Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.482 – 490

REVIEW ARTICLE

Efficient Image Retrieval Using Different Content Based Image Retrieval Methods: A Review

Priyanka Srivastava[1], Dr. K.S.Patnaik[2]

[1]Lecturer, Computer Science Department, Usha Martin Academy, India

[2]Associate Professor, Computer Science and Engineering Department, Birla Institute of Technology, Mesra, India

[1] Priyanka12_1987@rediffmail.com; [2] ktosri@gmail.com

ABSTRACT- In this paper, we reviewed different methods used in Content-based image retrieval, to resolve the problem of efficient and similar digital image retrieval from large databases with high precision. Uses of different features like texture, color, shape have been focused in different ways to implement better and faster retrieval of data through different CBIR methods. Methods like Statistical Fractal-scaled Product Metric (SFPM), Fractal-scaled Product Metric (FPM), local tetra patterns (LTrPs) have been focused to maximize the accuracy of CBIR systems and to enhance similarity queries.

Keywords- Content-based image retrieval (CBIR), local binary pattern (LBP), local derivative pattern (LDP), local tetra pattern (LTrP), Statistical Fractal-scaled Product Metric (SFPM), Fractal-scaled Product Metric (FPM), Proto Object (PO), Relevance Feedback (RF)

I. INTRODUCTION

A. Motivation

The tremendous growth of digitization due to Web cameras, digital cameras, mobile phones and other electronic devices is making the image database management extremely tedious and clumsy task.

There exists an essential need for developing an efficient expert technique that can automatically search, sort and retrieve the desired image from the huge database.

Content-based image retrieval (CBIR) is one of the common methods for such applications. The fundamental step of CBIR is its feature extraction whose effectiveness depends upon the method adopted for extracting features from given images. The features in CBIR includes color, texture, shape, faces, spatial layout, etc., to represent and index the image database, classified as general features or primitive features such as color, texture, shape and domain-specific features or logical features such as human faces, fingerprints, etc.

Earlier, introduction of picture archiving and communication system (PACS) [4] has contributed to the effective use of digital images. But image retrieval in PACS is not fast and steady as expected by the user. Most solutions in PACS are based on metadata as image height, width, type of compression used to store it. Therefore, CBIR techniques have been exclusively researched in the last few years to retrieve a specific image as result by inputting query images irrespective of text based search and it manages digital picture archives based on their visual characteristics.

Now-a-days, different techniques of CBIR came in existence to remove the problem of partial occlusion (accurate object are not retrieved as it is hidden by other object), non rigid motion, over segmentation [2]. A non-rigid motion is a motion that is flexible. The result of various objects having non-rigid motion does not resemble the original object in some respect.

B. Background

CBIR, image retrieval is mainly performed for analyzing similarity among different images using distance function which compares the feature vectors to quantify how close (similar) two vectors are, focusing the queries. Feature vector are numerical value of different features of an image. Major issues in CBIR systems includes how to define a feature vector(signature) which denotes image content and how to determine the similarity between a pair of images based on their signatures.

The visual properties of an image such as color, texture, shape is being extracted by image segmentation algorithms [4] also called feature extractors or image descriptors, either globally or locally (region of interest). It computes a number of numeric values, representing features or signatures of the images. The signature is then processed through a function to measure the similarity between signatures of different images. Examples can include histogram for color, Haralick descriptors for texture, and Zernike moments for shape [4]. Zernike polynomials state a precise mathematical model to capture the global shape of images while preserving enough information. Zernike moments of an image are the projections of the image pixel onto a base function which does not changes with respect to image size, rotation, translation.

However, the use of large number of extractors leads to “dimensionality curse”, which reduce the query accuracy and retrieval speed. Therefore, suitable methods which combine feature extractors without degrading the results are needed.

A suitable approach is to combine multiple descriptors by estimating the similarity for each of them in a base level and then aggregates the partial similarities using a composition function.

