Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Multiple-instance content-based image retrieval employing isometric embedded similarity measure

John Y. Chiang[a,*], Shuenn-Ren Cheng[b]

[a]Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, Taiwan
[b]Department of Business Administration, Cheng Shiu University, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 16 November 2007
Received in revised form 4 May 2008
Accepted 23 June 2008

Keywords:
Content-based image retrieval (CBIR)
Multiple-instance learning
Fractal orthonormal basis
Diversity density

ABSTRACT

In image-based retrieval, global or local features sufficiently discriminative to summarize the image content are commonly extracted first. Traditional features, such as color, texture, shape or corner, characterizing image content are not reliable in terms of similarity measure. A good match in the feature domain does not necessarily map to image pairs with similar relationship. Applying these features as search keys may retrieve dissimilar *false-positive* images, or leave similar *false-negative* ones behind. Moreover, images are inherently ambiguous since they contain a great amount of information that justifies many different facets of interpretation. Using a single image to query a database might employ features that do not match user's expectation and retrieve results with low precision/recall ratios. How to automatically extract reliable image features as a query key that matches user's expectation in a content-based image retrieval (CBIR) system is an important topic.

The objective of the present work is to propose a multiple-instance learning image retrieval system by incorporating an *isometric embedded* similarity measure. Multiple-instance learning is a way of modeling ambiguity in supervised learning given multiple examples. From a small collection of positive and negative example images, semantically relevant concepts can be derived automatically and employed to retrieve images from an image database. Each positive and negative example images are represented by a linear combination of fractal orthonormal basis vectors. The mapping coefficients of an image projected onto each orthonormal basis constitute a feature vector. The Euclidean-distance similarity measure is proved to remain consistent, i.e., *isometric embedded*, between any image pairs before and after the projection onto orthonormal axes. Not only similar images generate points close to each other in the feature space, but also dissimilar ones produce feature points far apart.

The utilization of an isometric-embedded fractal-based technique to extract reliable image features, combined with a multiple-instance learning paradigm to derive relevant concepts, can produce desirable retrieval results that better match user's expectation. In order to demonstrate the feasibility of the proposed approach, two sets of test for querying an image database are performed, namely, the fractal-based feature extraction algorithm vs. three other feature extractors, and single-instance vs. multiple-instance learning. Both the retrieval results, execution time and precision/recall curves show favorably for the proposed multiple-instance fractal-based approach.

1. Introduction

The retrieval of digital images from an image database is an active research area due to the inefficiency of query processing utilizing traditional textual language. Most image retrieval paradigms fall between automated pixel-oriented information models and fully human-assisted database [1–4]. These approaches differ in application domain, visual features extracted, features discrimination criteria employed and query mechanisms supported. Feature vector characterizing image properties is generally composed of color, texture, shape and location information. Distance measure, e.g., n-dimensional Euclidean distance, is utilized to compute the similarity between different feature vectors. Query specification tools are provided to allow user-constructed sketches and weight assignments among different feature components, etc. As an example, the QBIC system allows the color, texture, or shape of an image or part of an image to be compared with feature vectors from database images using Euclidean similarity measure. The retrieval of similar images

[*] Corresponding author. Tel.: +886 934 151515; fax: +886 7 5254301.
E-mail address: chiang@cse.nsysu.edu.tw (J.Y. Chiang).

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.06.020

from a database corresponds to determining neighboring points in the proximity of the feature point of a query image.

The mapping of an image to the corresponding feature vector is a process of dimensionality reduction. By finding a lower-dimensional representation of an image, an effective feature vector is expected to contain vital characteristics of the original. The pitfalls associated with traditional approaches are two-fold, namely (1) representing features extracted are not powerful enough to discriminate between similar and dissimilar images, and (2) multiple features extracted from a single image do not necessarily match user's expectation. In this paper, an *isometric embedded* feature extraction method employing fractal orthonormal bases (FOB) is introduced. The features extracted from positive and negative example images are further combined by a multiple-instance learning paradigm to induce feature commonality to query image database.

In what follows, the proposed isometric embedded feature extraction method employing fractal orthonormal basis will be introduced first. The conservation of Euclidean distance-based similarity measure before and after the mapping onto orthonormal basis will be proved. Image pairs with long feature distance in the feature domain are guaranteed to be dissimilar ones in the image domain, while feature points close to each other correspond to similar images. Next, the procedure for combining multiple features extracted from positive and negative example images to forge a unified query key will be outlined. This multiple-instance learning procedure identifies common positive features and excludes negative ones to further clarify the user's searching criteria. The last section shows the effectiveness of this novel approach by comparing the retrieval results of (1) single-instance *vs.* multiple-instance learning by applying the same fractal-based feature extraction paradigm, and (2) multiple-instance retrieval by employing the proposed fractal orthonormal basis approach and other feature extraction techniques. Due to the preservation of distance relationship in both the image and feature domains for the fractal-based feature extraction method, and the derivation of commonality from multiple example images, consistent search results matching user's expectation are obtained in both tests.

2. Isometric embedded (FOB)

Even though similar images generally derive feature points close to each other, there is no guarantee that dissimilar images will map to distant feature points. For example, the comparison of color feature usually employs certain measure of histogram. Images with similar histogram distributions will be regarded as similar under this scheme. However, color histogram represents an image's global feature. With an analogous histogram distribution, the color within a dissimilar image might be locally distributed in a totally different manner. Using color feature as a measure of similarity between images is not powerful enough to exclude false-positive cases. Moreover, a query image might be rotational, scaling, shifted, or noise-corrupted variations of database images. A traditional retrieval algorithm might not be sufficiently robust to include similar database images of these variations, causing the occurrence of false dismissal.

