

Hybrid Image Retrieval System Using Markov Chain

S.Sabena, P.Yogesh

*Department of Information Science and Technology, CEG Campus, Anna University
sabenazulficker@gmail.com, yogesh@annauniv.edu*

Abstract

The need for efficient retrieval of the images is the need of the hour since the number of image collections available on the Internet is ever growing. The existing systems use the probabilistic approach based on Markov chain to find the relation between the image and the features of the image through the keywords. In this method, for the given text query, the Markov chain is constructed based on the keyword relation and the logical connection between keywords is quantified. This paper proposes a hybrid image retrieval method to improve the process of retrieving the desired images through queries that are based on multiple features of the image. In this method, for the given image query, the Markov chain is constructed with respect to the values of the features extracted from the image. The feature extraction considers the image colour histogram and texture value of the image. As a result, the image retrieval process is improved by considering both the visual and textual features of the image in the construction of the Markov chain.

Keywords: Text-Based Image Retrieval, Content-Based Image Retrieval, Color Moments, Markov chain, Hierarchical clustering.

Introduction

Image retrieval techniques are categorized into two: Text-Based Image Retrieval (TBIR) system and Content-Based Image Retrieval (CBIR) system. Text-based algorithms are keywords based. Annotated keywords are assigned to each image, when images are stored in a database. The annotation operation is time consuming and tedious. Besides being tedious, it is subjective in nature. Since it is subjective, the annotations are sometimes incomplete and hence may miss some features that are required for image retrieval. Content based image retrieval (CBIR) techniques have been proposed to overcome the limitations of TBIR systems. In a CBIR system, images are automatically indexed by their extracted features such as shape, texture, color, size etc. However, these features are considered as low level features since the semantics provided by them is not close to the human perception. Moreover,

extracting all visual features of an image is a difficult task. The semantic gap problem posed by the low level features shall be addressed by mapping the low level concepts to high level visual concepts. This is a difficult task. The semantic gap problem may be alleviated by using a combination of low level and high level concepts.

The proposed work in this paper uses the combination of low level and high level features to improve the performance. The rate of the accuracy of the proposed system has increased considerably in comparison to the text-based and content-based methods. This paper is organized as follows. Section 2 focuses on the related works in the field. In section 3, content-based image retrieval systems have been explained. In section 4, proposed hybrid technique has been explained. Section 5 presents the implementation and experimental results and finally, in section 6, the conclusions have been presented.

Related Work

Annotation-Based Image Retrieval (ABIR) system is an attempt to incorporate more efficient semantic content into both text-based queries and image captions. Markov chain is a probabilistic approach for annotation based image retrieval. In [1], authors follow an approach that constructs the Markov chain based on the caption terms of the image, to quantify the logical connection between the keywords. It fully explores the correlation among the labels of the image.

In some systems, a content-based approach is combined with a text-based approach. As an example, Blobworld system automatically segments each image into regions, which correspond to objects or parts of the objects in an image. In this system, users can view the results of the segmentation of both the query images and can realize from the returned results how the segmented features have influenced the retrieval results [2]. Query by Image and Video Content (QBIC) system supports queries based on example images. The visual features used in the system include colour, texture, and shape. In this system, colour is represented using a k-bin color histogram and the texture is described by an improved Tamura texture [3]. The VisualSEEK system uses both content-based and text-based queries. The system uses color and texture visual features. The color feature is represented by color set and the texture is represented as wavelet transform. The system establishes spatial relationship between image regions with respect to color. A binary tree was used to index the feature vectors [4].

Chabot uses a relational database management system called postgres, which supports search through a combination of text and color [5]. Photobook, computes features vectors for the image characteristics, which are then compared to compute a distance measure utilizing one of the system matching algorithms. System matching algorithms include Euclidean, mahalanobis, divergence, vector space angle, histogram, Fourier peak, wavelet tree distances and user-defined matching algorithms via dynamic code loading [6]. Photobook works by comparing features associated with the images and not the images themselves. In [7], a system has been presented which is a combination of text-based and content-based algorithms. For text retrieval, the Apache Lucene engine has been used and for content-based retrieval, images have

been segmented to different areas and regions and histogram has calculated for each section.