Another major issue is the gap between the low-level visual features and the high level semantic features within images. High level semantic feature retrieval requires the system to measure the similarity in a way human being would perceive or recognize [8]. It includes integrating low level features with other media, like audio, text, etc. which are more natural to human being, and therefore at a higher level. For a given query, the Relevance feedback (RF) is the most powerful tools to narrow the semantic gap problem [7] and thus improves the performance of a CBIR system. CBIR system first retrieves a list of different images according to a predefined similarity metrics, often defined by the distance between query vector and feature vectors of images in a database [15]. It focuses on the correlation between a user and a search engine by analyzing similar or dissimilar images with the query image, which can be positive and negative feedbacks by the user respectively [5], and the system will refine the query with new list of images but it is having several shortcomings. To address the challenges encountered by traditional RF approaches many studies have been attempted by resorting to the user information provided by a number of users. Therefore, besides the low-level visual features of images, each pair can also be related with a set of similar or dissimilar pairwise constraints decided by users. This new approach of utilizing user feedback log data [3] which can be achieved when the system can accumulate RF information provided by a number of users with similar and dissimilar pairwise constraints, for image retrieval is termed as collaborative image retrieval (CIR) whose basic purpose is to enhance the performance of CBIR system.

Different approaches of CBIR mainly works within three categories in a unified framework, i.e., unsupervised learning, supervised learning with explicit class labels, weakly supervised learning with pairwise constraints (or side information). Unsupervised learning method does not use any class label information and normally prohibit the internal distribution of data. Supervised learning approaches can effectively explore certain collections of data with explicit class labels. Almost all weakly supervised learning approaches [3] can only learn distance metric from the training data that are viewed in the forms of pairwise constraints (or side information), in which each pairwise constraint determines whether the corresponding two samples are similar or dissimilar for a defined task.

II. DIFFERENT METHODS USED FOR CBIR

A. Fractal-scaled Product Metric (FPM method)

Every feature extractors generate signatures compared by metrics which in turn generate metric space. n metric spaces can be aggregated as, $M_i = (S_i, \delta_i)$, $1 \leq i \leq n$, defining a metric over the Cartesian product of $M_1 \times M_2 \times \ldots \times M_n$, called product metric. The FPM method calculates the scale factors (or weights w_i) as given in equation 1 between the composed metrics based on correlation fractal dimension of each original metric space prepared by the extracted descriptors. By knowing correlation fractal dimension of a dataset, prediction of its properties which are similar to that of a dimensional dataset is done with approximately the same embedded dimension**.**

The composition distance function Δ between two images x, y is the aggregation of individual distances $\delta_i(x_i, y_i)$ as

$\Delta(x,y) = \sum_{i=1}^{r} w_i \,.\, \delta_i(x_i, y_i)$ equation (1)

Where w_i is the weight associated with respective descriptor.

Main focus of FPM [4] is to identify the role for the overall similarity calculation of each descriptor. The value of the intrinsic dimension focus the presence of correlations among the attributes of a dataset, the fractal dimension gives an estimate of a lower bound for the number of features needed in a similarity search to record the essentiality of data information. The FPM method performs normalization to shrink the effects of the range difference in the similarity computation by using the largest known distance $dmax_i$ for each descriptor.

FPM metric can be given as:

$$\Delta(x,y) = \sum_i^n D_{2i} . \frac{\delta(x_i, y_i)}{dmax_i} \quad \text{equation(2)}$$

Here, D_2 is correlation fractal dimension, $dmax_i$ is the largest known distance between all pairs of elements considering every descriptor, i denotes the descriptor but this approach is quadratic on the number of elements.

B. Statistical Fractal-scaled Product Metric (SFPM)

The Statistical Fractal-scaled Product Metric technique [4] maximizes the query accuracy with minimum number of features used to perform the query. This method combines statistical approach (association rule mining) to perform feature selection and fractal theory approach (FPM) to find the most suitable weight of features which speed up data searching.

first step involves the submission of images from the database to a set of feature extractors where image being represented by its feature vectors which then proceed to feature selection process through the StARMiner algorithm [4], inducing reduced set of features for each extractor. StARMiner, termed as Statistical Association Rule Miner. It produce more semantically significant patterns. Spatio-Temporal Association Rules (STARs) describes how objects move between regions over time [9]. The reduced feature vector's properties are analyzed using the fractal theory (FPM method), producing scale factors for multi-descriptors. Lastly, it is submitted to CBIR system for execution of the similarity queries asked by the user, thus improves the correctness of the result and the speed of the query execution [4].

