Image Retrieval Based on Fractal Signatures

John Y. Chiang Z. Z. Tsai
Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
E-mail:chiang@cse.nsysu.edu.tw E-mail: tsaitc@cse.nsysu.edu.tw

Abstract

The objective of the present work is to propose a novel method to extract a stable feature set representative of image content. Each image is represented by a linear combination of fractal orthonormal basis vectors. The mapping coefficients of an image projected onto each orthonormal basis constitute the feature vector. The set of orthonormal basis vectors are generated by utilizing fractal iterative function through target and domain blocks mapping. The distance measure remains 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 above statements are logically equivalent to that distant feature points are guaranteed to map to images with dissimilar contents, while close feature points correspond to similar images. Therefore, utilizing coefficients derived from the proposed linear combination of fractal orthonormal basis as key to search image database will retrieve similar images, while at the same time exclude dissimilar ones. The coefficients associated with each image can be later used to reconstruct the original. The content-based query is performed in the compressed domain. This approach is efficient for content-based query. Scaling, rotational, translation, mirroring and horizontal/vertical flipping variations of a query image are also supported.

Keywords:content-based image retrieval, fractal orthonormal basis, iterative function system

1. Introduction

The retrieval of digital image is an active area of research in computer science 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 schemes [1]. 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/or 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 be compared with feature vectors from database images using Euclidean similarity measure [2]. The retrieval of similar images from database corresponds to determine 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 the image, an effective feature vector is expected to contain vital characteristics of the original. The pitfall associated

with the traditional approach is that even though similar images generally derive feature points close to each other. However, 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 resembling histogram distribution will be regarded as similar under this scheme. However, even with analogous histogram distribution, the color within a dissimilar image or sub-image might be spatially distributed in a totally different manner. Using color feature as a measure of similarity between images is not powerful enough to exclude the 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 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, f_q of images i_1, i_2, i_3, i_q, respectively, are shown in Figure 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 one is called false-positive, while a similar image, e.g., image i_2, incorrectly excluded from the final search result is referred as 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.

Image retrieval by content allows a user to search image database by specifying the content of an exemplary image as the basis for retrieval [3]. In traditional content based indexing, content indices (colors, shapes or textures) for each image in the database are first extracted and appended to the image data as overheads. The corresponding feature vector of a query image is computed and compared to the stored feature vectors. Images most similar to the query are returned to the user. Given that images are usually coded in compressed format in a database, it would be more efficient if the compressed data can also be used directly as indices for content-based query. In our proposed scheme, each image is decomposed into a linear combination of fractal orthonormal basis. The coefficient of each term serves both as a feature component in the corresponding feature vector and compressed data. The content-based query followed is performed in the compressed domain. Contents of the image are embedded in the compressed data, which can be easily and efficiently used as indices for content-based image retrieval. The extracted feature vector, composed of linear coefficients, will be proved in the following section to preserve distance metric between the corresponding image points in the image domain.

In what follows, fractal orthonormal basis approach will be introduced first. The procedure of generating a set of fractal orthonormal basis for an ensemble of database images will be outlined. Next, the conservation of Euclidean distance 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, while feature points close to each other correspond to similar images. The last section shows the effectiveness of this novel approach using a butterfly image database

as an example. Due to the preservation of distance relationship in both the image and feature domains, consistent search results are obtained.

2. Fractal Orthonormal Basis Approach

Barnsley suggested that storing images as collections of transformations could lead to image compression [4]. Jaquin was among the first to publish a fractal image compression scheme by regular partitioning of the image [5]. 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 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 [6], Vines proposed a scheme by finding a set of basis vectors to best represent the image in the sense of achieving the higher fidelity with good compression [7,8]. Vines' method was intended to improve the decoded signal-to-noise ratio of fractal compression, no application to image database retrieval was ever suggested.

According to Vines' approach, a set of orthonormal basis vectors is created by 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 fractal orthonormal bases 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 $M_r = L^2{}_R$ be the length of the range vectors. Let $\boldsymbol{R}_I = \{\tilde{r}_i^{\,I}\}_{i=1}^{N_R}$ be the set of all range vectors in an image *I*. Three basis vectors $\tilde{v}_1, \tilde{v}_2$ and $\tilde{v}_3$, determined *a priori* according to Vines' approach, are orthonormalized later to form the first three of the required M_r orthonormal basis vectors, where $\tilde{v}_1 = \{1,1,\Lambda\ ,1\}^T$, the DC value, $\tilde{v}_2 = \{0,1,2,3,4,5,6,7,0,1,2,3,4,5,6,7,\Lambda\ ,0,1,2,3,4,5,6,7\}^T$, the tilt along the *x*-axis, and $\tilde{v}_3 = \{0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,\Lambda\ ,7,7,7,7,7,7,7,7\}^T$, the tilt along the *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 $\tilde{v}_1, \tilde{v}_2$ and $\tilde{v}_3$. 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:

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

where $\left|(s_i^k \cdot s_j^k)\right|$ 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

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

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 $[\tilde{v}_1, \tilde{v}_2, \tilde{v}_3, \tilde{v}_4, \Lambda, \tilde{v}_{M_r}]$. These basis vectors are further processed using the Gram-Schmidt procedure to obtain the corresponding fractal orthonormal basis matrix $\boldsymbol{Q} = [\tilde{q}_1, \tilde{q}_2, \Lambda, \tilde{q}_{M_r}]$. 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 $\tilde{w}_i^I = \boldsymbol{Q}^T \tilde{r}_i^I$ 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 $W_I = \{w_{ij}^I, i = 1, \Lambda, N_R, j = 1, \Lambda\ M_r\}$ 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 indicates that 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 $\boldsymbol{Q} = [\tilde{q}_1, \tilde{q}_2, \Lambda, \tilde{q}_{M_r}]$, with coefficients $\tilde{w}_{ij}^J \in \boldsymbol{W}_I, \tilde{w}_{ij}^J \in \boldsymbol{W}_J, i = 1\Lambda\ N_R, j = 1\Lambda\ M_r$, respectively.

$$d_{IJ} = \sum_{i=1}^{N_R} \left\| \sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J)\tilde{q}_j \right\| \cdot$$

$$= \sum_{i=1}^{N_R} \left(\sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J)^2 \tilde{q}_j^2 \right)^{1/2},$$

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

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

All cross-product terms are zeros.

$$d_{IJ} = \sum_{p=1}^{N_R} \left(\sum_{q=1}^{M_r} (w_{ij}^I - w_{ij}^J)^2 \right)^{1/2}$$

$$= \sum_{i=1}^{N_R} \left\| \sum_{j=1}^{M_r} (w_{ij}^I - w_{ij}^J) \right\| = d_{W_I W_J},$$

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., $d_{IJ} \le \varepsilon$, implies the proximity of the corresponding feature points in the feature domain, $d_{W_I W_J} \le \varepsilon$. The above statement is logically equivalent to "if feature vectors W_I and W_J are distant to each other, then image *I* is also distant to 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 that dissimilar image points imply feature points far apart. Therefore, the search in near proximity of the feature point of a query image will not return dissimilar images. This property makes sure that no false-positive will 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.

Even though any orthornormal basis set can be utilized to construct the feature space, a compact, efficient representation of an image that leads to concentrations of energy in as few coefficients as possible is preferred. Image energy concentrated in as low-dimensional subspace as possible is highly desirable due to lower computation complexity required in feature comparison process and fewer truncation errors incurred in ignoring less significant terms. The directions of axes for the aforementioned fractal orthonormal basis vectors are chosen with the largest correlation to the other range vectors in the ensemble of database images. The linear coefficients by projecting an image onto the proposed orthonormal space and the frequency components by transforming the same image by Fourier transform are compared. The projected coefficients and frequency components are first ranked according to their magnitudes, respectively. The accumulated ranked power spectrum starting with the largest coefficients or frequency components are tabulated and normalized, as shown in Figure 2. Much fewer fractal orthonormal coefficients are needed to constitute the same amount of energy in comparing with those derived by using Fourier kernels.

3. Experimental Results

In order to demonstrate the power of the proposed fractal orthonormal basis approach, a database consisting of butterfly images is constructed. A total of 1013 butterfly images with natural or uniform backgrounds are downloaded from websites http://www.thais.it/entomologia/, http://turing.csie.ntu.edu.tw/ncnudlm/index.html, http://www.ogphoto.com/index.html, http://yuri.owes.tnc.edu.tw/gallery/butterfly, http://www.mesc.usgs.gov/resources/education/butterfly, and http://mamba.bio.uci.edu/~pjbryant/biodiv/bflyplnt.htm. All images acquired are trimmed down to 320×240 pixels with 24 bits of depth per pixel. 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. Figure 3 shows the fractal orthonormal basis matrixes $\boldsymbol{Q}_R, \boldsymbol{Q}_G$, $\boldsymbol{Q}_B$ derived by following the