Most of these works suffer from the semantic gap problem and they are not able to achieve the goal of retrieving the images according to the human perception. In this paper, we propose a hybrid framework that uses both low level and high level visual concepts to fill the semantic gap in the context of human perception.

The Proposed System

Since the Content Based Image Retrieval (CBIR) deals only in terms of low level visual concepts, it is unable to fill the semantic gap between the retrieval system and the human perception. Hence we propose the hybrid image retrieval system that can improve the image retrieval process. In this hybrid system, Markov chain is constructed for both visual features as well as textual features. Visual features are based on the image colour, texture and edge of the image. Textual features are based on the image keywords. Section 3.1 explain the hybrid image retrieval system using Markov chain. Section 3.2 describes the construction of Markov chain for caption terms, section 3.3 discusses the Markov chain construction for visual concepts and section 3.4 explains the optimization of Markov chain.

Hybrid image retrieval system using Markov chain:

At first the user gives the text or image query to the system, then the search responds with a list of images. If user gives a text query, the system downloads or retrieves images based on the caption term of the images, then the user accepts the returned images or refines new query instead. The user modifies the query based on the returned images and in this process the image retrieval is getting refined. However, in this process the Markov chain is constructed between the two consecutive queries and then system refines the results based on Markov chain. This process reduces the overhead involved in the process of query refinement. Suppose if the user gives the image query, then the system retrieves the images based on visual concepts of the images and the user accepts the returned images or refines the query with new images. So in this process, the Markov chain is constructed between two consecutive visual concepts and then the search results are refined based on the constructed Markov chain. The architecture of the proposed hybrid image retrieval system is given in figure 1.

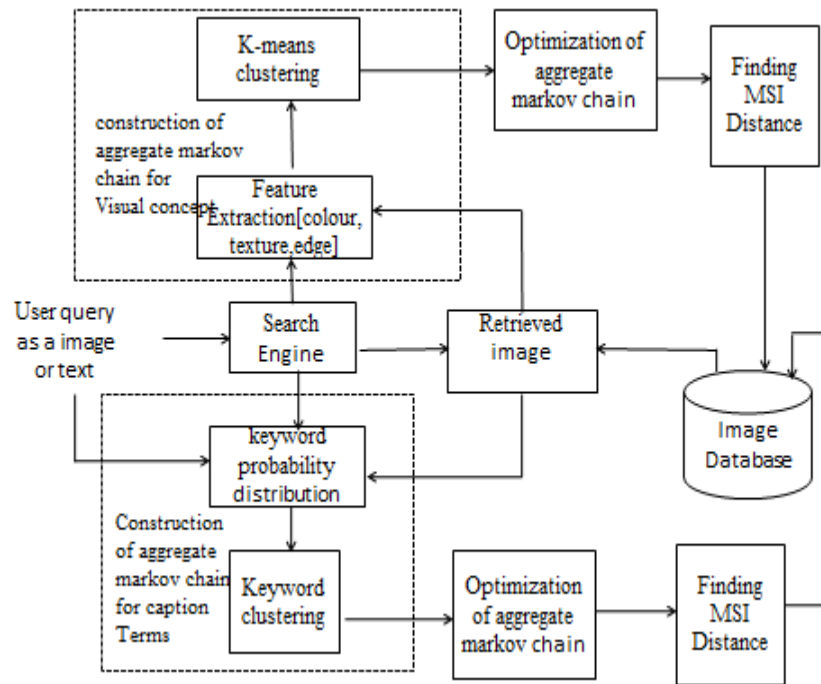


Figure 1: Hybrid Image Retrieval System

Construction of Aggregate Markov chain for textual features

The user submits text query to the system and then system retrieves the images based on the relationship between the query and caption terms. Suppose the retrieved results are not satisfied by the user then the user refines the results by giving new query. Section 3.2.1 explains the probability distribution of caption terms and section 3.2.2 describes the keyword clustering.

