SAJET

South Asian Journal of Engineering and Technology Vol.2, No.12 (2016) 1–8

ISSN Number (online): 2454-9614

A Survey on Fisher Criterion based Genetic Algorithm for Feature Selection Method

T.Sowmi, E.Saravana Kumar*

Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India.

*Corresponding Author: (E.Saravana Kumar)
E-mail address: saraninfo@gmail.com
Phone: +91 9894634885

Received: 27/12/2015, Revised:04 /02/2016 and Accepted: 05/02/2016

Abstract

With the presence of large amount of data in the database, retrieval process based on feature selection method became a major research issue in multimedia community because of its computational complexity and memory efficiency. In this paper, the combination of Fisher Criterion based Genetic Algorithm to find the fittest feature subset from the feature vector were surveyed. Different forms of hashing technique, the genetic algorithm improves the computational speed of estimating the fittest solution. The most consequently used extraction methods were GIST, SIFT and Bag-of-words to extract the feature vector from the images. By applying the Fisher criterion algorithm the fittest solutions were obtained and reduces the dimensionality of the image. The Extensive experiments were mostly carried out in Matlab using four commonly used datasets as MIR Flickr, CIFAR-10, NUS-Wide and SIFT-1M.

Keywords: Fisher Criterion, Genetic Algorithm, Fittest Solution, Hashing.

1.Introduction

Data mining predicts the relevant information from the large database and which became the most important technique in industry as well as in society. In recent years, Data mining plays a vital role in companies to focus on the needed information in their data warehouse. Data mining tool extracts the information exactly where the experts may also miss because it may lies outside their expectations [28]. This mining technique proceeds with the long process in terms of five technologies as Selection, Pre-processing, Transformation, Evaluation and Interpretation where the selection process proceeds with searching for patterns in the data and pre-processing handles the missing data fields [29]. From various kinds of mining techniques, Image based mining presented a great deal of attention among the researchers. Particularly Image based retrieval has aroused a research issue in multimedia community. Under data mining, Image mining is considered as one of its type. Image mining proceeds with predicting the relevant knowledge from the image which is not explicitly found in the images based on the feature extraction methods for image retrieval. Nowadays retrieving the relevant images for the given query image had undergone with various optimization techniques. In Image mining it has undergone two approaches. First is to extract the images from the database. Second is to combine the alphanumeric characters and to collect the images. Form various optimization techniques most frequently used technique was hashing.

In hashing based optimization technique for the collection of images in the datasets it creates the data matrix based on the feature extraction method. In general for Image retrieval it contains various image based extraction method which converts the image into single feature vector. Then by using the Linear Sequential Hashing (LSH) method [23] the decimal values are converted to binary string as 0 and 1 which normally reduces the dimensionality of the image. Neighbourhood Discriminant Hashing (NDH) technique [24] was used to update the transformation matrix in terms of projected data and Euclidean distance (i.e., Pixel difference). By

updating the transformation matrix new objective function was estimated for each iteration. Based on the estimation of Euclidean distance the relevant images are retrieved. But in this optimization technique the computational complexity was high as well as to improve the efficiency of true positive and false negative rates genetic algorithm is considered.

In Genetic algorithm (GA), Fisher Criterion (FIG) based optimization technique [1] was considered. This algorithm estimates the fitness function in terms of estimating the average weight for the test image as well as for all training examples by calculating the mean for all training images in the dataset. The fitness function is estimated for reducing the dimensionality of the image in terms of mean and weight calculation. In GA first it initialize the population in terms of chromosomes (i.e., Feature values) by considering the average weight of the individuals. Second is selection process which selects the parent among the chromosome by estimating the probability in terms of individuals. Third is crossover which finds the new individuals by considering the parent chromosome. Fourth is mutation which proceeds as same as the crossover process. Finally, by predicting the threshold value the process gets stop or it proceeds among certain iteration.

The remainder of this paper is followed as. The combination of fisher criterion and genetic algorithm are surveyed in Section 2. The feature extraction methods are described in Section 3 and the working of fitness function based fisher criterion are presented in Section 4. The description of genetic algorithm are declared in Section 5 and the survey on usage of datasets are described in Section 6. The research questions based on the retrieval process and their observation are followed in Section 7 and 8. The overall discussion of the survey are described in Section 9 and finally the conclusion is presented in Section 10.

2. Related Work

In this literature, the feature selection techniques are survived widely in means of fisher criterion based genetic optimization algorithm are presented as follows and the outline of the survey are described in table 1.

