Combining features using principle component analysis and independent component analysis for image retrieval

Mohanned M. Dawoud [a], Amin A. Shoukry [b] and Khaled M. Mahar [a]
[a] *College of Computing and Information Technology, Aast, Alexandria, Egypt*
e-mail: khmahar@aast.edu
[b] *Computer and Systems Engineering Department, Alexandria University, Alexandria, Egypt*
e-mail amin.shoukry@gmail.com

Research work shows that Independent Component Analysis (ICA) and Principle Component Analysis (PCA) are good variants of projection pursuit. In the present paper, a comparison between combined using PCA and ICA and uncombined features for Content Based Image Retrieval (CBIR) is performed. The paper benefits from ICA and PCA and uses a Hierarchical Self-Organizing Map (HSOM) for better images clustering. To assess the performance of the proposed technique, two features provided by MPEG-7 are used, Color Structure (CS), and Edge Histogram (EH). Research work showed that these two features give better performance among other MPEG-7 features such as color and texture features. The proposed technique is applied to two datasets. The results showed that using ICA in the combining process gives better performance than PCA, and PCA gives better performance than uncombined features.

أثبتت الأبحاث أن تحليل المكونات المستقلة شكل آخر جيد من متابعة الإسقاط. فى هذا البحث يتم جمع الخصائص باستخدام تحليل المكونات المستقلة أو تحليل المحاور الرئيسية بحيث يكون الناتج مجموعة من الخصائص الجديدة التى لها نفس عدد وطول الخصائص الأصلية وتستخدم فى عملية استرجاع الصور. هذه الخصائص تستخدم فى المستويات الخاصة بالمخططات الهرمية ذاتية التنظيم، بحيث يستخدم العنصر الرئيسى فى المستوى الأول (العناصر الرئيسية تحتوى على الجزء الأهم من البيانات) والثانى فى المستوى التالى وهكذا. لذلك تكون النتيجة معتمدة على مزيج من كل الخصائص فى كل مستوى من المخطط الهرمى ذاتى التنظيم. توضح التجارب التى أجريناها فى هذا البحث أن ربط الخصائص يطى نتائج أفضل من كل من الخصائص المجمعة - باستخدام تحليل المحاور الرئيسية أو تحليل المكونات المستقلة- والخصائص الغير مجمعة ولكن بتكلفة تنفيذ مضاعفة فى عملية البحث. كذلك توضح النتائج ان جمع الخصائص باستخدام تحليل المكونات المستقلة يعطى نتائج أفضل من جمعها باستخدام تحليل المحاور الرئيسية والذى بدوره يحقق نتائج أفضل من استخدام الخصائص الغير مجمعة.

Keywords: Component Analysis, Content Based Image Retrieval, Hierarchical Self-Organizing Map, Color Structure, Edge Histogram

1. Introduction

Content-Based Image Retrieval (CBIR) is the application of computer vision to the image retrieval problem; that is, the problem of searching for digital images in large databases. Very intensive research efforts have been done in this field due to the rapid increase of number and usage of multimedia files on the internet. Therefore, many CBIR systems came to existence.

The framework depicted in fig. 1 can conceptually describe most image retrieval systems [11]. However, different techniques of querying, relevance feedback, features extraction, matching, indexing data structures, and result presentation have been applied in these systems.

One of the indexing techniques is the use of Hierarchical Self-Organizing Map (HSOM) as proposed in [8]. The HSOM consists of many levels; each level has its features and learning rate. In [5], several parallel SOMs have been used with the user Relevance Feedback (RF) technique to adapt to the user's preferences. In [4], structure, color, and texture have been combined to improve the performance of image retrieval. Amato et al. [1] showed that using Color Structure (CS) gives better results than the other MPEG-7 color features, and that using Edge Histogram (EH) gives better performance than Homogeneous Texture (HT).

In this paper, a technique, that benefits from the previous techniques without increasing the length of the feature vector, is

© Faculty of Engineering Alexandria University, Egypt.