Caption terms probability distribution

The user implicitly relates retrieved images to query by assuming Markov chain transitions. If the user relates the image I_i to query q_i , where the caption terms cp_2 follows cp_1 and if this pattern occurs m times then the one step transition probability $p(cp_1, cp_2)$ is being updated using recurrent formula as shown in equation 1.

$$P_i(cp_1, cp_2) = \frac{TP_i(cp_1, cp_2) + t}{T + t} \quad (1)$$

The probability of $p(cp_1, cp_2)$ depends on the T caption terms and the new probability based on $T + t$ caption terms. (What is small t ?)

This procedure constructs a Markov chain where each caption term corresponds to a state. Each time a caption term is presented the state counter is advanced. If another caption term follows in the same query then the interstate link between both caption terms is advanced. Markov chain for each image is denoted by $\pi(i)$.

Keyword clustering

User connects the definite caption terms together and then forms the new query implicitly. These caption terms are related to each other regardless of the images that may or may not be picked by this user. To solve this zero-frequency problem, the keyword space is clustered into similar keywords. For this purpose, the Aggregate Markov Chain (AMC) is constructed in this process. AMC is constructed by using all the queries posed by the users and the selected images as the keywords. The purpose of the AMC is to model keyword relevance. Markov kernel is used to cluster the keywords. Kernel of the Markov chain is composed of query keywords.

Construction of Aggregate Markov Chain for visual concepts extraction

The user submits the image query, then the system retrieves the images based on the visual concepts of images. Feature extraction is used to extract visual concepts from the images. The visual features like Color histogram, Color moment, Gabor filters and wavelet transforms are used to represent the visual concepts of the image. So the image is represented by n features. The feature vector of image is given by equation (2)

$$I = (\vec{F}_1, \vec{F}_2, \dots, \vec{F}_n) \quad (2)$$

The concept vector of j^{th} feature is represented in equation (3).

$$\vec{F}_j = \{f_1^j, f_2^j, \dots, f_d^j\} \quad (3)$$

To extract the various visual concepts involved in constructing the Aggregate Markov Chain, we have employed the following visual features.

Color histogram

The main method of representing color information of images in CBIR systems is through color histograms [8]. A color histogram is a type of bar graph, where each bar represents a particular color of the color space being used. Statistically, a color histogram is a way to approximate the joint probability of the values of the three color channels. The most common form of the histogram is obtained by splitting the range of the data into equally sized bins. Then for each bin, the number the colors of the pixels in an image that fall into each bin are counted and normalized to total points, which gives us the probability of a pixel falling into that bin. One of the main drawbacks of the color histogram is that it does not take into consideration the spatial information of pixels. Thus very different images can be considered similar because they have similar color distributions even though the spatial information of pixels in these images may not be similar. An improvement of the color histogram method includes the cumulated color histogram proposed in [11]. Their results demonstrated the advantages of the proposed approach over the conventional color histogram approach. However the approach has the disadvantage that in the case of multi-dimensional histograms there is no clear way to order bins.

Color moments

Color moments are one of the best color descriptors. Most of the color distribution information is captured by the three low-order moments. We assume that an image has N by M pixels. The first-order moment (μ) calculates the mean color, the second-order moment (σ) calculates the standard deviation, and the third-order moment calculates (θ) the skewness of color. These three moments are extracted using the following mathematical formulation.

$$\mu = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M f_{ij}$$

$$\sigma = \left(\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M (f_{ij} - \mu_{ij})^2 \right)^{\frac{1}{2}}$$

$$\theta = \left(\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M (f_{ij} - \mu_{ij})^3 \right)^{\frac{1}{3}}$$

where f_{ij} is the value of pixel in the i^{th} row and j^{th} column of the image.