procedures outlined in Section II for a training set of 100 butterfly images in the database. 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 $\tilde{v}_1 = \{1,1,\Lambda\ ,1\}^T$, 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 three most significant coefficients per color component are employed in later Euclidean distance computation, the remaining less significant coefficients are considered with 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 process is very efficient. Figure 5 demonstrates the retrieval result by using a typical butterfly image (scientific name: *abpiercani*) as query. The images retrieved are arranged according to the degree of similarity from left to right, top-to bottom. The scientific name of the butterfly is listed on top of the image. After providing a query image, a user can choose a subset of R, G, or B color components as matching indices for feature discrimination. Only the coefficients corresponding to the selected color planes will be included in the calculation of Euclidean distance. The brightness factor can also be selectively turned on or off by including or excluding the coefficient corresponding to $\tilde{q}_1$ in the similarity computation to counter the influence of changing light intensity between images.

A user can also specify a sub-region of a query image for retrieval. Since each image is coded with a range block size of 8×8, a sub-region with integer multiples of 8×8 pixels can be expressed as a partitioning of non-overlapping range blocks, as shown in Figure 5 (a). However, if the specified sub-region is not integer multiples of 8×8, as illustrated in Figure 5 (b), then a mask with the largest integral multiples of 8×8 that can be fit into the sub-region is applied from the top-left corner with an increase of one pixel horizontally or vertically toward the bottom right corner. On each iterative step, only the range vectors under current mask are formulated as a superposition of orthonormal basis vectors. The coefficients derived are used as signature in the later matching process.

The proposed fractal orthonormal basis approach is also scale- and rotational-invariant. Users can specify the range of scales and rotation angles of the query image. Variations of the query image multiplied or rotated by different scaling factors and rotation angles are coded. If the size of the query region after scaling and rotation operations is not integral multiple of 8×8, then the aforementioned mask will be applied. The coefficients corresponding to each scale factor and rotation angle of the range vectors under the current mask are compared with those of all database images. Figure 6 illustrates the retrieval result by specifying a sub-region with scaling factors ranging from 0.8 to 1.2 and a 30 degree increment of rotation angle. In comparison with Figure 4, since a brown dark pattern on the wing of the butterfly is specified for searching, images with possible slight scaling (0.8 ~ 1.2) and orientation difference (every 30 degree) of the marking are included in the retrieval set.

4. Conclusions

A feature extracting method that preserves distance measure in the image and feature spaces is provided. The fractal orthonormal basis set introduced can better summarize image contents with fewer orthonormal axes than those of Fourier kernels. Lower computation requirements and truncation errors are obtained in comparison with other orthonormal decomposition techniques. In retrieving similar images from database, only few coefficients are required to be evaluated in the computation of Euclidean distance. The retrieval efficiency in terms of computation complexity and speed is very high.

A database consisting of butterfly images collected from existing websites is constructed to demonstrate the power of this approach. The feature discrimination procedure by calculating the Euclidean distance between the corresponding linear coefficients can retrieve shift-, rotation-, and scale-variations of the query image, as specified by the user through the query interface. Contents of the image extracted are embedded in the compressed data that can be easily and efficiently used as indices for content-based image retrieval. Logic predicates, e.g., AND, OR, NOT, or spatial constraints might be further imposed on plural number of sub-regions of a query image to proceed more complicated image retrieval applications.

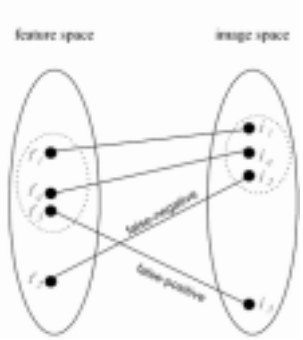
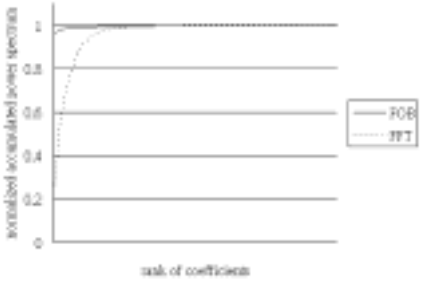


Figure 1. The distance relationship between image points and corresponding feature points is not preserved through most feature extraction process. Image-feature pair $\{i_2, f_2\}$ and $\{i_3, f_3\}$ illustrates the case of false-negative and false-positive, respectively.