The corresponding feature vectors f_1,f_2,f_3, and f_q of images i_1,i_2,i_3, and i_q, respectively, are shown in Fig. 1. The derived feature points in the feature domain might not preserve the same spatial distance relationship as their counterparts in the image domain. When an image i_q is used for querying a database, i_1,i_3 will be included in the search result due to the proximity of points f_1,f_3 with f_q in the feature space. However, image i_2 will be excluded since point f_2 is considered as too distant from f_q. A dissimilar image, e.g., image i_3, mistakenly classified as similar is called a false-positive, while a similar image, e.g., image i_2, incorrectly excluded from the final search result is referred to as a false-negative. Being unable to provide stable distance measure, most systems try to minimize false-negative results at the expense of an increased number of false-positives. A compact, perceptually relevant representation of an image content that preserves the distance relationship in terms of similarity metric in both image and feature spaces is highly desirable.

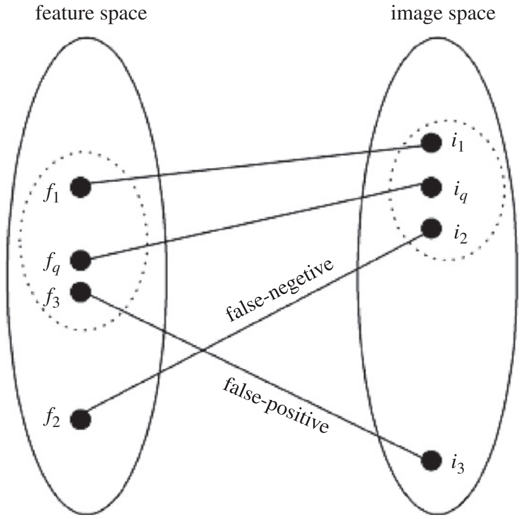

Fig. 1. The distance relationship between image points and corresponding feature points is not preserved through most feature extraction processes. Image-feature pairs $\{i_2,f_2\}$ and $\{i_3,f_3\}$ illustrate the false-negative and false-positive cases, respectively.

Barnsley suggested that storing images as collections of transformations could lead to image compression [5]. Jacquin was among the first to publish a fractal image compression scheme by regular partitioning the image [6]. Jacquin's method is based on partitioned iterated function systems and affine transform acting locally rather than globally. The accurate coding of a range block is dependent upon there being a self-similar domain block in the codebook. Because this piecewise self-similar model is an approximation of real-world data, there is no guarantee that a perfect mapping can be found. Observing that the iterative function system (IFS) coding technique seems to have a limit in the accuracy that an image can be coded, Vines proposed a scheme by finding a set of basis vectors to best represent the image in the sense of achieving higher fidelity with good compression [7]. Vines' method was intended to improve the decoded signal-to-noise ratio of fractal compression. However no application of Vines' approach to image database retrieval was ever suggested.

According to Vines' approach, a set of orthonormal basis vectors is created by the Gram–Schmidt procedure and the range blocks are coded by projecting the block elements onto this basis. The principle in determining the orthonormal set is to create a basis that allows each range block to be accurately represented with a minimum number of the basis vectors. These FOB are derived from the domain vectors. With these vectors, the range blocks can be encoded with a simple projection operation, and the map parameters will be the corresponding weights for this orthonormal basis.

For a range block of size $L_R \times L_R$, let $M_r = L_R^2$ be the length of the range vectors, and $\boldsymbol{R}_I = \{\tilde{r}_i^I\}_{i=1}^{N_R}$ be the set of all range vectors $\tilde{r}_i^I$, $i = 1 \sim N_R$, in an image I. Three basis vectors v_1,v_2, and v_3, determined a priori according to Vines' approach, are orthonormalized later to form the first three vectors of the required M_r orthonormal basis vectors, where $\tilde{v}_1 = \{1,1,\ldots,1\}^{\mathrm{T}}$, the DC value, $\tilde{v}_2 = \{0,1,2,3,4,5,6,7,0,1,2,3,4,5,6,7,\ldots,0,1,2,3,4,5,6,7\}^{\mathrm{T}}$, the tilt along x-axis, and $\tilde{v}_3 = \{0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,\ldots,7,7,7,7,7,7,7,7\}^{\mathrm{T}}$, the tilt along y-axis.

The remaining basis vectors will be chosen to span the $(M_r - 3)$-dimensional subspace S^0 perpendicular to the subspace spanned by a priori vectors v_1, v_2, and v_3. At the k th iteration, the i th projected range vector is denoted as s_i^k that resides in a corresponding subspace S^k. The optimal basis vector direction is determined by taking the s_i^k vector with the largest correlation to all of the other s_i^k vectors, i.e., the vector s_i^k maximizes the following equation is selected:

$$\sum_{j=1, j\neq i}^{N_R} |(s_i^k \cdot s_j^k)|,$$

where $|(s_i^k \cdot s_j^k)|$ is the absolute value of the inner product of s_i^k and s_j^k.

Once each basis vector direction is determined, the remaining s_i^k vectors are projected onto the subspace orthogonal to s_l^k by the following projection operator:

$$P_{S_k} = I - s_l^k (s_l^{k^{\mathrm{T}}} s_l^k)^{-1} s_l^{k^{\mathrm{T}}}.$$

The chosen basis vector direction is saved as t_k and the process is repeated until the necessary $M_r - 3$ vectors are obtained. In this manner, the set of $M_r - 3$ orthogonal vectors, $\{t_i\}_{i=1}^{M_r-3}$, that best represents the subspace S^0 is determined. A search is then performed through the domain vectors to find the best set of domain vectors for these direction vectors. The domain vector with the largest component in the direction of the direction vector is selected. Because it is possible that one domain vector has the largest component on more than one direction vector, each domain vector is only allowed to be used once.

The three fixed vectors $\tilde{v}_1, \tilde{v}_2, \tilde{v}_3$ and the $M_r - 3$ domain vectors form a set of M_r vectors that span the space of the range vectors. If the selected $M_r - 3$ domain vectors are denoted sequentially as $\{\tilde{v}_i\}_{i=4}^{M_r}$, then the set of fractal basis vectors is equal to $[\tilde{v}_1, \tilde{v}_2, \tilde{v}_3, \tilde{v}_4, \ldots, \tilde{v}_{M_r}]$. These basis vectors are further processed using the Gram–Schmidt procedure to obtain the corresponding fractal orthonormal basis matrix $\boldsymbol{Q} = [\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_{M_r}]$. Features representing an image are then extracted by projecting range vectors into each fractal orthonormal basis.