Texture feature

Texture is an important property in image retrieval and is a regional descriptor in the retrieval process. The texture descriptor provides measures, such as smoothness, coarseness and regularity [13, 17, 18]. Texture description algorithms are divided into different categories, such as structural and statistical. Statistical methods include Fourier power spectra, co-occurrence matrices, Tamura features and word decomposition, Markov random field, fractal model, and filter-based techniques, such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity [16, 20, 21, 22]

Gabor filters

Gabor filters consists of a group of wavelets each of which capturing energy at a specific resolution and orientation. Therefore, Gabor filters are able to capture the local energy of the entire signal or image. The Gabor filter has been widely used to extract image features, especially texture features [18]. Daugman discovered that Gabor filters provide optimal Heisenberg joint resolution in visual space and spatial frequency. For this reason, Gabor filters have been successfully employed in many applications including image coding, texture segmentation, retina identification, document analysis, target detection, fractal dimension measurement, line characterization, edge detection, image representation, and others.

Wavelet Transform

Another multi-resolution approach called wavelet transform has been most widely used in many aspects of image processing. A wide range of wavelet-based tools and

ideas have been proposed and studied for noise removal from images, image compression, image reconstruction, and image retrieval. The multi-resolution wavelet transform has been employed to retrieve images in [19]. The wavelet features do not achieve high level of retrieval accuracy. Therefore, various methods have been developed to achieve higher level of retrieval accuracy using wavelet transform. Wavelet features are computed from discrete wavelet coefficients.

Hierarchical clustering:

After feature extraction, the images are grouped into a set of clusters based on visual concepts. A cluster can be regarded as a representative and discriminative feature hidden in the training images. Clustering can be achieved by using canopy algorithm [11]. The clusters are formed based on the visual distance between the images in the database. A cluster contains a set of images and an image is annotated by a set of caption terms. The canopy is an unsupervised pre-clustering algorithm, often used as pre-processing step for the K-means algorithm or the Hierarchical clustering algorithm. The canopy algorithm is mainly used to speed up clustering operations on large data sets, where using another algorithm directly may be impractical due to the size of the data set. For each visual concept, the images are grouped into a set of clusters. The images are grouped into set of clusters $\{c_1, c_2, c_3, \dots, c_n\}$ for each feature, and each image is projected as a set of caption terms.

Optimization of Aggregate Markov chain

For text query or image query the AMC (Aggregate Markov Chain) is used to cluster the caption term space or visual concept space and define explicit relevance links between the caption terms or visual terms by means of clustering. This clustering task is linked to the convergence characteristics of the AMC chain by evaluating the series

$$CP_D(m) = \sum_{x=0}^m P_D^x$$

where P_D is the AMC kernel. A suitable termination condition stops the series at the desired m where the slow convergence has taken over, but not before the rapid convergence has finished [1]. The value of the determinant is used as a termination condition since the clusters in the rows will drop its rank and the determinant will become close to zero. For image query the optimization is based on the canopy based k-means clustering. This is used to find the termination of the Markov chain.

Finding MSI Distance:

Let x and y be two images represented by their respective steady state probability row vectors π_x, π_y respectively. Let CP^T be the covariance matrix of the zero-mean transpose expected fractional occupancies matrix of the Aggregate Markov Chain (AMC) calculated at the desired m . Then the Markovian Semantic Indexing (MSI) distance between images x and y is defined as,

$$\begin{aligned} d(x, y) &= (\pi_x(x) - \pi_x(y)) \Sigma CP^T (\pi_x - \pi_y)^T \\ &= \delta_{xy} \Sigma ([CP]^T \delta_{xy})^T \end{aligned}$$

(Equations and expressions are not clearly visible.)

Experimental Results

In this section, we present the results of our experiment. Hybrid image retrieval system is proposed to improve the image retrieval process. The proposed system is compared with both caption term based system and visual concept based retrieval system. The evaluation of the proposed system is based on the precision and recall and also on the average precision. The experiment has been performed on the ground truth images in [12]. There are 23 classes in that database. Arbogreen, Australia, Barcelona, barcelona2, Cambridge, Cambusinfall, Cannonbeach, Cherries, Clolumbia george, Football, Geneva,Greenlake, Greenland, Icpr2004. Imageset, Indonesia, Iran, Italy ,Japan, Leaflestrees, Sanjuans, Springflowers ,Swissmountains, Yellowstone are the 23 classes in the database. In that for experimental purpose we have taken the Geneva and Cambusinfall classes. There are 75 images in these two classes. The “building with sky” is the example query used for evaluating our proposed approach. For experiment purpose we take only top 10 images in the retrieval results.