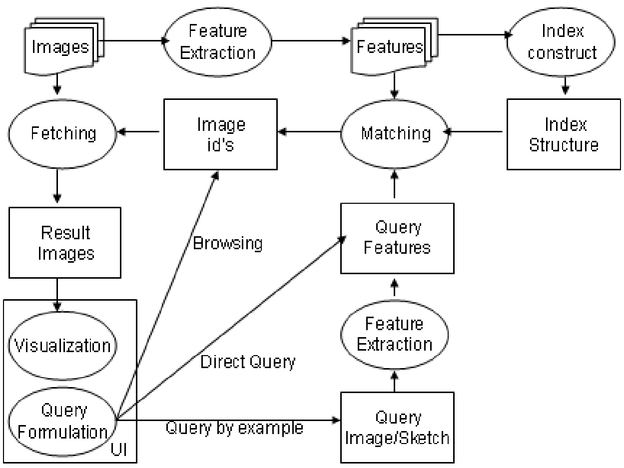

Fig. 1. Content-based image retrieval framework.

proposed. Concatenating normalized features as in [4] gives better performance, but increases the length of the feature vector, which increases the complexity of the search process. So, in this paper, Component Analysis (PCA) [9] and/or Independent Component Analysis (ICA) [3] are used to combine features while keeping the length of the feature vector fixed. A comparison between combined and uncombined features is made using the F-measure as a performance measure [10].

The proposed technique has been tested on two datasets, the first has 10221 images of size 128×128 consisting of six classes, and the second has 1376 images of size 128×128 consisting of four classes. The classification of the images is used to measure the success of an image query.

The remainder of the paper is organized as follows: Section 2 briefly presents the CS extraction process, Section 3 presents the EH extraction process. Section 4 describes the proposed technique, Section 5 presents the results, and finally, Section 6 provides the conclusions.

2. Color structure extraction

CS feature is usually used in CBIR since it takes into consideration the structure of the color beside the color itself. CS is extracted in three steps:

1- The image color space is converted to Hue-Maximum-Minimum-Difference (HMMD) color space [7].

2- The pixels of the image are nonlinearly quantized into one of four color space quantization operating points (256, 128, 64, and 32 bins) [7].

The quantization shown on table 1 is built using the following steps. Firstly, the HMMD color space is divided into five subspaces (0, 1, 2, 3, and 4). The subspaces division is performed on the HMMD values by using the following intervals: [0, 6), [6, 20), [20, 60), [60,110), and [110, 255]. Secondly, each color subspace is uniformly quantized along the hue and sum axes where the number of quantization levels is shown in table 1 (for both hue and sum parameters). Finally, the values that contain the color structure histogram are normalized to the 8-bit code value, so that the range of the histogram is converted from [0, 1] values to [0, 255].

3- A structuring element is used to visit all locations in the image. For each color found inside the structuring element, the corresponding value in the histogram is incremented by one as shown in fig. 2. The increment depends only on the occurrence of the color not on the count of occurrences.

The structuring element can be of size 4×4 or 8×8. It has been determined experimentally that the optimal scale is 8×8 [7]. To reduce the computational overhead, the distance between the structuring points is increased with the image size. This method is equivalent to sub-sampling the image by powers of 2, and then using a structuring element of 8×8 pixels. The sub-sampling factor of an image with width W and height H, respectively, is given by $k=2^p$ [6] where:

Table 1
HMMD color space quantization for CS

No. of cells	Subspace	0	1	2	3	4
256	Hue	1	4	16	16	16
	Sum	32	8	4	4	4
128	Hue	1	4	8	8	8
	Sum	16	4	4	4	4
64	Hue	1	4	4	8	8
	Sum	8	4	4	2	1
32	Hue	1	4		4	4
	Sum	8	4		1	1