The coding of a given range vector $\tilde{r}_i^I$ of image I with a set of weight $\tilde{w}_i^I$ is equivalent to $\tilde{w}_i^I = \boldsymbol{Q}^{\mathrm{T}} \tilde{r}_i^I$ or $\tilde{r}_i^I = \boldsymbol{Q} \tilde{w}_i^I$. The previous two equations define the basic encoding and decoding process. An image I with range vector set $\boldsymbol{R}_I = \{\tilde{r}_i^I\}_{i=1}^{N_R}$, the set of weights $W_I = \{w_{ij}^I,\ i = 1, \ldots, N_R,\ j = 1, \ldots, M_r\}$ can be derived according to the fractal orthonormal basis matrix $\boldsymbol{Q}$. The set of weights W_I serves both as a feature vector and compression coefficients of image I. From the perspective of image database retrieval, the weight matrix W_I represents the signature of image I and a distance metric d_{IJ} is employed to measure the similarity of images I and J based on feature points W_I and W_J in the M_r-dimensional space. The weight matrix W_I is also utilized in the later decompression process to reconstruct the original image from the image coding/decoding perspective.

According to the above paradigm, an image I is partitioned into non-overlapping range blocks $\boldsymbol{R}_I = \{\tilde{r}_i^I\}_{i=1}^{N_R}$. Each range block $\tilde{r}_i^I$ is decomposed into a linear combination of orthonormal basis vectors by employing the same fractal orthonormal basis matrix $\boldsymbol{Q}$. The set of coefficients for all range blocks, W_I, is the signature for image I used in the retrieval of image database. Since the original image can be reconstructed by employing the feature set W_I with high S/N ratio, W_I therefore is a good representation of image I with little information loss. The similarity measure between two images I, J is determined by comparing a distance metric d_{IJ} between W_I and W_J. Next, we will show that the distance metric employing Euclidean measure is isometric embedded in both image and feature domains, i.e., the proximity of two image points I, J in the image space is indicative of corresponding feature points W_I, W_J, and vice versa.

Proposition. *The Euclidean distance between images I and J in the image domain and that of the corresponding feature vectors W_I and W_J, derived by projecting range blocks of I and J onto a set of orthonormal basis vectors, are equivalent.*

Proof. The Euclidean distance d_{IJ} between images I and J can be formulated as $d_{IJ} = \|I - J\|$, or expressed in terms of range blocks

$$d_{IJ} = \sum_{i=1}^{N_R} \|\tilde{r}_i^I - \tilde{r}_i^J\|,$$

where $\tilde{r}_i^I \in \boldsymbol{R}_I, \tilde{r}_i^J \in \boldsymbol{R}_J$. Each range block $\tilde{r}_i^I$ and $\tilde{r}_i^J$ of image I and J can be further represented as a linear combination of M_r orthonormal basis vectors $\boldsymbol{Q} = [\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_{M_r}]$, with coefficients $\tilde{w}_{ij}^I \in \mathbf{W}_{I,} \tilde{w}_{ij}^J \in \mathbf{W}_J,\ i = 1 \ldots N_R,\ j = 1, \ldots, M_r$, respectively. Again

$$\begin{aligned} d_{IJ} &= \sum_{i=1}^{N_R} \left\| \sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J) \tilde{q}_j \right\| \\ &= \sum_{i=1}^{N_R} \left(\sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J)^2 \tilde{q}_j^2 \right)^{1/2}, \end{aligned}$$

Since all basis vectors are orthonormal, i.e.,

$$\tilde{q}_j^2 = 1, \quad \tilde{q}_i \cdot \tilde{q}_j = 0 \quad \forall i, j \in \{1 \cdots M_r\},\ \ i \neq j.$$

All cross-product terms are zeros.

$$\begin{aligned} d_{IJ} &= \sum_{p=1}^{N_R} \left(\sum_{q=1}^{M_r} (w_{ij}^I - w_{ij}^J)^2 \right)^{1/2} \\ &= \sum_{i=1}^{N_R} \left\| \sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J) \right\| \\ &= d_{W_I W_J}. \end{aligned}$$

The above proposition states that the Euclidean distance measure remains the same after the projection of points in image space into a set of orthonormal basis vectors in the feature domain. The image space and the feature space are "*isometric*" to each other. From this, we can conclude that the closeness of two image points in the image space, i.e., $d_{IJ} \leqslant \varepsilon$, implies the proximity of the corresponding feature points in the feature domain, $d_{W_I W_J} \leqslant \varepsilon$. The above statement is logically equivalent to "if feature vectors W_I and W_J are distant from each other, then image I is also distant from image J." Since similar images are mapped to close feature points and only points close to the feature point of a query image will be included in the retrieval results, images corresponding to distant features points will be excluded. This property suggests that false-negative cases are unlikely to occur. Therefore, employing the proposed paradigm will not falsely ignore any similar images based on the Euclidean metric in the feature space. Similar objects will be included in the final retrieval set. Another facet of the above proposition reveals that the proximity of feature points in feature domain indicates the closeness of image points in image space. This statement is equivalent to stating that dissimilar image points imply that feature points are far apart. Therefore, searching in near proximity of the feature point within a query image will not return dissimilar images. This property ensures that no false-positives occur. Utilizing the coefficients of the linear combination of an orthonormal basis set as feature vectors will retrieve consistent database retrieval result excluding both false-positive and false-negative cases. □

3. Multiple-instance learning

Even though the above-mentioned FOB approach can extract isometric preserving features contained within an image, yet an image commonly contains a plurality of concepts, i.e., feature clusters, scattered throughout the image space. For example, several foreground objects present in front of a non-uniform background are commonly observed in a natural scene. Given a single image to query an image database is ambiguous in terms of which concept the user intends to be included in the query. Take waterfall scene as an example, the image might contain concepts such as waterfall itself, rocks surrounding the waterfall and the plants growing around the waterfall, etc. Most existing content-based image retrieval (CBIR) systems require either global information is used or a user must explicitly indicate which part of the image is of interest. Given a reliable feature extractor, our next goal is to take a small set of positive and negative example images, each associated with extracted feature clusters the user desired or unwanted. The positive concepts are then integrated and the negative ones excluded to automatically generate the query key for a CBIR system.