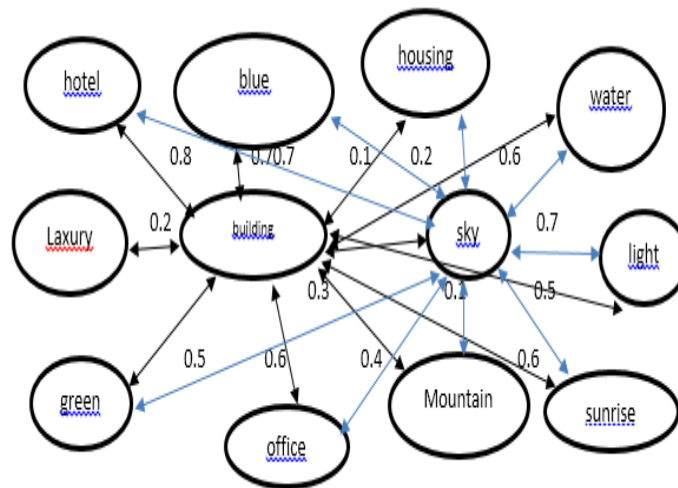


Figure 2: The transition diagram for Experiment

The distance between the top 10 retrieved images is calculated and also that images are ranked for all the methods. We conducted retrieval experiment among the students in our college. We record the user query, click-through data and URL. The Markovian chains are constructed according to the click-through data of users from the retrieval results for different queries. Then the aggregate Markov chain is constructed from the constructed Markov chains. The aggregate Markovian process of the experiment is shown in fig.2. This shows the cluster for the concept with transition probability for the example query. In this transition diagram we omit the relationship that fall below 0.1 transition probability based on [1]. Table 1 shows the distance between the top 10 images. The first column shows the initial list of images and the











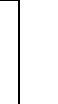
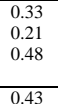
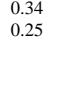
second column shows the ranking of images based on visual concepts. The third and fourth column shows the ranking of images based on caption terms and hybrid method respectively. The application of the MSI distance for the hybrid system returns the best retrieval results based on Markovian network. Images 1, 2 and 3 are closely related to the query terms. It shows nearly correct order for images based on the benchmark list. The images have the building and the sky since the blue key terms are also ranked with high priority. The proposed system is able to rank the images using either the visual concept or caption terms alone. However the hybrid system ranks the images including both visual and caption terms in a better way.

Table 2 shows the distance table between the images. In this table we can observe the better results of the hybrid system when compared to the systems that use either visual concept or caption term alone. In the visual concept based method, the relationship between the visual concepts and the image is represented in the form of a frequency matrix. The visual concept method needs deeper qualitative inference to represent relationship between the visual concepts. But in the caption term based method, the relationship between the caption terms gives better and easily inferable results. The proposed system includes both the visual concept/query and image/visual concepts covariance matrix to improve the retrieved results in terms of the precision and recall.

Table 1: The Rank of 10 Images Based on Distance

IMG	MSI for VC	MSI for CP	MSI for HYBRID
2	8	1	3
3	11	9	2
4	1	6	5
5	7	4	7
6	9	2	6
7	3	8	4
8	2	11	8
9	5	10	10
10	6	5	9
11	10	7	11

Table 2: The distance value for visual concept, caption terms and hybrid approach

										
	0.33 0.21 0.48	0.12 0.15 0.23	0.43 0.46 0.43	0.27 0.34 0.32	0.12 0.32 0.23	0.38 0.32 0.32	0.14 0.23 0.23	0.56 0.45 0.42	0.54 0.65 0.45	0.19 0.23 0.34
	0.43 0.34 0.25	0.14 0.17 0.32	0.33 0.21 0.48	0.31 0.21 0.48	0.23 0.21 0.48	0.34 0.21 0.48	0.32 0.43 0.47	0.33 0.21 0.48	0.44 0.21 0.56	0.21 0.21 0.28