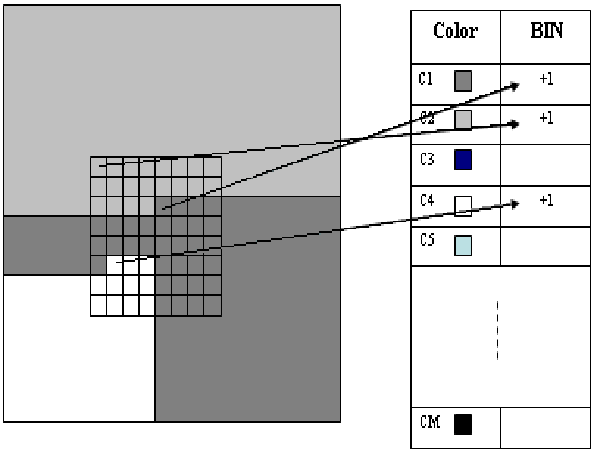

Fig. 2. Accumulation of CS histogram where M is quantized cells number.

$$P = max\{0, round(0.5 * log_2(W*H) - 8)\}. \quad (1)$$

The spatial extent of the structuring element is given by E×E, where E=8×k. In this paper, a quantization of 64 bins is used, which means that the feature length is 64.

3. Edge histogram extraction

According to [1], the use of EH gives better performance than HT. Extraction of the EH from an image is summarized in the following steps for an image size of 128×128:

1. Divide the image into 16×16 non-overlapping blocks. Hence, the number of blocks is 16.
2. For each block, a 4 bins histogram is constructed to count the number of edges; of type horizontal/vertical/ 45 diagonal/ 135 diagonal; in each sub-block of size 2×2 pixels.
3. The final EH will be a 64 bins histogram of the 16 blocks × the 4 bins histogram per block.

The filter coefficients for vertical, horizontal, 45-degree diagonal, and 135-degree diagonal respectively, are shown in fig. 3. Each edge magnitude mv or m_h or m_{d-45} or m_{d-135} for each image sub-block is obtained as shown in eq. (2), below:

$$m_d = \left| \sum_{k=0}^{3} a_k \times f_d(k) \right|, \quad (2)$$

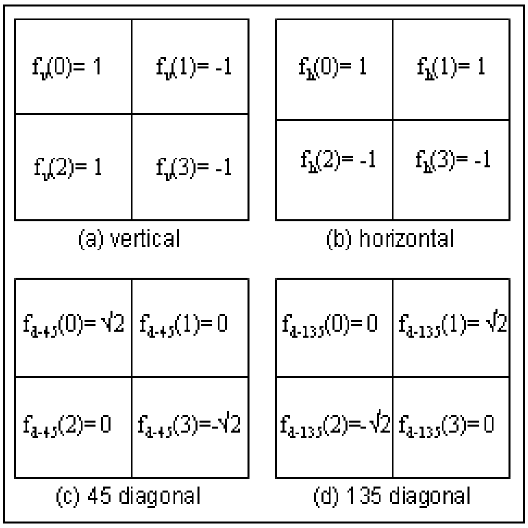

Fig. 3. Filter coefficients for edge detection.

Where'd' refers to a direction and 'a' refers to a gray level.

If the maximum value among the four edge magnitudes is greater than a threshold (T_{edge}), then the image sub-block is considered to have the corresponding edge type. Otherwise, the image block contains no edge.

4. The proposed technique to combine features

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Low-order components often contain the "most important" aspects of the data. PCA is theoretically the optimum linear transformation for a given data in least square sense.

In this paper, features are combined and used in an HSOM for indexing. So, the two stages, "Feature extraction" and "Index construction" shown in fig. 1 are detailed.

4.1. Feature extraction

As shown in fig. 4, for each image in the dataset, n features are extracted. These features may be color, texture, edges or any other feature. Using one of these features directly, (for example color feature) in the first level of the HSOM may result in putting similar images -according to the second feature (edge) - in different clusters. In other words, the clustering depends on the feature used first. Therefore, using PCA or ICA to combine features will generate the same number of features with the same length, but each one of the new features is a combination of all used features. Note that the Low-order components of PCA and ICA often contain the "most important" aspects of the data. So, the first new features (Feature 1 in fig. 4) will be the lowest component; the second new feature (Feature 2 in fig. 4) is the next component, and so on.

4.2. Index construction

The feature extraction process generates n features for each image in the dataset. These features are used to cluster the images using HSOM to speedup the indexing process.