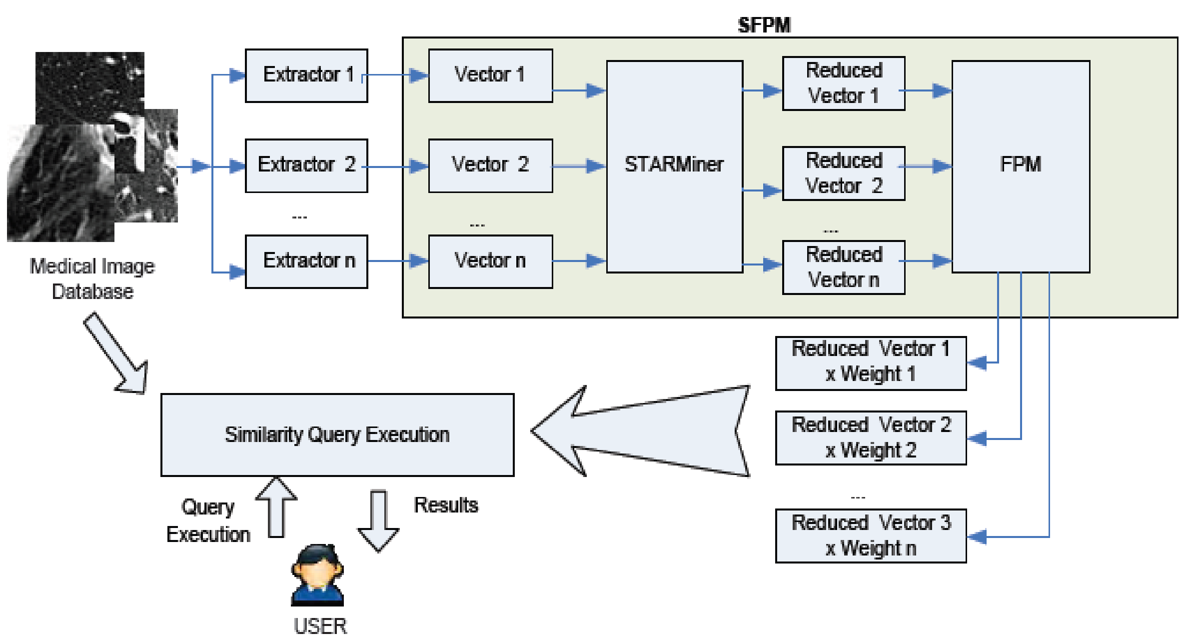

Fig. 1 SFPM method illustration

SFPM starts its process by StARMiner, which does a significant reduction in the number of features of the feature vectors. StARMiner algorithm is used to determine statistical association rules from a dataset (image) to select the most relevant features by removing noise and redundant features, making a new and enhanced sketch of the images with query precision.

Let T be the collection of data of medical images, x an image class, $T_x \in T$ where T_x is the subset of images of class x and F is a feature. Let $\mu_F(Z)$ and $\sigma_F(Z)$ be the mean and standard deviation of the values of feature F respectively, of subset of images Z. StarMiner uses three edges, $\Delta\mu_{min}$ as the minimum allowed difference between the average of the feature F in images from class x and in the remaining dataset, γ_{min} as the minimum confidence to reject the hypothesis H_0 and $\Delta\sigma_{max}$ as the maximum standard deviation of F allowed in a given class. StARMiner algorithm follows rules of the form x → F, if the following equations satisfied [4].

$$\mu_f(T_x) - \mu_f(T - T_x) \geq \Delta\mu_{min}$$

$$\sigma_f(T_x) \leq \Delta\sigma_{max}$$

$$H0 : \mu_f(T_x) = \mu_f(T - T_x)$$

Here, x → F relates a feature F with a class x, where values of F have statistically distinct behavior in images of class x. It indicates that F is a prominent feature to distinguish images of class x from the remaining image as it is having particular and uniform behavior in images of a given class.