In multiple-instance learning [8–11], each training example image corresponds to a bag containing a plurality of instances. Each instance corresponds to a point in feature space. A bag is labeled negative if all the instances in it are negative. On the other hand, a bag is labeled positive if there is at least one instance in it which is positive. Note the instances within a bag are not labeled themselves, only the bag is labeled positive or negative. Each bag is therefore an ambiguous example image. From a collection of labeled bags, the learner tries to induce common concepts that will label unseen bags correctly. Even though Maron [10] applied multiple-instance learning on an image database system to classify images of natural scenes, yet the feature extractor (referred as bag generator in Ref. [10]) employed does not possess isometric-embedded property. How to extract representative image features to be applied in the following multiple-instance commonality induction process remained an open question, according to Maron Our work intends to answer this question by providing a novel isometric-embedded feature extractor. The isometric-embedded features extracted from both positive and negative example images are further forged through multiple-instance learning procedure, as suggested by Maron, to derive a unified query key for retrieving relevant images from database. In the previous section, the fractal orthonormal feature extractor is proved to be isometric embedded in both image and feature domains. Given this desirable property, the ambiguity between feature clusters existing within the same image is further clarified with the help of multiple-instance learning. The isometric-embedded features extracted from both positive and negative example images are then forged through multiple-instance learning procedure to derive a unified query key for retrieving relevant images from database.

Given an arbitrary combination of positive and negative example images specified by the user, finding the common features existing among all positive images while excluding features found in negative images is computationally complex due to local maxima and the size of search space. In this paper, a diverse density (DD) approach is employed to induce the instance (feature) commonality among multiple bags (example images) [8]. DD is a general tool with which to learn from multiple-instance examples, not a feature extractor itself. The DD algorithm derives areas in feature space that are close to at least one instance from every positive bag and far from every negative instance. The algorithm searches the feature space for points with high DD. Once the point with maximum DD is found, a new image is classified positive if one of its features is close to the maximum DD point.

Denote the i th positive bag as B_i^+, its j th feature instance B_{ij}^+, and the negative counterparts B_i^- and B_{ij}^-, respectively. Given a set of positive and negative example images, a common feature instance corresponds to locating a feature point t that maximizes $\Pr(t|B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-)$ in feature space. By applying Bayes' rule and a uniform prior, the above equation is equivalent to $\arg\max_t \Pr(B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-|t)$ as follows:

$$\begin{aligned}
&\Pr(t|B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-)\\
&\quad=\frac{\Pr(t,B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-)}{\Pr(B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-)}\\
&\quad=\frac{\Pr(B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-|t)\Pr(t)}{\Pr(B_1^+,\ldots,B_n^+,B_1^-,\ldots,B_m^-)}.
\end{aligned}$$

Assume further conditional bags independence given feature point t, it can be decomposed into $\arg\max_t \prod_i \Pr(B_i^+|t)\prod_i \Pr(B_i^-|t)$

In order to instantiate the above general diversity density formulation, the noisy-or model [12] and Gaussian-like distribution are utilized. The probability $\Pr(t|B_i^+)$that not all points missed the positive instances can be represented as

$$\begin{aligned}
\Pr(t|B_i^+) &= \Pr(t|B_{i1}^+,B_{i2}^+,\ldots)\\
&= 1-\prod_j(1-\Pr(t|B_{ij}^+))\\
&= 1-\prod_j(1-\exp(-\|B_{ij}^+-t\|^2)),
\end{aligned}$$

and the probability $\Pr(t|B_i^-)$ that all points missed the negative instances

$$\Pr(t|B_i^-)=\prod_j(1-\Pr(t|B_{ij}^-))=\prod_j(1-\exp(-\|B_{ij}^--t\|^2)),$$

respectively, where $\|\|^2$ is the Euclidean distance measure in the feature space.

Finding an appropriate location in feature space is accomplished through diversity density procedure by finding the best weighting of instances. Each instance in the feature space is generated by projecting the image into the FOB. The utilization of an isometric-embedded fractal-based technique to extract reliable image features, combined with a multiple-instance learning paradigm to derive relevant concepts, can produce desirable retrieval results that better match user's expectation.

4. Experimental results

In order to demonstrate the power of the proposed fractal orthonormal basis approach and multiple-instance query framework, two different sets of test are performed, namely (1) single-instance *vs.* multiple-instance query by employing the same novel fractal-based feature extractor, and (2) multiple-instance retrieval by employing the proposed fractal orthonormal basis approach *vs.* other feature extractors. The first test intends to demonstrate that even though with a powerful fractal-based feature extractor, multiple-instance learning is highly efficient in summarizing the commonality among positive and negative example images. The second set of experiments illustrates the importance of a reliable feature extractor by utilizing different feature extraction schemes with the same multiple-instance paradigm. The multiple-instance query results are analyzed through precision–recall figures and the FOB approach consistently obtains the highest precision–recall value.

Since it is not clear to the authors how to derive the total number of correct images, necessary for the calculation of precision–recall values, contained in a public image database, e.g., COREL, a custom database consisting of tropical fish images with variable size is constructed. A total of 2636 images with natural or uniform backgrounds are downloaded from 14 websites, as shown in Fig. 2.