Xiabi Liu et al (2015) proposed feature selection method and genetic optimization algorithm [1]. They selected the fittest features among the various extraction method interms of bag-of-words, wavelet transform method and histogram. They also used classifiers as SVM, KNN and Naive Bayes to label the images in the dataset. Based on that classification the retrieval process was processed. Among these classifiers SVM classifier performed best in classifying the labels among the classes. But KNN and Naive Bayes classifiers lagged in their performance.

Jiang Suhua et al (2015) determined two Dimension Threshold Image Segmentation Based on Improved Artificial Fish-Swarm algorithm to improve the segmentation threshold algorithm [2]. While compared with other algorithm swarm produced the better segmentation result. But in this algorithm the optimal solution was not obtained for segmentation technique. E.Saravana Kumar et al (2012) proposed the Performance and Accuracy related issues of Content-Based Image Retrieval [27]. They showed that the image retrieval was based on feature selection and usually the image was represented in the feature vectors.

Feiping Nie et al (2008) used trace ratio criterion [3] for feature selection. In this selection algorithm they proposed two popular feature selection algorithm as Laplacian score and fisher score. They directly optimized the score of the selection technique for the feature subset. For some cases the optimal solution was not obtained. Li Zhuo and Jing Zheng (2008) designed a genetic algorithm based wrapper feature selection method for classification of hyper spectral images using support vector machine [4]. At same time the feature subsets and SVM kernel parameters were optimized by using this technique. GA-SVM method was significant in reducing the computational complexity and also improves the classification accuracy. But this algorithm regretted more computer memory. Medical image analysis in artificial neural network was proposed by J. Jiang, P. Trundle et al (2010). They showed the basic concept of neural network for easy understanding of working principle in the artificial neural network [5]. Non-rigid registration of medical images were used to register a patient's data to an anatomical atlas. But in common the neural network was difficult to interpret and analyse. In some cases they define the process of transforming the input to output in simple manner.

Anup R. Aswar, Kunda P. Nagarikar et al (2014) proposed Texture-Based Identification and Characterization of Pneumonia Patterns in Lung [6]. Identification and characterization of diffuse parenchyma lung disease (DPLD) patterns was difficult. So they presented an automated scheme for volumetric quantification of interstitial pneumonia (IP) patterns which is considered as a subset of DPLD. This algorithm gave a deep understanding of feature selection technique. FCM considered images as separate points. Because the fuzzy function does not considered the spatial dependence.

3. Feature Extraction Method

In this survey, the image based feature extraction method are revised which includes the most commonly used extraction method as GIST, SIFT, Pixel and Bag-of-word as follows.

Matthijs Douze, Herve Jegou et al (2009) described the evaluation of GIST descriptors for web-scale image search [7]. They proposed this technique to show the search accuracy and global GIST descriptors for

feature extraction. This descriptor was used to retrieve the initial set of images from the same landmarks and it was mainly used for the image completion.

Table 1. Combination of Fisher criterion based Genetic Algorithm and Feature Extraction method

Year	Publication title	Methodology	Performance metrics	Datasets	Experimental method
2004	The Distinctive Image Features from Scale-Invariant Key points	Detects the keypoints in the images and filters the keypoints using Laplacian function.	Recall, code length and Precision	GIST1M	Matlab
2008	Trace Ratio Criterion for Feature Selection	Used two popular feature selection algorithm as Laplacian score and fisher score to estimate the fittest solution.	Precision and Recall	GIST-1M and GIST-75M	Matlab
2008	A Genetic Algorithm Based Wrapper Feature Selection Method for Classification of Hyperspectral Images using Support Vector Machine (SVM)	Optimized feature subset and SVM kernel parameters at the same time.	Precision and Recall	CIFAR-10	Matlab
2009	The Evaluation of GIST descriptors for web-scale image search	Used to retrieve the initial set of images from the same landmarks and it was mainly used for the image completion.	Mean Average Precision (MAP)	MIR Flickr	Study
2010	Medical image analysis with artificial neural networks	Non rigid registration also used to register the patients data.	Precision and Recall	CIFAR10 and Flickr image	Matlab
2013	Traffic scene classification using GIST descriptor	It focused on the outline of the image and the properties by discarding their relationships and scene of the local objects.	Precision and Recall	NUS-WIDE	Study
2014	Texture-Based Identification and Characterization of Pneumonia Patterns in Lung	Proposed automated scheme for volumetric quantification of interstitial pneumonia (IP) patterns.	Recall	Photo Tourism and MIR-Flickr	Matlab
2015	Recognizing Common CT Imaging Signs of Lung Diseases Through a New Feature Selection Method Based on Fisher Criterion and Genetic Optimization	Used feature selection method to obtain the fittest solution and used classifiers to label the classes.	Precision and Recall	80 million images from the Internet	Study
2015	Two Dimension Threshold Image Segmentation Based on Improved Artificial Fish-Swarm Algorithm	Improved the image threshold segmentation problem.	Precision and Recall	Caltech 101 and Photo Tourism	Matlab