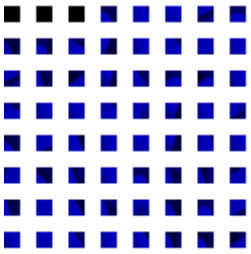


References

1. Gupta, A. and Jain,R. Visual Information Retrieval, Commun. ACM 40, 5 (May 1997), 70-79.
2. Flickner, M., Shawney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Perkovic, D., Steel, D., and Yonker, P. Query by image and video content: The QBIC system, IEEE Comput. 28, 9 (Sept 1995), 23-31.
3. Rui, Y., Huang, T. S., and Chang, S-. F. Image retrieval: Current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation (March 1999).
4. Barnsley, M. F. Fractals Everywhere, Academic Press, San Diego (1988).
5. Jacquin, A. A fractal Theory of Iterated Markov Operators with Applications to Digital Image Coding, PhD thesis, Georgia Institute of Technology (August 1989).
6. Cheung, K.-M. and Shahshahani M. A comparison of the fractal and JPEG algorithms, TDA Progress Report 42-107 (Nov 1991), 21-26.
7. Fisher, Y. Ed. Fractal Image Compression, Theory and Application, Springer-Verlag, New York (1995), 199-214.
8. Vines, G. and Hayes, M. H. Nonlinear Address Maps in a One-Dimensional Fractal Model, IEEE Trans. Signal Processing 41, 4 (April 1993), 1721-1724.

Figure 2. Normalized accumulated ranked power spectrum of the proposed fractal orthonormal basis (FOB) approach and Fourier transform, starting with component with the largest magnitude. Much fewer fractal orthonormal coefficients are needed to constitute the same amount of energy in comparing with those derived by using Fourier kernels.

(a) (b)



(c)

Figure 3. The 64 8×8 fractal orthonormal basis vectors of (a) R, (b) G, and (c) B color components, respectively, derived from an ensemble of 100 butterfly database images. The size of each vector is enlarged by two for ease of observation..

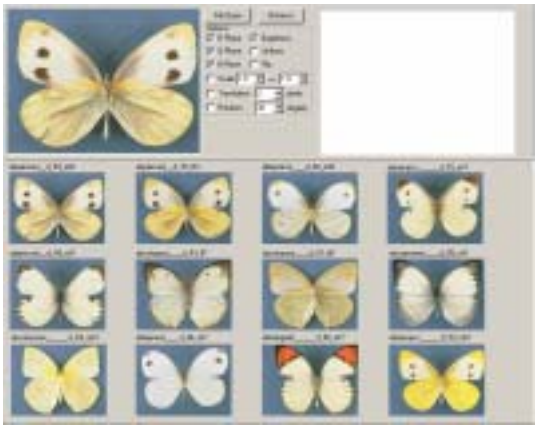
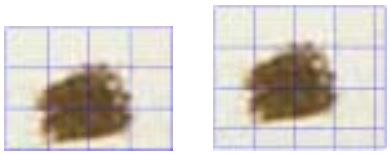


Figure 4. An image retrieval example. The features from R, G, B color components and brightness level of the query image are all selected. The rectangular area in the upper right-hand corner provides an enlarged viewing window for the image retrieved.



(a) (b)

Figure 5. An enlarged view of the selected sub-region of a query image with size that is (a) integral multiples of 8×8, (b) not integral multiples of 8×8. For the case of (b), a mask with the largest integral multiples of 8×8 that can be fit into the sub-region is applied from top-left toward the lower-right corner with an increment of one pixel is applied. On each iteration, the coefficients derived are used as features of the sub-region to compare with those of database images.

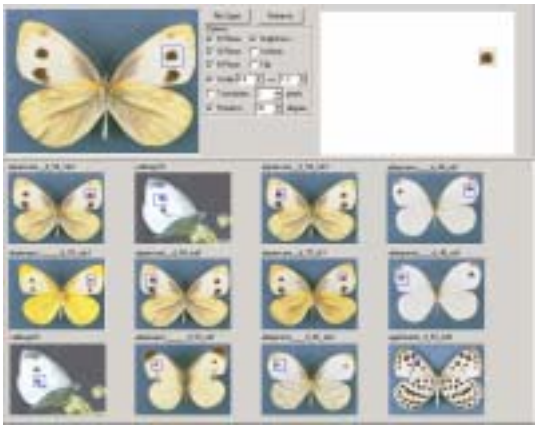


Figure 6. Retrieval results based on a sub-region of a query image with scaling factors 0.8 through 1.2 and rotation angles every 30 degrees.