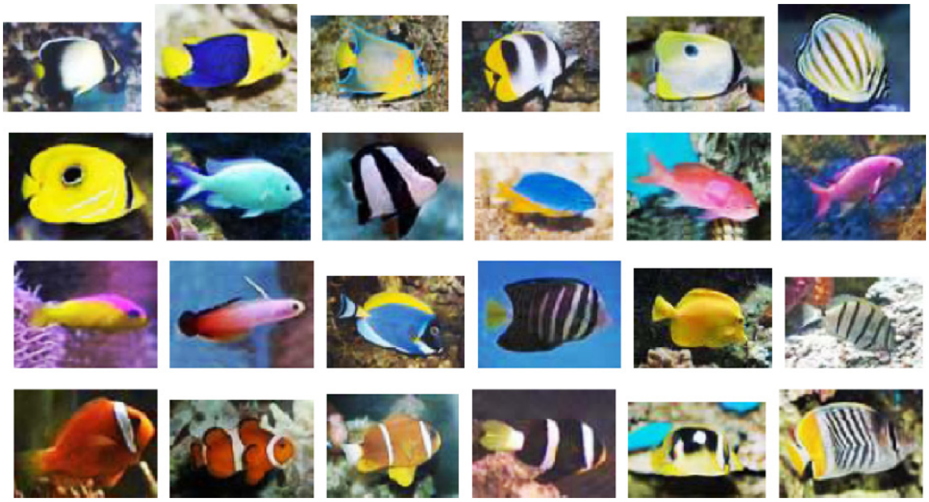

Fig. 2. Images contained in the tropical image database constructed.

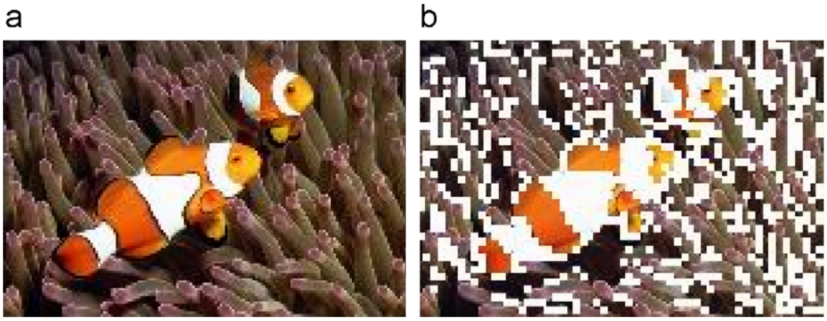

Fig. 3. (a) An image in the database and (b) major feature clusters, composed of foreground yellow-white stripes of clown fish and background anemone, extracted by utilizing fractal orthonormal bases.

Each image is partitioned into non-overlapping range blocks with size 8 × 8. The R, G, B color components are processed independently to determine the fractal orthonormal basis in each color plane. The fractal orthonormal basis matrixes $\boldsymbol{Q}_R, \boldsymbol{Q}_G$, and $\boldsymbol{Q}_B$ derived are 64-dimensional each. The 64 fractal basis vectors of each color plane are composed of uniform, edge or texture regions. The coefficient corresponding to the vector $\tilde{q}_1$, the orthonormalized version of the first a priori vector $\tilde{v}_1 = \{1, 1, \ldots, 1\}^{\mathrm{T}}$, is considered as the brightness level of a specific color component within an image.

A total of 64 coefficients for each color component are derived by projecting a range vector into an orthogonal space with 64 dimensions. A color range vector can therefore be losslessly reconstructed by employing 192 linear coefficients. Since the energy is highly concentrated in relatively few numbers of axes, most coefficients are negligible in the later similarity comparison stage. Only the three most significant coefficients per color component are employed in later Euclidean distance computation, whereas the remaining less significant coefficients are considered as zero values. Since all projection coefficients of database images are calculated only once and stored as compression coefficients, the computation of similarity measure involves only the derivation of feature coefficients for the query image, subtraction of matching coefficients and summation of all squared differences. Therefore, the retrieval procedure is very efficient and proceeded in the compressed domain.

Fig. 3 demonstrates the power of the proposed FOB. The major features extracted are composed of the foreground yellow-white stripes of clown fish and the background anemone. If a user expects to retrieve relevant clown fish images from database, employ this specific image as a query might still retrieve undesirable images containing anemone as the major constituent object. Applying multiple-instance learning, another positive image with clown fish in different background, or negative image with the same background can enhance the common concepts existing among positive images or exclude negative ones, and induce the desirable feature clusters to retrieve expected results.

4.1. Single-instance vs. multiple-instance

Fig. 4(a) is the positive example image used in the single-instance test. One more positive image Fig. 4(b) and one negative image Fig. 4(c) are added in the following multiple-instance test. Fig. 4(d) demonstrates the single-instance retrieval result, while Fig. 4(e) shows the multiple-instance counterpart by employing two positive example images and one negative image. Comparing Figs. 4(d) and (e) it clearly demonstrates that multiple-instance learning induces the common characteristics of white and yellow stripes existing in two clownfish images, while excluding the dark texture background in the negative image. A more satisfactory result is obtained for the multiple-instance case.

In order to further demonstrate that feature commonality can be further consolidated by adding more example images, a third positive image, as shown in Fig. 5(a), is included in the multiple-instance query, in addition to the original two positive images and one