Let T be the collection of data of medical images, x an image class, $T_x \in T$ where T_x is the subset of images of class x and F is a feature. Let $\mu_F(Z)$ and $\sigma_F(Z)$ be the mean and standard deviation of the values of feature F respectively, of subset of images Z. StarMiner uses three edges, $\Delta\mu_{min}$ as the minimum allowed difference between the average of the feature F in images from class x and in the remaining dataset, γ_{min} as the minimum confidence to reject the hypothesis H_0 and $\Delta\sigma_{max}$ as the maximum standard deviation of F allowed in a given class. StARMiner algorithm follows rules of the form x → F, if the following equations satisfied [4].

$$\mu_f(T_x) - \mu_f(T - T_x) \geq \Delta\mu_{min}$$

$$\sigma_f(T_x) \leq \Delta\sigma_{max}$$

$$H0 : \mu_f(T_x) = \mu_f(T - T_x)$$

Here, x → F relates a feature F with a class x, where values of F have statistically distinct behavior in images of class x. It indicates that F is a prominent feature to distinguish images of class x from the remaining image as it is having particular and uniform behavior in images of a given class.

Let T be the collection of data of medical images, x an image class, $T_x \in T$ where T_x is the subset of images of class x and F is a feature. Let $\mu_F(Z)$ and $\sigma_F(Z)$ be the mean and standard deviation of the values of feature F respectively, of subset of images Z. StarMiner uses three edges, $\Delta\mu_{min}$ as the minimum allowed difference between the average of the feature F in images from class x and in the remaining dataset, γ_{min} as the minimum confidence to reject the hypothesis H_0 and $\Delta\sigma_{max}$ as the maximum standard deviation of F allowed in a given class. StARMiner algorithm follows rules of the form x → F, if the following equations satisfied [4].

Let T be the collection of data of medical images, x an image class, Tx T where Tx is the subset of images of class x and F is a feature. Let μF (Z) and σF (Z) be the mean and standard deviation of the values of feature F respectively, of subset of images Z. StarMiner uses three edges, Δμmin as the minimum allowed difference between the average of the feature F in images from class x and in the remaining dataset, γmin as the minimum confidence to reject the hypothesis H0 and Δσmax as the maximum standard deviation of F allowed in a given class. StARMiner algorithm follows rules of the form x → F, if the following equations satisfied [4].

Let T be the collection of data of medical images, x an image class, $T_x \in T$ where T_x is the subset of images of class x and F is a feature. Let $\mu_F(Z)$ and $\sigma_F(Z)$ be the mean and standard deviation of the values of feature F respectively, of subset of images Z. StarMiner uses three edges, $\Delta\mu_{min}$ as the minimum allowed difference between the average of the feature F in images from class x and in the remaining dataset, γ_{min} as the minimum confidence to reject the hypothesis H_0 and $\Delta\sigma_{max}$ as the maximum standard deviation of F allowed in a given class. StARMiner algorithm follows rules of the form x → F, if the following equations satisfied [4].

$$\mu_f(T_x) - \mu_f(T - T_x) \geq \Delta\mu_{min}$$

$$\sigma_f(T_x) \leq \Delta\sigma_{max}$$

$$H0 : \mu_f(T_x) = \mu_f(T - T_x)$$

Here, x → F relates a feature F with a class x, where values of F have statistically distinct behavior in images of class x. It indicates that F is a prominent feature to distinguish images of class x from the remaining image as it is having particular and uniform behavior in images of a given class.

C. Local Tetra Pattern

For texture classification and its retrieval, Local Tetra Pattern (LTrP) [1] emerged as a silver line. The LTrP describes the spatial distribution of the local texture by the use of direction of the center gray pixel g_c. The Local Binary Pattern(LBP), Local Derivative Pattern(LDP), and Local Ternary Pattern(LTP) extract information based on the distribution of edges, which is having only two directions i.e. positive direction or negative direction. But local tetra patterns (LTrPs) focus on four direction code i.e. it differentiates the edges in more than two directions [11]. It calculates the directionality of all pixels by the use of $0^{\cdot}$ and $90^{\cdot}$ derivatives of LDPs. LBP and LTP determines the relationship between the referenced pixel and its surrounding neighbors by computing gray-level difference. But LTrP encodes the relationship between the referenced pixel and its neighbor pixel based on the directions being calculated using the first-order derivatives in vertical and horizontal directions. It present a statistical view of the texture retrieval problem by combining the two related tasks, named as feature extraction (FE) and similarity measurement (SM) [11].

