

An Exploration of Subspace Models and Transformation Techniques for Image Classification

Anusha.T.R¹
PG Student

Hemavathi.N¹
PG Student

Shakunthala. C. H¹
Assist. Professor

¹Dept of ECE, SJB Institute of Technology, Bangalore-560060, India.

anushatr21@gmail.com, hema15sg@gmail.com.

Abstract—In computer vision, image retrieval is a technique which uses visual contents to search images from large scale image databases according to user's interest. In typical image retrieval systems, the visual contents of the images in the database are extracted and described by feature vectors. In this paper, we explore a comparative study for image retrieval system using transformations techniques like DCT to extract low level features. This is then applied to DWT to extract even more low frequency components. The dimensionality reduction is achieved by using PCA from which the feature vectors are extracted which are classified using different distance metrics. The different dataset used in this paper are Caltech-101, Caltech-256, Corel-1K and Corel-10K. Feature vector for the test image is compared with those of the train images. In this experiment, we compared 4 distance measures and their modifications between feature vectors with respect to the recognition rates. The experimental results revealed that the proposed technique produces the better recognition rate compared to other benchmark techniques.

Keywords— Content Based Image Retrieval (CBIR), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), similarity measures.

I. INTRODUCTION

Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Text-based image retrieval [1] uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Obviously, annotating images manually is a cumbersome and expensive task for large image databases, and is often subjective, context-sensitive and incomplete. As a result, it is difficult for the traditional text-based methods to support a variety of task-dependent queries.

The difficulties faced by text-based retrieval became more and more severe. The efficient management of the rapidly expanding visual information became an urgent problem.

This need formed the driving force behind the emergence of content-based image retrieval techniques. Content-based image retrieval [1], a technique which uses visual contents to search images from large scale image databases according to user's interests, has been an active and fast advancing research area since the 1990s. Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image [15].

Color features are linked to the chromatic part of an image. A color histogram provides allotment of colors which is achieved by damaging image color and obtaining the numbers of pixels that fit into every color. Thus the image's color histogram is examined and saved in the database. Retrieval of those images has been done in the matching process whose color allotment matches to the example query [2, 16]. In Texture features, dissimilarity in brightness with high frequencies in the image spectrum are characterized. While making a distinction between areas of the images with same color, these features are very useful. Measures of image texture such as the degree of contrast, directionality, regularity and randomness can be obtained using second-order statistics [18].

In shape, either the global form of the shape or local elements of its boundary, shape features can be differentiated. Global form of the shape: like the area, the extension and the major axis orientation. Local elements of its boundary: like corners, characteristic points or curvature elements [2, 5]. In spatial units like points, lines, regions and objects and their allocation in an image, spatial relationships can be articulated. Spatial features can be classified into directional and topological relationships. Directional relationships: like right, left, above, below together with a distance and topological relationships: like disjunction, adjacency, containment or overlapping of entities [4, 16].

In recent years, the research of developing image retrieval systems has attracted a lot of attention from many different fields. The prior work on various transformation technique, feature extraction techniques, distance measures have been developed. Keerti Keshav Kanchi [6] has developed an algorithm for the facial expression recognition system, which uses two-dimensional discrete cosine transform (2D-DCT) for image compression and the self organizing map(SOM) neural network for recognition purpose on AT&T database which has highest recognition rate. Jianmin et al. [7] proposed a simple, low-cost and fast algorithm to extract dominant colour features directly in DCT domain without involving full decompression to access the pixel data Prabhakar Telagarapu et al. [8] proposed Image Compression Using DCT and Wavelet Transformations by which it is concluded that

overall performance of DWT is better than DCT on the basis of compression rates. This paper can further be extended for line singularities with new transform named Ridgelet Transform.