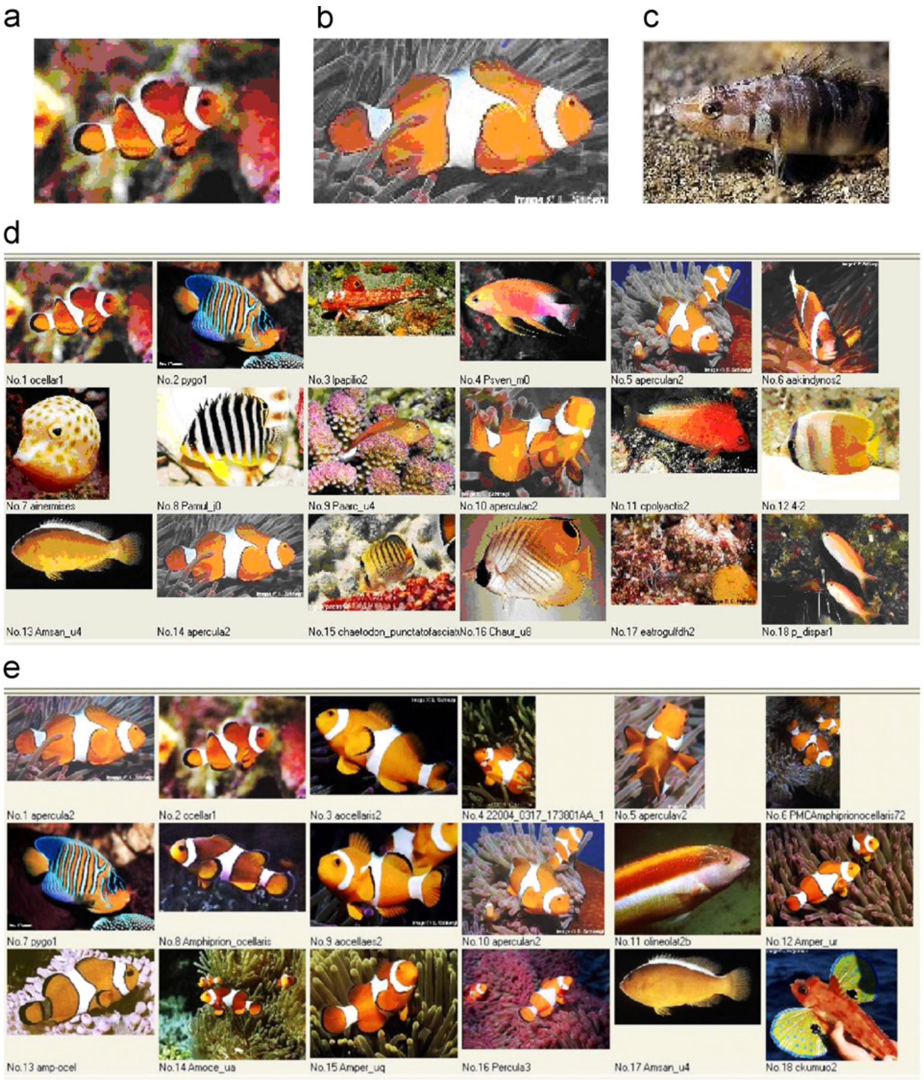

Fig. 4. (a) Positive image 1, (b) positive image 2, (c) negative image, (d) single-instance retrieval result by using (a) as a query, and (e) multiple-instance retrieval result using two positive images (a), (b) and one negative image (c).

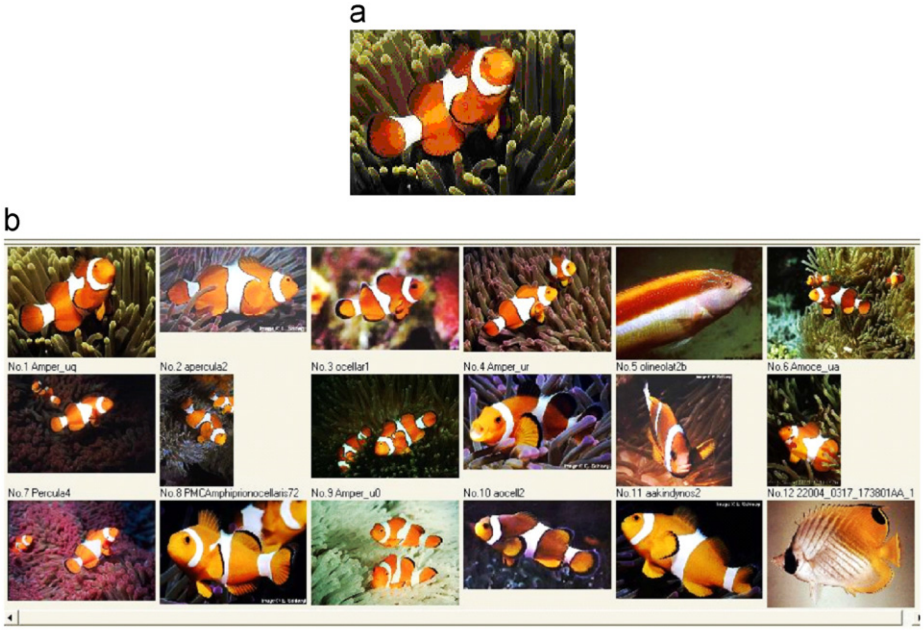

Fig. 5. (a) Positive image 3, (b) multiple-instance retrieval result using three positive images Figs. 4(a), (b) and 5(a) and one negative image Fig. 4(c).