Suppose an image as I, the first-order derivatives along $0^{\cdot}$ and $90^{\cdot}$ directions are denoted as $I^1_e(g_p)$ where $e=0^{\cdot},90^{\cdot}$. Let g_c be the center pixel of image I, and g_h and g_v denote the horizontal and vertical neighborhoods of g_c, respectively**.** The first-order derivatives at the center pixel g_c can be calculated as:

$$^1{}_{0^{\cdot}}(g_c) = I(g_h) - I(g_c)$$

$$^1{}_{90^{\cdot}}(g_c) = I(g_h) - I(g_c)$$

LBP is an efficient texture operator which labels the pixels of an image by using the neighborhood of each pixel and considers the result as a binary number. To capture texture information of the face appearance, LBP descriptor is incorporated into multiscale heat-kernel.

Provided a center pixel in the image, the LBP value is computed by analyzing its gray value with its neighbors, based on

$$LBP_{P,R} = \sum_{P=1}^{P} 2^{(P-1)} \times f_1(g_p - g_c)$$

$$f_1(x) = \begin{cases} 1, & x > 0 \\ 0, & else \end{cases}$$

Where g_c is the gray value of the center pixel, g_p is the gray value of its center neighbors, P is the total number of neighbors, and R is the radius of the neighborhood.

Zhang *et al.* introduced the feasibility and effectiveness of LDPs for face recognition using high order local patterns. It considers the LBP as the nondirectional first-order local pattern operator and extended it to (n-1)th order derivative. The LDP contains more detailed discriminative features in comparison of LBP**.**

LDP involves two different phases. In the first phase, image of the face is decomposed into different orientation and scale using multiscale and multiorientation Gabor filters. In the second phase, LBP analysis is used for neighboring relationship description not only in image space but also in different scale and orientation responses [1].

D. Content-Based Scalability with JPEG 2000

Image being characterized by mathematical curves and straight lines and which is being used at all resolutions and at all sizes is termed as scalable image. A prominent example which provides scalability with respect to quality, resolution and color component in the transfer of images is JPEG2000 (an image compression standard and coding system). An image is composed of several object regions. Initially, identification and description of semantic object region is done, and then addition of information to what is being coded for meaningful interpretation to both humans and computers. The two main areas are, object- based coding and region-based coding.

However scalability with respect to semantic content is done by region based bit allocation mechanism within the JPEG2000 using the concept of proto objects which also bridges the semantic gap between object-based and region-based approaches. Proto Object is defined as a volatile unit of visual information which can be bounded into a coherent and stable object when accessed by focused attention.

Initially an input image is partitioned into different proto-objects (a region which contains semantically meaningful physical object) and background regions (region that contains no object of interest) by simulating visual focus of attention on salient proto-objects [6] and then separately encodes these regions using JPEG2000 coding systems.

An object-based coding of generic scene content requires segmentation methods which can be defined as the classification or clustering an image into several parts or regions according to the features of image [10], to determine pixels with markup tags specifying what each set of pixels represents. Another option is lower level process of region-based coding, which focus on image regions as semantic unit defined by their statistical properties. Statistically homogeneous regions can be taken out by pixel-level segmentation techniques and the shape of a region needs to be coded using shape coding algorithms. To transmit the background separately from the foreground, there is the requirement of refined motion analysis and prediction strategies which is not an efficient method.

The proposed JPEG2000 system to support the content-based scalability, as shown in Fig. 3, we first fragment an input image into PO regions and BG regions, and then amend both the construction of an operational RD curve in the coding pipeline and the implementation of an efficient rate control scheme in terms of PO regions. Two major modifications being made in standard JPEG2000 system are as follows: 1) Use of PO region segmentation instead of tile partition. 2) Analyzing he quality layer in terms of PO regions.