The wavelet representation decomposition defines an orthogonal multi resolution representation called a wavelet representation. Stephane G. Mallat [9] proposed a theory for multi resolution signal decomposition. It is computed with a pyramidal algorithm based on convolutions with quadrature mirror filters. Jon Shlens [10] proposed a tutorial on PCA derivation, discussion and SVD that clearly explains the magic behind black box. This paper focuses on building a solid intuition for how and why PCA works, further it derives principals behind PCA. Fazal Malik et al. [11] proposed Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain. The proposed method is tested by using Corel image database and the experimental results shows the method has robust image retrieval for various distance metrics with different histogram quantization in a compressed domain. Jinjun Wang et al. [12] proposed Locality-constrained Linear Coding for Image classification. The paper introduces an approximation method to further speed-up the LLC computation, and an optimization method to incrementally learn the LLC codebook using large-scale training descriptors which shows better classification accuracy.

The paper is organized as follows: In section II, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Principal component analysis (PCA), Classification using Distance Metrics are explained. Experimental Results and Performance Analysis are discussed in Section III. Conclusions are drawn at the end.

II. PROPOSED METHODOLOGY

In this paper, the recognition rate is determined by comparing the feature vectors of test and trained images. DCT, DWT and PCA are used for feature extraction and different distance metrics are used for classification. We initially apply DCT algorithm on four popular and widely used benchmarking image datasets (Caltech_101, Caltech_256, Corel_1k, and Corel_10k) to obtain low level features. Later, DWT is applied on the output of DCT to obtain even more low frequency components. PCA is applied to construct feature vector by extracting low frequency components from DCT and DWT. Finally, for classification purpose we use four different similarity distance measure techniques in reduced feature space using PCA. Following subsections furnishes the detailed explanation of DCT and DWT, feature extraction and classification stages.

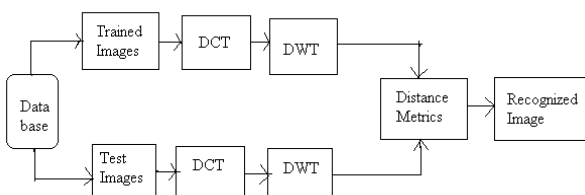


Fig.1 Block Diagram of Image Recognition System

Initially, the images are separated as train and test images from the database, 2D-DCT is applied for the trained images and this DCT co-efficient are given as input to Haar DWT which is followed by PCA and thus the features are extracted. The same procedure is followed for test images. Once the feature vectors are extracted the test and train images are compared using different distance metrics which is used for classification to obtain the best match and corresponding recognition rate.

A. DCT:

The discrete cosine transform (DCT)[7,13] is an algorithm widely used in different applications. The most popular use of the DCT is for data compression, as it forms the basis for the international standard lossy image compression algorithm known as JPEG. The DCT has the property that, for a typical image, most of the visually significant information about the image is concentrated in just a few coefficients. Extracted DCT coefficients can be used as feature vectors useful for image classification. The DCT transforms images from the spatial domain to the frequency domain. Since lower frequencies are more visually significant in an image than higher frequencies, the DCT discards high-frequency coefficients and quantizes the remaining coefficients. This reduces data volume without sacrificing too much image quality.

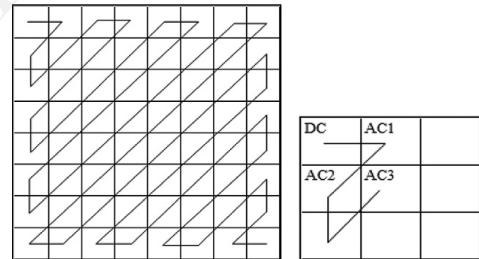


Fig.2 DCT block co-efficient in zigzag order

The 2D-DCT of an $M \times N$ matrix A is defined as follows:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\left(\frac{\pi(2m+1)p}{2m}\right) \cos\left(\frac{\pi(2n+1)q}{2n}\right), \quad 0 \leq p \leq M-1, 0 \leq q \leq N-1 \quad (1)$$

The values B_{pq} are the DCT coefficients. The DCT is an invertible transform, and the 2D Inverse-DCT is defined as follows:

$$A_{mn} = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \alpha_p \alpha_q B_{pq} \cos\left(\frac{\pi(2m+1)p}{2m}\right) \cos\left(\frac{\pi(2n+1)q}{2n}\right), \quad 0 \leq m \leq M-1, 0 \leq n \leq N-1 \quad (2)$$

