moments.

$$\mu(W,\theta,\sigma_x,\sigma_y) = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} FGab(x,y,W,\theta,\sigma_x,\sigma_y)$$

$$std(W,\theta,\sigma_x,\sigma_y) = \sqrt{\sum_{x=1}^{X} \sum_{y=1}^{Y} \left|\left|FGab(x,y,W,\theta,\sigma_x,\sigma_y)\right| - \mu(W,\theta,\sigma_x,\sigma_y)\right|^2}$$

$$skew = \frac{1}{XY} \times \sum_{x=1}^{X} \sum_{y=1}^{Y} \frac{\left(FGab(x,y,W,\theta,\sigma_x,\sigma_y) - \mu(W,\theta,\sigma_x,\sigma_y)\right)^3}{std(W,\theta,\sigma_x,\sigma_y)}$$

The feature vector is constructed using $\mu(W,\theta,\sigma_x,\sigma_y), std(W,\theta,\sigma_x,\sigma_y)$and $Skew$ as the feature components.

Sobel Edge Detector

An edge is defined by a discontinuity in gray-level values. The pixel's gray-level which value is similar to other around pixel's gray-level, there is probably not an edge at that point [5].

Sobel method is applied to perform edge detection. The Sobel edge detector use two masks with 3x3 sizes, one estimating the gradient in the x-direction and the other estimating the gradient in the y-direction [6].The gradient of x-direction(G_x) and y-direction(G_y) is calculating by3×3 matrix. The matrix is

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

The x-direction(G_x) and y-direction $(G_y)$are calculated and combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient. to compute the edge magnitude which given by:

$$|G| = \sqrt{G_x^2 + G_y^2}$$

Shape

Shape is an important visual feature and it is one of the primitive features for image content description. Shape based image retrieval is the measuring of similarity between shapes represented by their features.

The overall performance of shape descriptors can be divided into qualitative and quantitative Performances. The qualitative characteristicsinvolve their retrievalperformance based on the captured shape details for representation. Their quantitative performance includes the amount of data required to be indexed in terms of number of descriptors, in order to meet certain qualitative standards as well as their retrieval computational cost [7] [8].

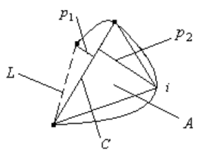

Figure 1: shape and measures to compute features

Shape descriptors can be divided into two main categories: regionbased and contour-based methods. Region-based methods use the whole area of an object for shape description. Contour-based methods use only the information present in the contour of an object.To extract and store a set of shape features fromthe contour image and for each individual contour.Thesefeatures are:

1. Circularity

$$Cir = \frac{4pA}{P^2}$$

2. Aspect Ratio

$$ar = \frac{p_1 + p_2}{C}$$

3. Discontinuity Angle Irregularity

$$dar = \sqrt{\frac{\sum|\theta_i - \theta_{i+1}|}{2\pi(n-2)}}$$

4. Length irregularity

$$lir = \frac{\sum|L_i - L_{i+1}|}{K}$$

Where *K=2P for n>3* and *K=P for n=3*

5. Complexity

$$com = 10^{-\frac{3}{n}}$$

6. Right Angleness

$$ra = \frac{r}{n}$$

7. Sharpness

$$sh = \sum \frac{\max\left(0.1 - \left(\frac{2|\theta-\pi|}{\pi}\right)^2\right.}{n}$$

8. Directedness

$$dir = \frac{M}{\sum P_i}$$

Where n is number of sides of polygon enclosed bysegment boundary. A is area of polygon enclosed bysegment boundaryP is perimeter of polygon enclosed by segment boundary.C is length of longest boundarychord.p1, p2 are the greatest perpendicular distances fromlongest chord to boundary, in each half-space either side of line through longest chord. θ_i is discontinuity anglebetween$(i-1)^{th}$and i^{th} boundary segment.r is numberof discontinuity angles equal to a right-angle within a specified tolerance.Mis total length of straightlinesegments parallel to mode direction of straight-linesegments within a specified tolerance.

IV. Feature Selection

Feature selection methods try to pick a subset of features that are relevant to the target concept. It is gives a better sight into the real-world classification problem [10] [11].

Idealized: find the minimally sized feature subset that is necessary and sufficient to the target concept [12].

Classical: select a subset of M features from a set of N features, M<N, such that the value of a criterion function is optimized over all subsets of size M [13].

Improving Prediction accuracy: the aim of feature selection is to choose a subset of features for improving prediction accuracy or decreasing the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features [11].

Approximating original class distribution: the goal of feature selection is to select a small subset such that the resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution given all feature values [11].