This segmentation also provides accurate rate control scheme including scalability with respect to content. Walther et al. gave the origin of PO region segmentation. By extracting particular image region around the focus of attention, a PO can be obtained. Then updated saliency map is obtained by setting the saliency value of the locations belonging to the PO region to zero**.** Second PO region is obtained by performing the same PO region finding process on the updated saliency map. The above process repeats until the maximum saliency value of the updated saliency map becomes less than the predefined threshold. Finally, the pixels not assigned to any PO regions in the saliency map belong to BG regions.

III. CONCLUSIONS

From the above mentioned review, it has been resulted that SFPM have decreased the feature vector size up to 65% and improved its query precision up to 27% as compared to FPM approach. Thus it is well-suited to improve the quality of content based query in CBIR systems. JPEG 2000 rate control approach efficiently reduces the computational complexity and memory usage and produce high quality image. Performance analysis determines that LTrP improves the retrieval result from 70.34% to 75.9% in terms of average precision. But still more focus is required to enhance the CBIR methods which can include semantic features irrespective of low level features.

REFERENCES

[1] S. Murala, R. P. Maheshwari and R. Balasubramanian, “Local Tetra Patterns: A New Feature Descriptor for Content-Based Image Retrieval,” *IEEE Transactions on Image Processing*, vol. 5, pp. 2874-2886, May 2012.

[2] J. Xue, C. Li and N. Zheng, “Proto-Object Based Rate Control for JPEG2000: An Approach to Content-Based Scalability,” *IEEE Transactions on Image Processing*, vol. 20(4), pp. 1177-1184, April 2011.

[3] L. Zhang, W. Lipo and L. Weisi “Conjunctive Patches Subspace Learning With Side Information for Collaborative Image Retrieval,” *IEEE Transactions on Image Processing,* vol. 21(8), pp. 3707-3720, Aug. 2012.

[4] R. Bueno, M. X. Ribeiro, A. J. M. Traina and C. Traina, “Improving Medical Image Retrieval through Multi-Descriptor Similarity Functions and Association Rules”, *IEEE 23rd International Symposium,* pp. 309-314, Oct. 2010.

[5] L. Zhang, L. Wang and W. Lin, “Semisupervised Biased Maximum Margin Analysis for Interactive Image Retrieval,” *IEEE Transactions On Image Processing,* vol. 21(4), pp. 2294-2308, April 2012.

[6] K.S. Patnaik, A. Mustafi, "Perceptual resemblance of facial images: a near set approach", *Indian Journal of Computer Science and Engineering*, vol. 1(3), pp. 152-156, 2010.
[7] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener and C. Roux, "Fast Wavelet-Based Image Characterization for Highly Adaptive Image Retrieval", *IEEE Transactions On Image Processing*, vol. 21(4), pp.1613-1623, April 2012.
[8] C. Zhang and T. Chen, "From Low Level Features To High Level Semantics", *Handbook of Video Databases: Design and Applications*, 2003.
[9] F. Verhein, "Mining Complex Spatio-Temporal Sequence Patterns", *In SDM*, vol. 9, pp. 605-616, 2009.
[10] M. X. Ribeiro, A.G. Balan, J.C. Felipe, A.J. Traina and Jr.C. Traina"Mining Statistical Association Rules to Select the Most Relevant Medical Image Features," In *Mining complex data* , pp. 113-131, 2005.
[11] S. Anitha, "Content Based Image Retrieval Using Local Tetra Pattern," *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development*, vol. 3(1), pp.219-228, Mar 2013.
[12] V. G. Ukinkar, "A New Content Based Image Retrieval Technique Using Local Tetra Patterns: A Review", *International Journal Of Pure And Applied Research In Engineering And Technology*, Vol. 1(8), pp.687-697, 2013.
[13] A. Azeem, M. Sharif, M. Raza and M. Murtaza, "A Survey: Face Recognition Techniques under Partial Occlusion," *Int. Arab J. Inf. Technol.*, vol. 11(1), pp. 1-10, 2014.
[14] H. Zhang, Z. Chen, W.Y. Liu and M. Li, "Relevance Feedback in Content-Based Image Search", *PRIS*, 2001.
[15] R. Bueno, D.S. Kaster, A.A. Paterlini and A.J. Traina, "Unsupervised scaling of multi-descriptor similarity functions for medical image datasets", *22nd IEEE International Symposium*, pp.1-8, Aug. 2009.