For large scale database it creates the patches among the images by finding the similar region in the images. They also developed different strategies for the compression of the GIST descriptors. Ivan Sikiric and Karla Brkic (2013) investigated the traffic scene classification using GIST descriptor [8]. It focused on the outline of the image and the properties by discarding their relationships and scene of the local objects. It showed better performance in the classification of the scene setups. The images were represented in the set of statistical properties as naturalness or roughness and openness.

The Distinctive Image Features from Scale-Invariant Key points was determined [9] by David G. Lowe (2004). They presented the SIFT descriptor to extract the reliable keypoints from the image. This process involved filtering the scalable keypoints and also performed the orientation technique interms of location and 3D recognition. Jialu Lin proposed the Bag-of–words model for image retrieval process. They showed that the keypoints are detected and the similar points are formed in patches. Each patches describes certain visual words which has been called as vocabulary generation. The difference in using this feature in text and image indicates that in text based on the language context the text words are sampled naturally but in image retrieval the visual words were described as the outcome of quantization. Karen Glocer, Damian Eads and James Theiler (2005) designed the online feature selection for pixel classification [10]. They proposed to calculate the mean for the pixel values in the images to normalize in the feature vector. The pixel values vary in the range from 0-255 values. They applies the Gaussian smoothing operator to find the weightage of the raw input image.

4. Fitness Function based Fisher Criterion

The fitness function [21] was estimated to find the fittest solution (i.e., optimal feature subset). The estimation of fitness function are survived by considering the various authors techniques as follows.

Marryam Murtaza et al (2014) proposed the Fisher's Criterion and Linear Discriminant Analysis for Face recognition [11]. They showed that this algorithm overcomes the inadequacy of Linear Discriminant Analysis (LDA) and Maximum Margin Criterion (MMC) which is a form of conventional LDA. It fought against the singularity of within class scatter matrix under the reasonable computational cost. LDA was considered as a supervised batch classifier which converts the high dimensional input data to the low dimensional data. It performed high computational cost in terms of large amount of data. The number of samples in the intra class is smaller than the dimensionality of the samples. The combination of LDA and MMC reduces the computational complexity in the feature free subspace. Using the minimum Redundancy Maximum Relevance (mRMR) algorithm, the computational complexity was reduced by reformulating within the class scatter matrix.

Quanquan Gu et al (2010) generalized the fisher score for feature selection. Fisher score was determined as the supervised feature selection strategy [12]. It selects the optimal features independently based upon their score which has been estimated by the fisher criterion. It formed the suboptimal feature subset regarding the scores of the feature selection which maximize the traditional fisher score. The filter based fisher criterion was usually derived as a binary selection of features which maximize the performance of the selection process. The fisher score was calculated in terms of considering the distance between local points. From the fisher score the top ranked n numbers were selected because the scores were determined independently it neglects the combination of feature. The selection procedure deals with heuristic algorithm which was suboptimal solution. But this process cannot handle redundant features. Zhi-Wei Hou et al (2010) proposed Kernelized Fuzzy Fisher Criterion based clustering Algorithm [13]. Clustering was performed in kernel free space where normally clustering takes in the Euclidean space. The segmentation algorithm found that the global optimum solution instead of local solution [14] by Jing Kong (2009) which was considered as a practical and effective segmentation algorithm.

5. Genetic Algorithm

Genetic algorithm was considered as a Heuristic search algorithm which undergoes 5 terminologies to obtain the optimal feature subset [22]. The survey on genetic algorithm are described as follows.

Bir Bhanu et al [15] described the self-optimizing Image segmentation system using a genetic algorithm. The genetic algorithm incorporates with the self-optimizing technique to adapt the segmentation process. Generally the segmentation problem was considered as an optimization problem and it was difficult task of any automated image understanding process. To overcome the segmentation problem, the genetic algorithm efficiently searches the hyperspace of segmentation parameter. Genetic algorithm was designed efficiently to find the approximate global maximum solution. They used existing high quality individuals with simple recombination technique. In this process after getting the image it analyze and finds the characteristics of the image and passes the information along with the external variables to the genetic learning component.