The values α_p and α_q in (1) and (2) are given by:

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leq p \leq M-1 \end{cases} \quad (3)$$

$$\alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leq q \leq N-1 \end{cases} \quad (4)$$

The $M \times M$ transform matrix T is given by:

$$T_{pq} = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0, 0 \leq q \leq M-1, \end{cases} \quad (5)$$

$$\sqrt{\frac{2}{M}} \cos\left(\frac{\pi(2q+1)p}{2M}\right), 1 \leq p \leq M-1, 0 \leq q \leq M-1$$

B. DWT

The discrete wavelet transform (DWT) [14], is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It is a tool that separates data into different frequency components, and then studies each component with resolution matched to its scale.

DWT is computed with a cascade of filtering followed by a factor 2 sub sampling. *H* and *L* denotes high and low-pass filters respectively, $\downarrow 2$ denotes sub sampling.

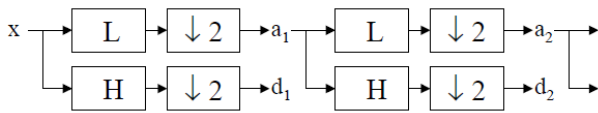


Fig.3 DWT Tree

Outputs of this filter are given by equations (6) and (7)

$$a_{j+1}[p] = \sum_{-\infty}^{\infty} l[n-2p]a_j[n] \tag{6}$$

$$d_{j+1}[p] = \sum_{-\infty}^{\infty} h[n-2p]a_j[n] \tag{7}$$

Elements a_j are used for next step (scale) of the transform and Elements d_j , called wavelet coefficients, determines output of the transform. $l[n]$ and $h[n]$ are coefficients of low and high-pass filters respectively. One can assume that on scale $j+1$ there is only half from number of a and d elements on scale j . This causes that DWT can be done until only two a_j elements remain in the analyzed signal. These elements are called scaling function coefficients.

DWT algorithm for two-dimensional pictures is similar. The DWT is performed firstly for all image rows and then for all columns (Fig.4).

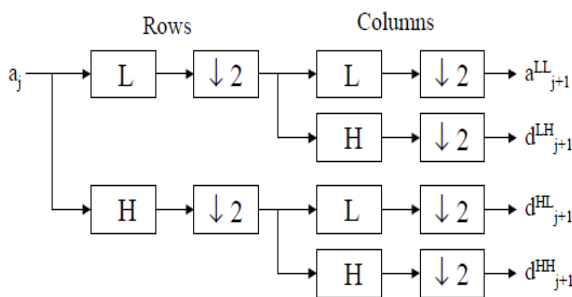


Fig.4 Wavelet Decomposition for 2D picture

The main feature of DWT is multiscale representation of function. By using the wavelets, given function can be analyzed at various levels of resolution. The DWT is also invertible and can be orthogonal. To compute the wavelet features in the first step Haar wavelet is calculated for whole image. As a result of this transform there are 4 sub band images at each scale (Fig5).

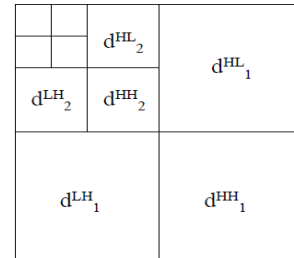


Fig.5 Sub band Images

Sub band image d^{LL} is used only for DWT calculation at the next scale. For the given image, the maximum of 8 scales can be calculated. The Haar wavelet is calculated only if output sub bands have dimensions at least 8 by 8 points. In the next step, energy of d^{LH} , d^{HL} and d^{HH} is calculated at any considered sale in marked ROIs.

$$E_{subband, scale} = \{\sum_{x,y \in ROI} (d_{x,y}^{subband})^2\} / n \tag{9}$$

Where n is the number of pixels in ROI, both at given scale and sub band.