	0.33 0.21 0.48	0.12 0.15 0.23	0.43 0.46 0.43	0.27 0.34 0.32	0.12 0.32 0.23	0.38 0.32 0.32	0.14 0.23 0.23	0.56 0.45 0.42	0.54 0.65 0.45	0.19 0.23 0.34
	0.33 0.23 0.45	0.14 0.23 0.32	0.32 0.40 0.34	0.23 0.43 0.45	0.22 0.34 0.34	0.23 0.32 0.32	0.14 0.23 0.23	0.56 0.45 0.42	0.54 0.65 0.45	0.19 0.23 0.34
	0.36 0.24 0.40	0.12 0.15 0.23	0.43 0.46 0.43	0.27 0.34 0.32	0.12 0.32 0.23	0.62 0.32 0.73	0.14 0.23 0.23	0.56 0.45 0.42	0.54 0.65 0.45	0.19 0.23 0.34
	0.47 0.56 0.23	0.17 0.15 0.33	0.36 0.54 0.43	0.27 0.31 0.45	0.39 0.32 0.28	0.19 0.30 0.98	0.54 0.66 0.62	0.59 0.45 0.42	0.54 0.65 0.45	0.18 0.23 0.22
	0.33 0.21 0.48	0.12 0.15 0.23	0.43 0.46 0.43	0.27 0.34 0.32	0.12 0.32 0.23	0.38 0.32 0.32	0.14 0.23 0.23	0.56 0.45 0.42	0.22 0.64 0.32	0.17 0.22 0.34
	0.33 0.21 0.48	0.12 0.15 0.23	0.43 0.46 0.43	0.27 0.34 0.32	0.12 0.32 0.23	0.38 0.32 0.32	0.17 0.22 0.23	0.38 0.43 0.47	0.32 0.11 1.23	0.19 0.23 0.34
	0.30 0.21 0.48	0.16 0.20 0.23	0.32 0.42 0.36	0.32 0.34 0.23	0.12 0.26 0.42	0.38 0.23 0.32	0.14 0.32 0.42	0.56 0.32 0.40	0.52 0.65 0.45	1.19 1.23 1.34
	0.34 0.36 0.23	0.11 0.14 0.23	0.43 0.46 0.43	0.22 0.34 0.32	0.12 0.38 0.23	0.38 0.40 0.32	0.22 0.30 0.32	0.56 0.45 0.32	0.54 0.63 0.45	0.19 1.09 1.34

Table 3 shows the precision Vs recall values for all the three methods. From table 3 we observe that the caption based method provides 5% improved results when compared to the visual concept based retrieved results. We also observe that the proposed hybrid approach provides 12% improvement against the caption based approach and 17% improvement against the visual concept based system.

Table 3: The precision vs recall value

System	Ground Truth Images	Precision Vs Recall
Markov chain for caption terms	90%	0.83
Markov chain for visual concepts	90%	0.78
Hybrid image retrieval system using Markov chain	90%	0.95

Conclusion

Hybrid image retrieval system improves the image retrieval process by Markov chain approach. In this approach, Markov chain is constructed for visual concepts as well as caption terms. By combining both features of the image, it improves the image retrieval process. Image retrieval based on CBIR alone fails to meet the user's needs due to the semantic gap. So in hybrid system the Markov chain is constructed for visual concept as well as caption terms. Visual concepts are based on the colour, texture and edge of the image. Caption terms are based on image keywords. This image retrieval is enhanced and effective since the hybrid system reflects the user preferences and satisfies user expectations in a better way. It achieves better precision Vs recall compared to existing systems.