Attakitmongcol K and Srikaew A (2005) proposed a new approach for optimization in watermarking by using genetic algorithm [16]. They used discrete Multiwavelet transform to propose the spread spectrum

image watermarking algorithm. They improved the visual quality of watermarked images and robustness of the watermark. Khaled Loukhaoukha et al (2010) described Multi-objective Genetic Algorithm for Image watermarking based on singular value decomposition and Lifting Wavelet transform [17]. Multiple Scaling Factors were used to achieve the highest robustness without losing watermark transparency. But determining the optimal values for Multiple Scaling Factors was quite difficult.

6. Datasets

The most commonly used datasets in image based classification are described as follows. The MIR Flickr Retrieval Evaluation presented a collection of comprised 25000 images which has been collected from the Flickr website by Mark J. Huiskes and Michael S. Lew (2008) in social photography [18]. The images are redistributable and contains both image contents and image tags which was used for research purposes as well as for community. The color images in Flickr website belongs to generic domain and with high quality. Antonio Torralba, Rob Fergus and William T. Freeman (2008) derived 80 million tiny images: a large dataset for non-parametric object and scene recognition [19].

The images in the dataset were of 32X32 pixels and represented in independent labels using Wordnet lexical database. The categories of image provided the comprehensive of all object categories and scenes. The semantic information in the Wordnet were used in conjunction with the nearest neighbour method which performs the object classification technique. The accuracy depend on the engine used and the specificity of the term used for querying. T.S. Chua, J.Tang and Y. Zheng (2009) survived a real-world Web image database from national university of Singapore [20]. NUS's Lab created a web image dataset for Media Search.

7. Research Method

The survey is processed under image based retrieval process in means of feature selection method. The research questions are described as follows which provides the clarity for the survey.

7.1 Research Questions

The research questions is based on fisher criterion for feature selection and also about genetic algorithm. They are described as follows.

7.1.1. Why genetic algorithm is preferred for image based retrieval process?

Motivation: To define the stopping criterion in more efficient manner and in order to achieve the convergence effectively.

7.1.2. What are the process in GA?

Motivation: GA is a search heuristic approach. Genetic algorithm proceeds with five terminologies as population initialization, selection, crossover, mutation and termination.

7.1.3. Why fitness function is estimated?

Motivation: To identify the fittest solution among the features for the reduction of dimensionality.

7.1.4. What are the function of FIG?

Motivation: Calculates the mean and average weight among the classes to estimate the fitness function.

7.1.5. What are the parameters considered for retrieval process?

Motivation: For image retrieval process most commonly used parameter were precision and recall.

8. Observation

From the survey on feature selection method, the research answers are obtained for the research questions and are described as follows.

8.1. Why genetic algorithm is preferred for image based retrieval process?

Genetic algorithm is considered to solve the optimization problem which comes under the evolutionary criterion. This algorithm is mainly preferred for the search technique. While comparing with conventional criterion method evolutionary method is suitable for searching technique [22]. Because genetic algorithm is efficiently applicable for searching technique in large state space, n dimensional state space or multi model state space and also applicable even the inputs are slightly changed. Genetic Algorithm offers significant searching process among many optimization technique as depth-first, breath-first and heuristic. GA pretend to survive the fittest among the individual over consecutive generation.

8.2. What are the process in GA?

Genetic Algorithm is a search heuristic approach which is used to generate the useful solutions to optimization problems and search techniques. This algorithm belongs to the evolutionary algorithm which generates candidate solutions in terms of fitness solution and which is applicable for solving both constrained and unconstrained problems. Genetic Algorithm is followed in 3 forms. First is to define the objective function, second is to implement the genetic representation and third is to implement the genetic operators. The procedure of estimating the fitness function became time consuming and it affects the GA efficiency [1]. In GA the strings

are represented in form of binary string as 0 and 1. Each bit in the string represents the feature in the images. In previous work the string 1 represents that the feature is selected and 0 represents that the feature is discarded. To vary from this approach newly weight has been calculated in terms of estimating the mean among the classes.