The main criteria of the feature selection are classification accuracy should not decrease and the selected features class distribution should be close to the class distribution of all the features. Feature selection methods searches for best features using some evaluation function [14].

The usefulness of selecting subsets of variables that together have good predictive power, as opposed to ranking variables according to their individual predictive power [15]. In literature, feature selection methods are wrappers, filters, and embedded methods. Wrappers utilize the learning machine of interest as a black box to score subsets of variable according to their predictive power. Filters select subsets of variables as a pre-processing step, independently of the chosen predictor. Embedded methods perform variable selection in the process of training and are usually specific to given learning machines.

There are four basic steps in a typical feature selection method[16] [17]. 1. A generation procedure to generate the next candidate subset; 2. An evaluation function to evaluate the subset under examination; 3.A stopping criterion to decide when to stop; and 4.A validation procedure to check whether the subset is valid.

V. Fisher Score

Fisher score is one of the most widely used supervised feature selection methods. The key idea of Fisher score is to find a subset of featuressuch that in the data space spanned by the selected features, the distances between data points in di□erent classes are as large as possible, while the distances between data points in the same class are as small as possible. In particular, given the selected m features, the input data matrix$X \in \mathbb{R}^{d\times n}$ reduces to $X \in \mathbb{R}^{m\times n}$ [18].

Then the Fisher Score is computed as follows,

$$F(Z) = tr\{(\bar{S}_b)(\bar{S}_t + \gamma I)^{-1}\}$$

whereγ is the positive regularization parameter, $\bar{S}_b$is called between-class scatter matrix and $\bar{S}_t$is called as total scatter matrix. They defined as

$$\bar{S}_b = \sum_{k=1}^{c} n_k\,(\bar{\mu}_k - \bar{\mu})(\bar{\mu}_k - \bar{\mu})^T$$

$$\bar{S}_t = \sum_{i=1}^{n} (z_i - \bar{\mu})(z_i - \bar{\mu})^T$$

where$\bar{\mu}_k$and n_k are the mean vector and size of the k-th class respectively in the reduced data space, i.e., z.

The overall mean vector of the reduced data is $\bar{\mu} = \sum_{k=1}^{c} n_k\,\bar{\mu}_k$. Let μ_k^j and σ_k^j are the mean and standard deviation of k-th class corresponding to the j-th feature. Then the Fisher score of the j-th feature is

$$F(x^j) = \frac{\sum_{k=1}^{c} n_k(\mu_k^j - \mu^j)^2}{(\sigma^j)^2}$$

where $(\sigma^j)^2 = \sum_{k=1}^{c} n_k(\mu_k^j)^2$. The feature with top score are selected by the Fisher Score.

Algorithm 1:Generalized Fisher Score for Feature Selection

Input:C and m;

Output:V and Ω;

Initialize $V = \frac{1}{n}1_n 1_c^T$and $t = 1$;

Find the most violated constraintp^1, and set $\Omega_1 = \{p^1\}$***repeat***

Initialize $\lambda = \frac{1}{t}1$;

repeat

Solve for V using below Equation

$$V = \left(\frac{1}{\gamma}\sum_{t=1}^{|p|}\lambda_t \sum_{j=1}^{d} p_j^t K_j + I\right)^{-1} H$$

under the current λ;

Solve for λ using gradient descent as in

$$\nabla_{\lambda_t g}(\lambda, V) = -\frac{1}{2\gamma}\, tr\left(V^T \textstyle\sum_{j=1}^{d} p_j^t K_j\, V\right);$$

until converge

Find the most violated constraintp^{t+1} and set

$\Omega_{t+1} = \Omega_t \,\cup\, p^{t+1}$;

$t = t + 1$;

untilconverge.

Maximum Relevance And Minimum Redundancy (mRMR)

The mRMR method used for feature selection. Mutual information (MI), which measures the mutual dependence of two variables, is used to quantify both relevance and redundancy in this method. MI is defined as following [19]

$$I(X,Y) = \iint p(x,y) log\frac{p(x,y)}{p(x)p(y)}$$

where X, Y are vectors, p(x, y) is the joint probabilistic density,p(x)andp(y)arethemarginalprobabilisticdensities.Given M data points drawn from the joint probability distribution $(x_i, y_i),\ i = 1, 2, \ldots M$ the joint and marginal densities can be estimated by the Gaussian kernel estimator as following

$$p(x,y) = \frac{1}{M}\sum \frac{1}{2\pi h^2} e^{-\frac{1}{2h^2}((x-x_i)^2 + (y-y_i)^2)}$$

$$p(x) = \frac{1}{M}\sum \frac{1}{2\pi h^2} e^{-\frac{1}{2h^2}(x-x_i)^2}$$

$$p(y) = \frac{1}{M}\sum \frac{1}{2\pi h^2} e^{-\frac{1}{2h^2}(y-y_i)^2}$$

whereh is a tuning parameter that controls the width of the kernels.