Of course, ROIs are reduced in successive scales in order to correspond to sub band image dimensions. In a given scale the energy is calculated only if ROI at this scale contains at least 4 points. Output of this procedure is a vector of features containing energies of wavelet coefficients calculated in sub bands at successive scales. The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Its scaling function $\phi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The Haar wavelet has several notable properties:

1. Any continuous real function with compact support can be approximated uniformly by linear combinations of $\phi(t), \psi(2t), \phi(4t), \dots, \phi(2^n t), \dots$, and their shifted functions. This extends to those function spaces where any function therein can be approximated by continuous functions.
2. Any continuous real function on $[0, 1]$ can be approximated uniformly on $[0, 1]$ by linear combinations of the constant function $\mathbf{1}, \psi(t), \psi(2t), \psi(3t), \psi(4t), \psi(2^n t), \dots$ and their shifted functions.
3. Orthogonality is of the form

$$\int_{-\infty}^{\infty} 2^{-\frac{n+n_1}{2}} \psi(2t)(2^n t - k), \psi(2^{n_1} t - k_1) dt = \delta_{n,n_1} \delta_{k,k_1}$$

C. PCA

Principal Component Analysis(PCA)[10] was invented by Karl Pearson. It involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. PCA is used to reduce the dimensionality of the data while retaining as much information (but no redundancy) as possible in the original dataset. It is a simple method for

extracting relevant information from huge data set. It is a powerful tool for analyzing data.

The plan for PCA is to take our data and rewrite it in terms of new variables so that our “new data” has all the information from the original data but the redundancy has been removed and it has been organized such that the most important variables are listed first. Since high correlation is a mark of high redundancy, the new data should have low, or even better, zero correlation between pairs of distinct variables. To sort the new variables in terms of importance, we will list them in descending order of variance.

D. Classification

Most statistical analysis classifiers require a distance metric to measure the distance between data points, distance metric is a vital element in clustering. The distance between the feature vector of train image to the feature vector of the query image is computed and compared using statistical metrics. This section will look at some statistic metrics commonly used for making comparisons and classification decisions.

In the proposed work, four distance measures such as Manhattan, Mean square error, Euclidean distance, and Angle based distance metrics are used to classify the image into one of the basic expressions. Let $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ be feature vectors of length n , $p > 0$, $z_i = \sqrt{1/\lambda_i}$ where λ_i is corresponding eigen values ($i = 1 \dots n$), thus the distances between feature vector of train and test images are obtained.

III. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

In this paper, the recognition rate is determined by comparing the feature vectors of test and trained images. DCT, DWT and PCA are used for feature extraction and different distance metrics are used for classification. We initially apply DCT algorithm on four popular and widely used benchmarking image datasets (Caltech-101, Caltech-256, Corel-1k, Corel-10k) to obtain low level features. Later, DWT is applied on the output of DCT to obtain even more low frequency components. PCA is applied to construct feature vector by extracting low frequency components by DCT and DWT. Finally, for classification purpose we use four different similarity distance measure techniques in reduced feature space using PCA.

The dataset which are used in this paper are explained below:

Corel-1K was initially created for CBIR applications comprising 1000 images classified into 10 object classes and 100 images per class. Corel-1K dataset containing natural images such as African tribal people, horse, beach, food items, etc [15, 17]. We experimented the proposed method using 15, 30 images per category as training and remaining for testing. With these the performance of the proposed method is evaluated and compared with the existing methods.

Caltech-101 dataset [22] contains 9144 images in 101 classes including animals, vehicles, flowers, etc, with significant variance in shape. The number of images per category varies from 31 to 800. As suggested by the original dataset and also by many other researchers [12,21], we

partitioned the whole dataset into 15, 30 training images per class and the remaining are test images per class, and the performance is measured using average accuracy over 101 classes.

Caltech-256 dataset [20] consists of images from 256 object categories and is an extension of Caltech-101. It contains from 80 to 827 images per category. The total number of images is 30608. The significance of this database is its large inter-class variability, as well as larger intra-class variability than in Caltech- 101. Moreover there is no alignment amongst the object categories.

Corel-10K dataset consists of images from 108 object categories which includes birds, flowers, fruits etc, with significant variance in shape. The number of images per category varies from 36 to 100 [19]. We partitioned the whole dataset into 15, 30 training images per class and the remaining are test images per class, and the performance is measured using average accuracy over 108 classes.