8.2.1 Genetic Process

GA proceeds with 5 steps in terms of initialization, selection, crossover, mutation and termination. In first step the population is initialized based on the individual chromosome in the classes by considering the feature vectors. Then to select the parent chromosome (i.e., fittest chromosome) among the individual the probability is estimated by predicting the fitness values for each individual in the classes. To generate new offspring from the parent chromosome crossover operation is performed. In crossover operator the two individual chromosome (i.e., two parent chromosome) mates to generate two new individuals. The fitness value is calculated for the newly generated individuals. But the size of the search space is static so the fitness value of new individuals are compared with the remaining individuals and the low fittest value are replaced with the new individuals. Mutation operator is processed as same as the crossover operator which eliminates the low probability among the new individuals. Finally termination operator is processed by fixing the iteration process or else by determining the threshold value based upon the fitness function.

8.3. Why fitness function is estimated?

A fitness function is considered as a type of objective function in the field of genetic algorithm. Each chromosome is represented in binary string. After each iteration the concept of genetic algorithm is to delete the low design chromosomes with best design chromosomes [21]. Each iteration indicate the close specification which has been generated by the fitness function by considering the mean and weight of overall classes and the members in the classes. The fitness function determines the fittest values among the features which is used to reduce the dimensionality of the features in the image. Based on the fittest value which has been estimated by mean and weight of all classes the selection operation is performed. The fittest value which are leading as high are selected by the roulette wheel technique.

8.4. What are the function of FIG?

Feature selection was considered as one of the most important issue in pattern recognition, data mining and machine learning. Moreover FIG was used to identify the important (i.e., relevant and Irrelevant) feature subset from the feature vector which has been estimated by the extraction method. It tends to remove the noisy or irrelevant feature form the feature vector and calculates the mean and average weight for the images in the classes to find the optimal subset [25] which was commonly described by Cheng-Lin Liu (2007). However the fisher function can be simulated in means of considering the averages of diagonals and off-diagonal blocks of a kernel matrix.

8.5. What are the parameters considered for retrieval process?

In image retrieval process, based on the parameters the survival of the output was calculated. The output means that the true positive and false negative rates (i.e., comparison of the relevant and irrelevant images in the dataset). True positive rate means that the comparison of retrieved images with rest of the images in the dataset. False positive rate proceeds the comparison with the actual images in the dataset and with the predicted images.

9. Discussion

Image based retrieval process became the most considering task in many applications as medical science, retrieving the culprits, traffic scenes etc., so this process became the challenging task in multimedia community. Moreover retrieval technique proceeds with feature extraction method on extracting the features from the images based on their intensity, orientation and pixel values. From the extraction method a single feature vector was estimated for easy convergence where the values were represented in decimal described by T.Sowmi et al (2015). By applying the hashing function, it was converted to binary string [27]. In Hashing technique the computational speed for estimating the hashing function was too high. In simulation moreover the true positive and false negative rate will be slightly improved. So to improve the computational speed (i.e., total cost) Fisher criterion based genetic algorithm was considered. The process proceeds with finding the fittest solution among the features. Then fisher criterion was applied to estimate he fitness function in means of reducing the dimensionality of the image. The fitness function was commonly determined by the mean and average weight of the features. In combination with the genetic algorithm it undergoes 5 terminologies as described in section 5. Meanwhile the fitness function for all classes were predicted earlier itself. So the time was consumed on generating the optimal feature subset. By combining the fisher criterion with the genetic

algorithm the computational speed will be improved as well as true positive and false negative rates will also be improved.

10. Conclusion

Due to the development of multimedia community, the storage of data's are increased while comparing with earlier days. So retrieving the images based on the extraction method became complicated. For this complication, many optimization algorithm were used previously. Comparing with the computational speed of the hashing based optimization technique, the genetic algorithm will improve the complexity. From the reviewed papers, mostly the implementation was carried out in Matlab with commonly used four datasets.

References

[1] Xiabi Liu, Ling Ma and Li Song, "Recognizing Common CT Imaging Signs of Lung Diseases Through a New Feature Selection Method Based on Fisher Criterion and Genetic Optimization," IEEE Journal of Biomedical And Health Informatics, Vol. 19, No. 2, March 2015.

[2] Jiang Suhua, Liu Chunqiang and Wang Dongdong, "Two Dimension Threshold Image Segmentation Based on Improved Artificial Fish Swarm Algorithm," International Conference on Chemical, Material and Food Engineering (CMFE-2015).

[3] Feiping Nie, Shiming Xiang, Yangqing Jia, Changshui Zhang and Shuicheng Yan, "Trace Ratio Criterion for Feature Selection," Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (2008).

[4] Li Zhuo, Jing Zheng, "A Genetic Algorithm based Wrapper Feature Selection Method for Classification of Hyperspectral Images Using Support Vector Machine," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008.