The first feature (the first component of PCA or ICA), which contains the "most important" aspects of the data, is used at the first level of an HSOM. The second feature (next component of PCA or ICA) is used to refine the clusters in a new level of the HSOM, and so on. The resulting HSOM is stored using only the images ids.

To retrieve an image, the process is as follows:

1. Extract the same features (extracted for the dataset images) for the query image.
2. Combine them using PCA or ICA (also according to the same transformation used in the feature extraction process of the dataset images).
3. Classify the image in a cluster of the stored HSOM using the extracted features, each feature being used at the corresponding level.
4. Get the images id's of the cluster, if the number of the id's in that cluster is smaller than the requested number of resultant images, take the nearest neighbor clusters one by one until reaching the requested number.
5. Retrieve images using their ids and present them.

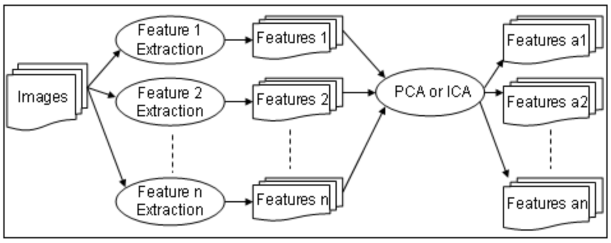

Fig. 4. Feature extraction and combination.

The matching of two features is performed using Euclidean distance. The Euclidean distance (D) between two vectors P = (p_1, $p2$, pn) and Q = (q_1, $q2$, qn) is defined as:

$$D = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad . \qquad (3)$$

5. Experiments and results

For the experiments, quantization with 64 bins is used in the CS extraction process to be of the same length as EH. In addition, a block of 8×8 for the structuring element is used. In the EH extraction, the total number of blocks is set to 4096 (4×4 sub-images, each sub-image is 16×16 blocks). Threshold is set to 11 according to [13].

The proposed technique is applied on two datasets. The first dataset (Dataset1) is used in [4]. It consists of 10221 images of size 128×128 classified in six classes (Man made, birds, bugs, mammals, flowers, and landscapes). The second dataset (Dataset2) is gathered from different sources [2, 4, 12]. It consists of 1376 images. All the images are normalized to the size 128×128. These images are classified into four classes (Nature, Buildings, Planes, and Cars).

The classification results of the images are used to quantify the success of images queries. The F-measure is used as a retrieval performance measure. As the value of the F-measure increases, more accurate retrieval performance is achieved and it is calculated as follows:

where F is the F-measure value, precision (P) and recall (R) are calculated as follows:

$$F(i) = 2(P \times R)/(P + R). \qquad (4)$$

$$P = \Pr ecision(i,j) = (N_{ij})/(N_j)\,. \quad (5)$$

$$R = \operatorname{Re} call(i,j) = (N_{ij})/(N_i)\,, \quad (6)$$

where, N_{ij} is the number of intersections between retrieved images and the relevant images (images in the same class) of the query image. N_j is the number of retrieved images. N_i is the number of relevant images.

According to the size of the datasets a 5×5 SOM is used at each level of the HSOM for the first dataset, and a 3×3 SOM for the second data set.

Fig. 5 (the upper for Dataset1 and the lower for Dataset2), shows a comparison between uncombined, combined using ICA, combined using PCA, and concatenated features (CS and EH). In uncombined features two cases are compared: (a) CS in the first level and EH in the second, (b) EH in the first level and CS in the second. In (c), the lower component of ICA is used in the first level and the higher in second level. The same is done in (d) but using PCA rather than ICA. In (e), CS and EH are normalized using Gaussian Normalization [4] before concatenating them. Each level of the HSOM uses this feature.

As shown in fig. 5, combined features give better performance than uncombined features. Concatenating features gives the best performance in these five cases. Using ICA for combination gives better performance than PCA.

Concatenating features lead to more complexity in feature extraction, training of the HSOM, and search time for retrieval. Since feature extraction and training are done offline and are done only one time before storing them, the complexity of using ICA or PCA in these two steps will not affect the online search process.

The search complexity for the four cases (a, b, c, and d) are the same, but the search process for the concatenated features will be duplicated according to the number of features used.