VI. Conclusion

Content based image retrieval to retrieve the images for diagnostic cases, to require less space and low bandwidth for transmit. Har transform technique is widely used for wavelet analysis.Edge and texture are extracted from compressed medical images using Sobel Edge Detection and Gabor Filter.Sobel edge detectors are very sensitive to the noise the pixels Image classification accuracy obtained using soft computing techniques on compressed medical image retrieval system. Fisher Score is used for feature selection.

References

[1] KamrulHasanTalukder, Koichi Harada, Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image, IAENG International Journal of Applied Mathematics, 36:1, IJAM_36_1_9.

[2] Haar A.: ZurTheorie der orthogonalenFunktionensysteme. MathematischeAnnalen, 69, 331–371. 1973

[3] Application of CART and IBL for Image Retrieval, VamsidharEnireddy and Dr.Kiran Kumar Reddi, International Journal of Computer Science and Telecommunications [Volume 3, Issue 12, December 2012]

[4] Ryszard S. Chora´, Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems, International Journal of Biology And Biomedical Engineering.

[5] Comparison of Canny and Sobel Edge Detection in MRI Images, Zolqernine OthmanHabibollahHaron ,Mohammed Rafiq Abdul Kadir.

[6] Robert P. Velthuizen, Lawrence 0. Hallt, Laurence

P. Clarke (1996), MRI Feature Extraction Using Genetic Algorithms, 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam, pp 1138-1139.

[7] Amanatiadi V.G. Kaburlasos A. Gasteratos S.E. Papadakis, ―Evaluation of shape descriptors for shapebased image retrieval‖,IET Image Process., 2011, Vol. 5, Iss. 5, pp. 493–499

[8] Belongie S., Malik, J., Puzicha, J., ―Shape matching and object recognition using shape contexts‖, IEEE Trans. Pattern Anal. Mach. Intell., 2002, 24, (4), pp. 509–522

[9] Feature Selection for Classification, M. Dash, H. Liu,

[10] Kohavi, R. and Sommetlield, D., Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. In: Proceedings of First International Conference on Knowledge Discovery and Data Mining, Morgan Kaufmann, 192-197, 1995.

[11] Koller, D. and Sahami, M., Toward optimal feature selection. In: Proceedings of International Conference on Machine Learning, 1996.

[12] .Dash,H.Liu "Feature Selection for Classification" in Intelligent Data Analysis 1 (1997) 131-156.

[13] Narendra, P.M. and Fukunaga, K., A branch and bound algorithm for feature selection. IEEE Transactions on Computers, C-26(9):917–922, September 1977.

[14] Application of Fisher Score and mRMR Techniques for Feature Selection in Compressed Medical Images VamsidharEnireddy, Dr.D.V.V.S.PhaniKumar, Dr.Gunna Kishore, International Journal of Engineering and Technology (IJET), e-ISSN : 0975-4024.

[15] An Introduction to Variable and Feature Selection, Isabelle Guyon, Andr´eElisseeff, Journal of Machine Learning Research 3 (2003) 1157-1182

[16] Siedlecki, W. and Sklansky, J., On automatic feature selection. International Journal of Pattern Recognition and Artificial Intelligence, 2:197–220.

[17] Langley, P., Selection of relevant features in machine learning. In: Proceedings of the AAAI Fall Symposium on Relevance, 1–5.

[18] Generalized Fisher Score for Feature Selection, QuanquanGu,Zhenhui Li Dept. Jiawei.

[19] Prediction of lysine ubiquitination with mRMR feature selection and analysis, YudongCai, Tao Huang, Lele Hu, Xiaohe Shi, Lu Xie, Yixue Li.

[20] Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., &Satyanarayanan, M. (2010).A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(1), 30-44

[21] Syam, B., Victor, J. S. R., &Rao, Y. S. (2013, March). Efficient similarity measure via Genetic algorithm for content based medical image retrieval with extensive features. In Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on (pp. 704-711). IEEE.

[22] Kumar, M. S., &Kumaraswamy, Y. S. (2012, March). An improved support vector machine kernel for medical image retrieval system. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 257-260). IEEE.

[23] Peter, N. B., Jooao, P. H., & David, J. K. (1997). EigenfacevsFisherfaceRecognitionusingclass specific linear projection.IEEE Transactions on pattern Analysis and machineIntelligence, 19(7), 711-720.

[24] Zhuang, X.-S, Dai, D.-Q, and Yuen, P.C, (2005). Face Recognition by Inverse Fisher Discriminant Feature. Springer-Verlag Berlin Heidelberg, 92-98.