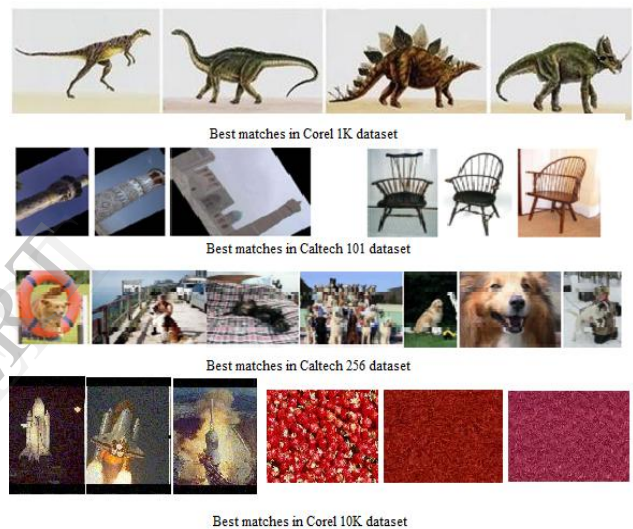


Fig.6. Best matches of different datasets.

These are the few images of different dataset namely Caltech-101, Caltech-256, Corel-1K & Corel-10K which have achieved a maximum recognition rate of 100%, 60%, 100%, 60% respectively.

Performance Analysis of Caltech and Corel datasets is shown below:

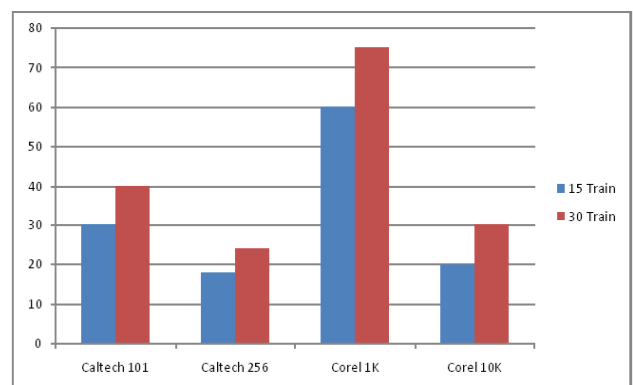


Fig.7 Recognition rate of various datasets.

In the proposed methodology, the recognition rate of 40%, 18%, 60% & 20% for 15 train images and 55%, 24%, 75% & 30% for 30 train images is achieved for Caltech-101, Caltech-256, Corel-1K & Corel-10K respectively. It is observed that the proposed methodology works efficiently on Corel-1K dataset with highest recognition rate.

Based on the below observations we can conclude that experimental results of proposed method is found to be better than the results proposed by Fei-Fei et.al. [22] and Serre et.al. [21] which has an average recognition rate of 18% and 30% respectively.

Comparison of proposed method with exiting methods:

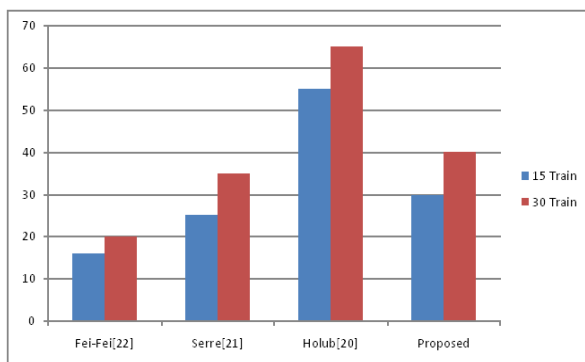


Fig.8 Recognition rate of various methods.