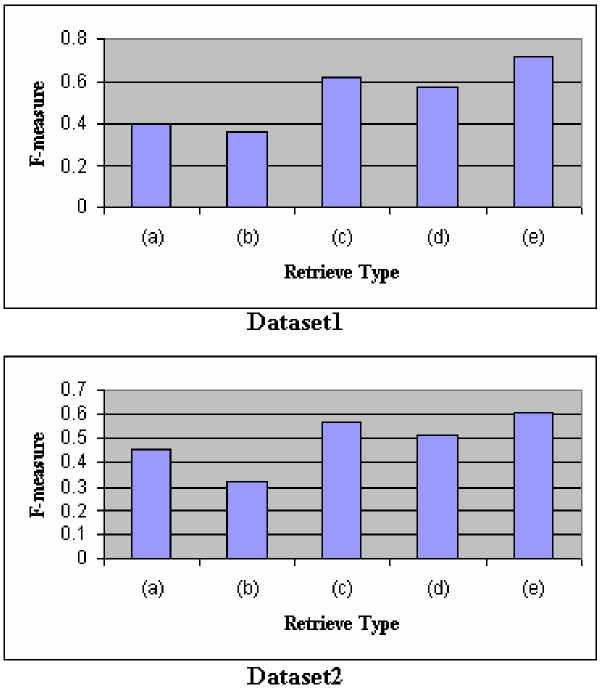

Fig. 5. F-Measure of retrieval, where a- CS then EH, b-EH then CS, c-ICA, d- PCA and e- Concatenated CS and EH.

6. Conclusions

In this paper, PCA and ICA are used to combine color structure and edge histogram. This combination gives better performance accuracy for retrieval. The additional complexity is in the feature extraction process, which is done -only once- offline. The used datasets are partitioned into classes to quantify the success of an image query. As the retrieval process depends on the clustering, the number of retrieved images may be more than the requested number. Adding techniques to rank and excluding the extra images is an ongoing area of our research.

References

[1] G. Amato, F. Falchi, C. Gennaro, F. Rabitti and P. Savino, "Improving Image Similarity Search Effectiveness in a Multimedia Content Management System", International Workshop on Multimedia Information Systems (2004).

[2] Freefoto free Pictures, http://www.freefoto.com [Accessed 2 Jul 2007].
[3] A. Hyvärinen and E. Oja "Independent Component Analysis: Algorithms and Applications", Helsinki University of Technology, Neural Networks, Vol. 13 (2004).
[4] Q. Iqbal and J.K. Aggarwal, "Combining Structure, Color and Texture for Image Retrieval: A Performance Evaluation", 16th International Conference on Pattern Recognition (ICPR) (2002).
[5] J. Laaksonen, "PicSOM--Self-Organizing Image Retrieval with MPEG-7 Content Descriptors", IEEE Transactions on Neural Networks, Vol. 13 (2002).
[6] B.S. Manjunath, "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11 (2001).
[7] B.S. Manjunath P. Salembier and T. Sikora, "Introduction to MPEG-7: Multimedia Content Description Interface" (2002).
[8] J. Pakkanen and J. Iivarinen, "Evaluating SOM as An Index in Content-Based Image Retrieval", the 10th Finnish Artificial Intelligence Conference (2002).
[9] J. Shlens, "A Tutorial on Principle Component Analysis: Derivation, Discussion and Singular Value Decomposition" (2003).
[10] M. Steinbach, G. Karypis and V. Kumar, "A Comparison of Document Clustering Techniques", Text Mining Workshop (2000).
[11] R.C. Veltkamp and M. Tanase, "'Content-Based Image Retrieval Systems: A Survey, Department of Computing Science", Utrecht University (2000).
[12] Viper Group-Multimedia Information Retrieval and Management http://viper.unige.ch/images/washington [Accessed 14 Sep 2007].
[13] C.S. Won, D.K. Park and S-J. Park, "Efficient use of MPEG-7 Edge Histogram Descriptor", ETRI Journal, Vol. 24 (2002).

Received April, 26, 2008
Accepted April 21, 2009