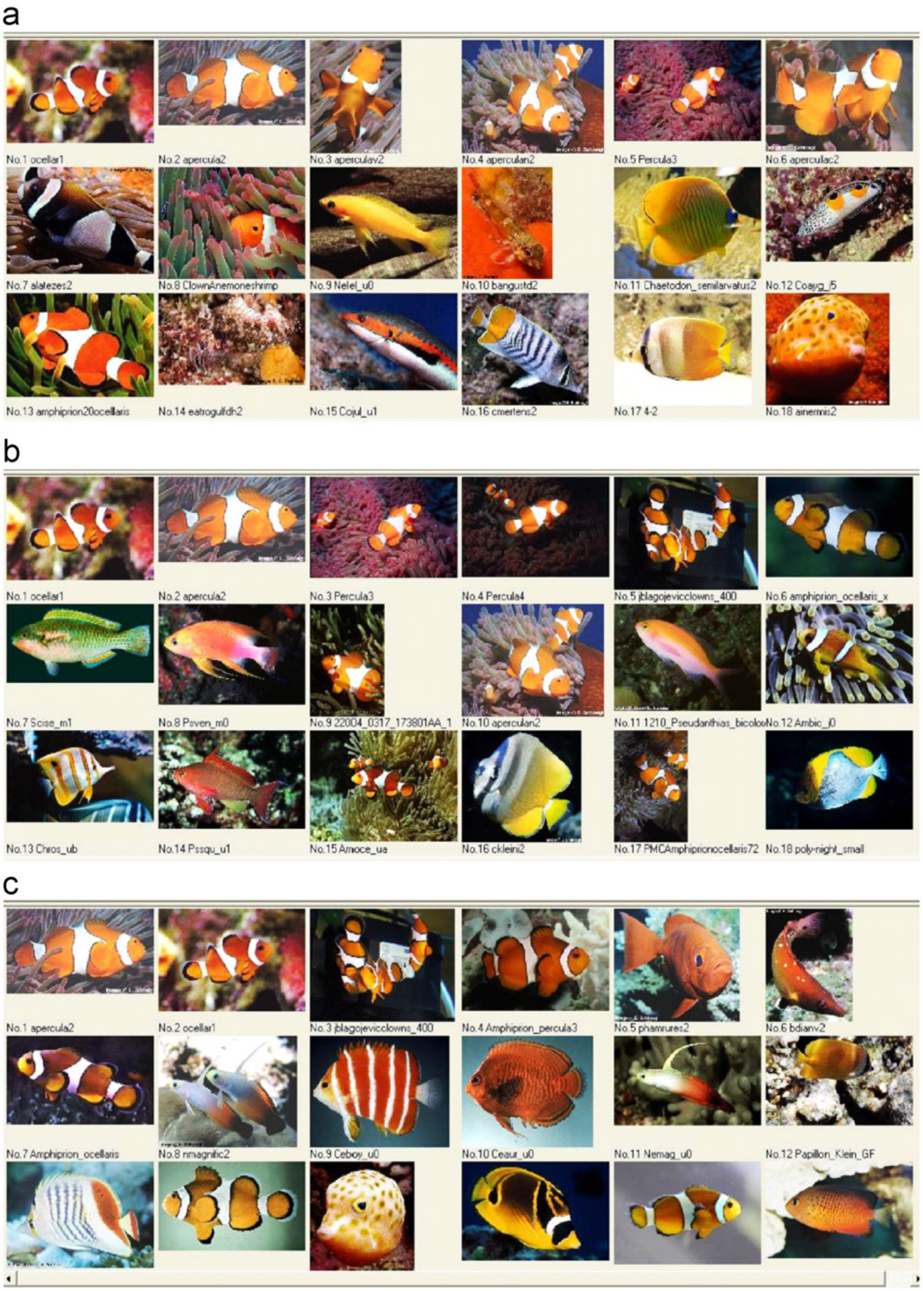

Fig. 6. Multiple-instance retrieval results by applying two positive images Figs. 4(a) and (b), and one negative image Fig. 4(c): (a) Panchanathan's texture-style codebooks with codebook size 256; codeword dimension 16; and L2-metric for histogram similarity. (b) Cohen's fractal-based image search methods. The color of thumbnail images is compared first to reduce search space. Prosperous thumbnail images are then enlarged. The difference between the original and enlarged images is coded by using fractal paradigm, and (c) Murtagh et al.'s fractal and multiscale approach. Mallat's dyadic wavelet transform, with three scale levels, is employed. The entropy of an image with different scales is calculated, with noise variance $\sigma = 1$.

negative image. The result in Fig. 5(b) shows even more improvement over that obtained in Fig. 4(e).

4.2. Fractal orthonormal basis vs. other feature extractors

Performance evaluation of multiple-instance query with three other feature extraction techniques is performed. They are Panchanathan's texture-style codebooks for content retrieval [13], Cohen's fractal-based image search methods [14], and Murtagh et al.'s fractal and multiscale approach [15]. Fig. 6 shows the retrieval results of applying these three approaches with the same multiple-instance combination of two positive images Figs. 4(a) and (b), and one negative image Fig. 4(c). Comparing with the retrieval result by applying FOB feature extractor, as shown in Fig. 4(e), FOB approach obviously compares favorably with those of other feature representation schemes.

The quantitative evaluation of the proposed FOB algorithm and three other methods implemented is also performed through plotting of precision–recall curve. Precision is defined as the ratio of the number of correct images retrieved to the total number of images retrieved. Recall is defined as the ratio of the number of correct images retrieved to the total number of correct images in the database. For our test, the total number of images retrieved is fixed as 18. The precision and recall ratio for four different cases, namely, one positive images, two positive and one negative images, three positive and

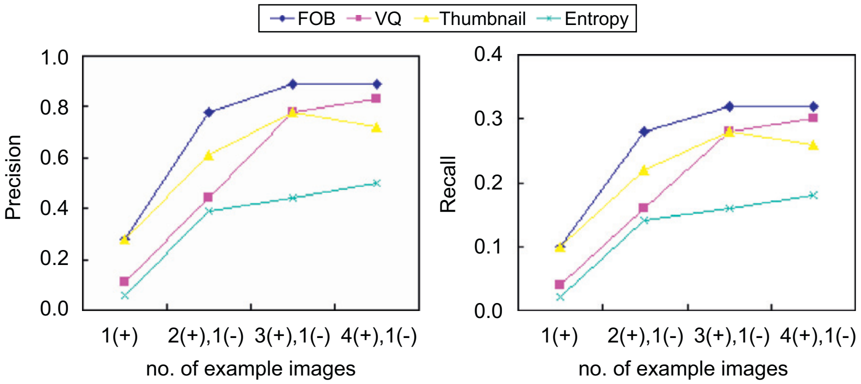

Fig. 7. Precision–recall curve of the proposed FOB scheme compared with the methods of Panchanathan [13], Cohen [14], and Murtagh [15].

Table 1
The execution time comparison of different methods in various processing steps of a CBIR system

	FOB	Panchanathan [13]	Cohen [14]	Murtagh [15]
Pre-processing	6 s Construction of fractal orthonormal basis	32 s Construction of a 256 × 16 codebook		
Indexing for a database with 2636 images	98.4 s Derivation of projection coefficients for all DB images	75.2 s finding the best match codebook entry for all DB images	32 952.4 s Thumbnail image generation + comparison of thumbnail images + IFS coded the difference between enlarged thumbnail and the original images	49 236.2 s Three-scale wavelet + IFS + entropy for all DB images
Commonality induction from query image(s)	1 (+): ≈ 0.03 s 2 (+)1(−): 1.1 s 3 (+)1(−): 2.1 s 4 (+)1(−): 3.8 s Derivation of projection coefficient for the query image(s) + learning	1 (+) ≈ 0.03 s 2 (+)1(−): 0.9 s 3 (+)1(−): 2.0 s 4 (+)1(−): 3.6 s Derivation of the best match codebook entry for the query image(s) + learning	1 (+): 13.2 s 2 (+)1(−): 42.7 s 3 (+)1(−): 55.1 s 4 (+)1(−): 69.8 sIFS-coded query image(s) + learning	1 (+): 19.5 s 2 (+)1(−): 59.3 s3 (+)1(−): 79.2 s 4 (+)1(−): 101.9 s Three-scale wavelet + IFS + entropy for the query image(s) + learning
Image retrieval similarity comparison with 2636 DB images	1 (+): 10.2 s 2 (+)1(−): 9.4 s 3 (+)1(−): 6.4 s 4 (+)1(−): 5.2 s	1 (+): 14.0 s 2 (+)1(−): 11.8 s 3 (+)1(−): 8.2 s 4 (+)1(−): 6.7 s	1 (+): 17.5 s 2 (+)1(−): 14.7 s 3 (+)1(−): 11.0 s 4 (+)1(−): 9.2 s	1 (+): 19.1 s 2 (+)1(−): 16.9 s 3 (+)1(−): 12.5 s 4 (+)1(−): 10.3 s