IV. CONCLUSION

This paper presents promising image representation methods called DCT and DWT. In this experiment, we addressed the problem of DCT for image retrieval. The proposed method extracts features from DCT and DWT based methods. For classification purpose, we explored different distance measure techniques and tested there superiority based on different images. The combination of DCT and DWT with Euclidean, Manhattan, Mean Square Error and Angle-based distance are performed. Based on the above observations we can conclude that DWT with various distance metrics gives better recognition compared to DCT with different distance metrics. The proposed method has better recognition rate compared to Fei-Fei et.al. [22] and Serre et.al. [21] which has an average recognition rate of 18% and 30% respectively. The proposed method is found to be competitive with Holub [20].

REFERENCES

- [1] Satrajit Acharya and M.R.Vimala Devi. Image retrieval based on visual attention model. *Procedia Engineering*, 30:542–545, 2012.
- [2] Stehling, R. O., Nascimento, M. A., and A. X. Falcao. On Shapes of Colors` for Content-based Image Retrieval. In *ACM International Workshop on Multimedia Information Retrieval (ACM MIR'00)*, 2000, 171-174.
- [3] M. Flickner et al., Query By Image and Video Content: The QBIC System. *IEEE Computer*, 28, 9 (1995), 23-32.

- [4] Jing, M. Li, H. J. Zhang and B. Zhang, An Effective Region-based Image Retrieval Framework, In *Proceedings of the Tenth ACM international conference on Multimedia*, 2002, 456-465.
- [5] M. Safar, C. Shahabi and X. Sun, Image Retrieval by Shape: A Comparative Study, In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME'00)*, 2000, 141-144.
- [6] Keerti Keshav Kanchi Facial Expression Recognition using Image Processing and Neural Network (IJCSSET) ISSN : 2229-3345 Vol. 4 No. 05 May 2013.
- [7] Jianmin Jiang, Ying Weng, PengJie Li, Dominant colour extraction in DCT domain, *Image and Vision Computing* 24 (2006) 1269–1277.
- [8] Prabhakar.Telagarapu, V.Jagan Naveen, A.Lakshmi, Prasanthi, G.Vijaya Santhi, Image Compression Using DCT and Wavelet Transformations, *International Journal of Signal Processing, Image Processing and Pattern Recognition* Vol. 4, No. 3, September, 2011
- [9] Stephane G. Mallat *IEEE transactions on pattern analysis and machine intelligence*. Vol. Ii. No 7. July 1989
- [10] Jon Shlens A tutorial on PCA derivation, discussion and SVD, Version 1, 25 march 2003
- [11] Fazal Malik, Baharum Baharudin Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Malaysia, 18 November 2012
- [12] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Ls, Thomas Huan, and Yihong Gong Akiir Media System, "Locality-constrained Linear Coding for Image Classification"
- [13] Stepan Obdrzalek and Jiri Matas, Image Retrieval Using Local Compact DCT-based Representation, *DAGM'03, 25th Pattern Recognition Symposium* September 10-12, 2003.
- [14] M. Kociołek, A. Materka, M. Strzelecki P. Szczypiński Discrete wavelet transform –derived features for digital image texture analysis, *Proc. of Interational Conference on Signals and Electronic Systems*, 18-21 September 2001, Lodz, Poland, pp. 163-168.
- [15] Sami Brandt, Jorma Laaksonen and Erkki Oja, Statistical Shape Features in Content-Based Image Retrieval, 2000 IEEE, PP. 1062-1065.
- [16] Guang Hai Liu, Content-based image retrieval using the local structures of color and edge orientation, *Spring world congress on engineering and technology*, 2 (2012) 438-441.
- [17] Tai sing lee, image representation using 2d gabor wavelets, *IEEE transactions on pattern analysis and machine intelligence*, vol. 18, and no. 10, october 1996.
- [18] K.J. Dana, B.van Ginneken, S.K.Nayar, and J.J.Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.*, 18(1):1–34, 1999.

- [19] Ben Steichen a, Helen Ashman b, Vincent Wade, Information Processing and Management 48 (2012) 698–724.
- [20] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report UCB/CSD-04-1366, California Institute of Technology, 2007.
- [21] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In CVPR, 2005.
- [22] L. Fei-Fei, R. Fergus, and P. Perona. An incremental bayesian approach testing on 101 objects categories. In Workshop on Generative-Model Based Vision, CVPR, 2004.

IJERT