[5] J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks," Computerized Medical Imaging and Graphics 34 (2010) 617–631.

[6] Anup R. Aswar, Kunda P. Nagarikar and P. D. Pawar, "Texture-Based Identification of Interstitial Pneumonia Patterns in Lung Multidetector CT," IETE 46th Mid Term Symposium "Impact of Technology on Skill Development" MTS-2015.

[7] Matthijs Douze, Herve Jegou, "Evaluation of GIST descriptors for web-scale image search," CIVR 09, July 8-10, 2009.

[8] Ivan Sikiric and Karla Brkic, "Classifying traffic scenes using the GIST image Descriptor," Proceedings of the Croatian Computer Vision Workshop, Year 1 September 19, 2013.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.

[10] Karen Glocer, Damian Eads and James Theiler, "Online Feature Selection for Pixel Classification," Appearing in Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005.

[11] Marryam Murtaza, Muhammad Sharif, Mudassar Raza, and Jamal Hussain Shah, "Face Recognition Using Adaptive Margin Fisher's Criterion and Linear Discriminant Analysis (AMFC-LDA)," The International Arab Journal of Information Technology, Vol. 11, No. 2, March 2014.

[12] Quanquan Gu, Zhenhui Li and Jiawei Han, "Generalized Fisher Score for Feature Selection," 2010.

[13] Zhi-Wei Hou, Liu-Yang Wang and Quan-Yin Zhu, "Kernelized Fuzzy Fisher Criterion based clustering Algorithm," Distributed Computing and Applications to Business Engineering and science., pp.87-91, 2010.

[14] Jing Kong, "A New Segmentation algorithm with the fisher criterion function," Computing, Communication, Control, and Management, 2009, pp. 63-67.

[15] Bir Bhanu, Sungkee Lee and John Ming, "Self-optimizing Image segmentation system using a genetic algorithm," 2009.

[16] Attakitmongcol K and Srikaew A, "A new approach for optimization in watermaking by using genetic algorithm," IEEE Transaction on Signal Processsing," DOI: 10.1109, Sep 2014.

[17] Khaled Loukhaoukha, Jean-yves Chouinard and Mohamed Haj Taieb, "Multi-objective Genetic Algorithm for Image watermarking based on singular value decomposition and Lifting Wavelet transform," Image and Signal Processsing, July 2010.

[18] M. J. Huiskes and M. S. Lew, "The MIR Flickr retrieval evaluation," in Proc. ACM Int. Conf. Multimedia Inf. Retr., 2008, pp. 39–43.

[19]A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1958–1970, Nov. 2008.

[20] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: A real-world Web image database from National University of Singapore," in Proc. ACM Int. Conf. Image Video Retr., 2009, Art. ID 48.

[21] Dong Xu and Shuicheng Yan. "Marginal Fisher Analysis and its Variant for Human Gait Recognition and Content- Based Image Retrieval," Image processing, Vol. 16, No. 11, 2007.

[22] Chih-chin Lai and Ying chuan chen, "A User- Oriented Image Retrieval System Based on Interactive Genetic Algorithm," Instrumentation and Measurement, Vol. 60, No. 10, 2011.

[23] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in Proc. ACM Symp. Theory Comput., 1998, pp. 604–613.

[24] Jinhui Tang and Zechao Li, "Neighborhood Discriminant Hashing for Large-Scale Image Retrieval," Image Processing, vol. 24, no. 9,

2015.

[25] Cheng-Lin Liu, "Feature Selection by combining Fisher criterion and Principal Feature analysis," Machine Learning and Cybernetics, 2007, pp.1149-1154.
[26] Mr.E. Saravana Kumar,Dr. A. Sumathi and K. Latha, "Performance and Accuracy related issues of Content-Based Image Retrieval," International Journal of Power Control Signal and Computation(IJPCSC) Vol3. No1. Jan-Mar 2012
[27] T. Sowmi and E. Saravana Kumar, "A Survey on Similarity Search for Large Scale Database," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 9, September 2015.
[28] Baker, R.S.J.D.,and Yacef, K.(2009), "The state of Educational Data Mining in 2009:A review and future vision" Journal of Educational Data Mining, Vol.1,No. 1,pp.3-17.
[29] Mercheron, A., and Yacef, K. (2005), "Educational Data Mining: a case study" in Proc. Conf. on Artificial Intelligence in Education Supporting Learning through Intelligent and Socially Informed Technology. IOS Press, Amsterdam, the Netherlands, pp. 467-474.