one negative images, and four positive images and one negative images are performed, as shown in Fig. 7. The multiple-instance query employing FOB approach consistently obtains the highest value. This undesirable effect of including false-positive images and excluding false-negative ones for three other image feature generation methods may be attributed to the feature extraction and similarity computation cannot be proved as isometric embedded operations.

The execution time comparison of the FOB with three other methods discussed above is also performed. The time required for different processing steps, including pre-processing, image database indexing, commonality induction from query image(s), and image database retrieval, is listed separately in Table 1.

5. Conclusions

Two key issues in CBIR applications, namely effective image feature extractor and feature cluster locator, are addressed in this paper. Lacking a good feature representation scheme for an image obviously cannot produce satisfactory retrieval result. However, a reliable image feature extractor alone might still result plural ambiguous concepts that might nullify the search results. A good feature extraction scheme combined with a multiple-instance learning paradigm will locate common feature clusters utilizing reliable image features extracted and produce search results meeting user's expectation.

The proposed fractal feature extraction method utilizes the coefficients of an image vector projecting onto a set of orthonormal axes as features. The distance relationship between image points and corresponding feature points is preserved through the feature extraction process. The reliable features extracted are further analyzed through a multi-instance learning stage to induce commonality among numerous features existing in multiple example images. The combination of a reliable feature extractor and feature cluster locator will generate query meeting user expectation, produce consistent database retrieval result and exclude both false-positive and false-negative cases.

The constraints regarding size, orientation, location, relative position, etc., can be further imposed upon the feature clusters extracted and incorporated in the similarity comparison. The isometric embedded similarity measure proposed can also explore the combination of others global or local features to produce an even more effective image feature extractor.

References

[1] V.P. Subramanyam Rallabandi, S.K. Sett, Knowledge-based image retrieval system, Knowledge Based Syst. 21 (2) (2008) 89–100.
[2] T. Leon, P. Zuccarello, G. Ayala, E. deVes, J. Domingo, Applying logistic regression to relevance feedback in image retrieval systems, Pattern Recognition 40 (10) (2007) 2621–2632.
[3] H. Chang, D.Y. Yeung, Kernel-based distance metric learning for content-based image retrieval, Image Visual Comput. 25 (5) (2007) 695–703.
[4] S.K. Saha, A.K. Das, B. Chanda, Image retrieval based on indexing and relevance feedback, Pattern Recognition Lett. 28 (3) (2007) 357–366.
[5] M.F. Barnsley, SuperFractals, Cambridge University Press, Cambridge, 2006 ISBN: 0-521844-932.
[6] A.E. Jacquin, Fractal image coding: a review, Proc. IEEE 81 (10) (1993) 1451–1465.
[7] G. Vines, Orthogonal basis IFS, in: Y. Fisher (Ed.), Fractal Image Compression, Theory and Application, Springer, New York, 1996, pp. 199–214, ISBN: 0-387942-114.
[8] O. Maron, Learning from ambiguity, Ph.D. Dissertation, Massachusetts Institute of Technology, 1998.

[9] A.L. Ratan, O. Maron, W.E.L. Grimson, T. Lozano-Perez, A framework for learning query concepts in image classification, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR'99), vol. 1, 1999, pp. 1423.

[10] O. Maron, A.L. Ratan, Multiple-instance learning for natural scene classification, in: Proceedings of the 15th International Conference on Machine Learning, 1998.

[11] Z.H. Zhou, M.L. Zhang, Solving multi-instance problems with classifier ensemble based on constructive clustering, Knowl. Inf. Syst. 11 (2) (2007) 155–170.

[12] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, CA, 1988.

[13] S. Panchanathan, C. Huang, Indexing and retrieval of color images using vector quantization, in: Proceedings of SPIE, Applications of Digital Image Processing XXII, vol. 3808, 1999, pp. 558–568.

[14] H.A. Cohen, Fractal image coding for thumbnail based image access, in: Proceedings of the Internationall Conference on Signal Processing Applications, ISSPA, Gold Coast, vol. 1, August 1996, pp. 158–161.

[15] F. Murtagh, A. Alexander, A. Bouridane, D. Crookes, J.G. Campbell, J.L. Starck, F. Bonnarel, Z. Geradts, Fractal and multiscale methods for content-based image retrieval, in: Proceedings of the Third UK Conference on Image Retrieval, Brighton, UK, 2000.

About the Author—JOHN Y. CHIANG received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 1985, the M.S. and the Ph.D. degrees in Electrical Engineering from the Northwestern University, Evanston, IL, USA, in 1987 and 1990, respectively. He is currently an Associate Professor of the Department of Computer Science and Engineering of the National Sun Yat-sen University, Kaohsiung, Taiwan. His areas of research include computer vision, image processing, and content-based image retrieval.

About the Author—SHUENN-REN CHENG received the B.S. degree in Statistics from Tung Hai University, Taichung, Taiwan, in 1988, the M.S. degree in Statistics from the New York St. Johns University, New York, USA, in 1993, respectively. He is an Assistant Professor of the Department of Business Administration of the Cheng Shiu University, Kaohsiung, Taiwan. His areas of research include experiment design and statistical analysis of patterns.