References

1. Konstantinos A. Raftopoulos, Member, IEEE, Klimis S. Ntalianis, Dionyssios D. Sourlas, and Stefanos D. Kollias, Member, "Mining User Queries with Markov Chains:Application to Online Image Retrieval"
2. Belongie, S., Carson, C., Greenspan, H. and Malik, J. (1998). Color And Texture-Based Image Segmentation Using Em and Its Application To Content-Based Image Retrieval, In Processing of 6th International Conference on Computer Vision.
3. Flickner, M., Sawhney, H., Niblack, W. and Yanker, p. (1995). Query By Image And Video Content: The Qbic System. Computer. 23–32.
4. Smith, J. R. and Chang, S. (1996). Visual seek: A Fully Automated Contentbased Image Query System. In Proceedings of 4th Acm International Conference on Multimedia. 211–218. s
5. Virginia, E. and Stonebraker, M. (2005). Chabot: Retrieval from a relational database of images. IEEE Computer. 28(9), 40-48.
6. Pentland, A., Picard, R. W. and Sclaroff, S. (1994). Photobook: Contentbased Manipulation of Image Databases, In Spie Storage And Retrieval Image and Video Databases. 2185, 34-47.
7. Demerdash, O., Kosseim, L. and Bergler, S (2008). CLaC at ImageCLEFphoto 2008, ImageCLEF Working Notes.
8. Andrysiak, T. and Chora'S, M. (2005). Image Retrieval Based on Heirarchical Gabor Filter. International Journal on Applied Mathematics and Computer Science. 15, 471–480.
9. Choras, R. (2007). Image Feature Extraction Techniques And Their Application For Cbir and Biometrics System, International Journal of Biology And Biomedical Engineering, 1, 6-16.
10. Akoushideh, A. and Shahbahrami, A. (2010). Accelerating Texture Features Extraction Algorithms using FPGA Architecture. International Conference on ReConFigurable Computing and FPGAs. 232-237.
11. Swain, M. J. and Ballard, D. H. 1991. Color indexing. International Journal of Computer Vision. 7(1), 11–32.

12. Rui, Y., Huang, T. S., and Chang, S.-F. 1999. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*. 10(1), 39–62.
13. Long, F., Zhang, H. J. and Feng, D. D. 2003. *Fundamentals of Content-based Image Retrieval*. Multimedia Information Retrieval and Management. D. Feng Eds, Springer.
14. S.Sabena, P.Yogesh, L.SaiRamesh, “Image Retrieval Using Canopy and Improved K-Mean Clustering” *International Journal on Computer Applications*, Sep 2011, PP. 15-19.
15. L.G.Shapiro, “Ground Truth Database”, <http://www.cs.washington.edu/research/imagedatabase/groundtruth/> Univ. of Washington, 2012
16. T.-T. Pham, N.E. Maillot, J.-H. Lim, and J.-P. Chevallet, “Latent Semantic Fusion Model for Image retrieval and Annotation,” *Proc. 16th ACM Conf. Information and Knowledge Management(CIKM)*, 2007 [14] K. Barnard and D. Forsyth, “Learning the Semantics of Words and Pictures,” *Proc. Int’l Conf. Computer Vision*, vol.2, pp408-415, 2001 [15] T. Hofmann, “Unsupervised Learning by Probabilistic Latent Semantic Analysis,” *Machine Learning*, vol. 42, no. 1/2, pp. 177-196, 2001
17. High-Dimensional Visual Vocabularies for Image Retrieval João Magalhães^{1,2}, Stefan Rügger^{1,2}. *SIGIR 2007*
18. Unifying Keywords and Visual Contents in Image Retrieval Xiang Sean Zhou and Thomas S. Huang University of Illinois at Urbana Champaign June 2002.
19. Improving Web Image Search by Bag-Based Reranking Lixin Duan, Wen Li, Ivor Wai-Hung Tsang, and Dong Xu, *Member, IEEE Transactions On Image Processing*, Vol. 20, No. 11, November 2011
20. Wei-Hao Lin, Rong Jin, Alexander Hauptmann, “Web Image Retrieval Re-Ranking with Relevance Model”, *IEEE/WIC International Conference On Web Intelligence(WIC’08)*, Halifax, Canada, 2003
21. Mingmin Chi, Peiwu Zhang , "Web Image Retrieval ReRanking with Multi-view Clustering" *WWW 2009*, April 20–24, 2009, Madrid, Spain