Efficient histogram for region based image retrieval in the discrete cosine transform domain

Amina Belalia[1], Kamel Belloulata[2], Shiping Zhu[3]

[1]School of Computer Science, RCAM Laboratory, Sidi Bel Abbés, Algeria

[2]Telecommunications Department, RCAM Laboratory, Sidi Bel Abbes University, Sidi Bel Abbés, Algeria

[3]Department of Measurement Control and Information Technology, School of Instrumentation Science and Optoelectronics Engineering, Beihang University, Beijing, China

ABSTRACT

Recently, several approaches of content based-image retrieval (CBIR), based on the characteristics of discrete cosine transform (DCT), such as decorrelation and concentration of energy in only a few coefficients, have been proposed. To reduce the semantic gap between high level search and low level patterns, a new concept based on region based search region-based image retrieval (RBIR) has been proposed. Recently, we proposed to use shape-adaptive (SA) DCT in a new RBIR system. In this paper, we propose an efficient histogram optimization suited to our region-based concept. This histogram takes into account the pattern's from the SA-DCT of the border blocks as well as the DCT coefficients of the internal blocks. Our proposed scheme has greatly improved the results compared to region-based reference methods. Region-based search is limited to the object of interest only, i.e. a region of the query image can only match a region of another image in the database.

This is an open access article under the [CC BY-SA](#) license.

Corresponding Author:

Kamel Belloulata

Department of Telecommunications, Electrical Faculty Engineering, Djillali Liabes University of Sidi Bel Abbes

Campus 3000 palces, Sidi Djillali, Sidi Bel Abbès, 22000, Algeria

Email: kamel.belloulata@univ-sba.dz

1. INTRODUCTION

Image indexing and search systems by content allow us to search for images from a database according to their visual characteristics [1]. These characteristics, also called low-level visual features are color, texture, shape [2], [3]. The semantic gap between low-level features and high-level semantic understanding of images is often hard to bridge [4]. The content based image retrieval (CBIR) systems in the compressed domain use a global descriptor [5]-[8], such as the histogram of the discrete cosine transform (DCT) coefficients [9], [10] or discrete wavelet transform (DWT) [11]-[13], to represent the images. Several methods have been developed. We cannot cite all the existing methods, but, we will try to summarize some works that are related to our approach. They used the histogram to extract information from the characteristic distribution, without having information about the location of its characteristics [14]. These global features are taken from the entire image and often are unable to find the local details in natural images. Therefore, non-similar images which vary in local detail may have a vector of similar characteristics. This situation creates difficulties for the low level feature's representation to have a solid relationship with the semantics of the image.

Region-based search systems attempt to fill gaps in content-based search systems [15]-[17]. A region-based search system applies segmentation [18]-[20] to images to decompose it into regions, which matches

objects if the segmentation is perfect [21]. The object representation is intended to be close to the perception of the human visual system (HVS). Since the search system identifies objects that are in the image, it will be easier for this system to recognize similar objects in different locations and with different orientations and sizes. Region-based search systems include the Netra [22] system, and the Blobworld [23] system. The Netra and Blobworld systems compare images region to region. The reason is to transfer part of the comparison task to the users. To search for an image, a user has the segmented regions of the image and is asked to select the regions which contribute to the search and also the attributes, for example color and texture, of the regions used to assess similarity. To measure the similarity between the images, Datta *et al.* [2], [21] proposed the integrated region matching (IRM) algorithm, which allows to match a region of an image to several regions of another image from the database. In other words, the mapping of regions between two images is a *many-to-many* relationship [24]. Therefore, the similarity between the two images is defined as the weighted sum of the distances, in feature space, between all the regions of the different images. Compared to search systems based on different regions, like Blobworld, the IRM approach decreases the impact of imprecise segmentation [20]. IRM incorporates the properties of all segmented regions so that the information on an image can be fully utilized. To increase the robustness against segmentation errors, IRM compares one region with multiple regions in another image. Each region is assigned an importance weight, which corresponds to the importance of the region. There are several ways to assign weight to a region. Some assume that all regions are equally important. In IRM, important objects in an image tend to occupy larger areas, called a pattern based on percentage of area. Another method called adaptive region matching (ARM), which consists in dealing with the similarity measurement problem in region-based image retrieval has been proposed in [25]. To decrease the negative influence of interfering regions and significant loss of information simultaneously in ARM method, a region importance index (RII) is calculated to find the semantic meaningful region (SMR). In addition, the ARM automatically performs the SMR-to-image search or the image-to-image search depending on whether it has a dominant region or not. On the other hand, the significant region-based image retrieval (SRBIR) model proposed by [26], identifies a region of importance in an image using a mechanism based on visual attention and represents this region using the color descriptors and the descriptors based on the curvelet transform. Recently, the integrated category matching (ICM) was adopted by Meng *et al.* [17] for similarity measurement combined with a centroid-based significance index method to match all merged regions in images according to their importance. For feature extraction in ICM method, a regional convolution mapping feature (RCMF) based on the convolutional neural networks (CNN) [27] was used. RCMF is further combined with the number and distribution of regions to represent the characteristics of merged regions. They employ the VGGNet19 model pretrained on ImageNet, which has smaller convolution kernels and a larger network depth.

In recent years the technology of handling images and video has undergone significant change and great majority of content is nowadays handled in *compressed* form. Lossy compression based on quantized block of DCT is a proven, highly efficient technique used in major compression standards (JPEG, MPEG1/2/4, H26X, VVC). This technique enables reduction of the size of the content to a small fraction of the original size while preserving well its perceptual quality. Since perceptual quality is also of major importance for pattern recognition, one can conclude that compression may provide interesting perspective for studying recognition problems. In this direction, Defee and Zhong [28] have proposed an approach to unify statistical and structural DCT information's image search. They have proposed a description of structural information into patterns by splitting the patterns into *square regions*. To make it possible to extract the characteristics of the regions arbitrarily formed for the region-based image retrieval (RBIR) systems, Liu *et al.* [29] have used a projections onto convex sets (POCS) in the wavelet domain. Recently, we have proposed a new region based image retrieval in the DCT domain by using a shape-adaptive DCT [30]. In this paper, we propose to optimize our scheme by using an efficient and adapted histogram which discriminates the border blocks patterns from the interior blocks patterns. Experimental results on two public Corel-1000 and Caltech-256 datasets show that the proposed method (PM) is very efficient for exploring the local similarity than the existing visual similarity approaches.

This paper is organized as follows: the section 2 will detail the principle of the ACs and DCs histogram patterns construction from the DCT and SA-DCT coefficients. In section 3, we describe several distances to measure the similarity between two histograms. Different criteria for the evaluation of our system will be developed in the section 4. One of the problems of an image-by-object search system is how to compare between two images, which will be developed in section 5. The detailed description of our RBIR system will be presented in the sections 6 and 7. In section 8, we will present the experimental results and analyze them. A

conclusion will be drawn in the section 9.

2. IMAGE DESCRIPTORS IN THE DCT DOMAIN

Suppose that I is the intensity of the image we want to transform and $\mathbf{x} = (x, y)$ are the spatial coordinates for the pixels of this image. Suppose $\{b_1, ..., b_N\}$ is the set of adjacent blocks, i.e., the collection of pixel coordinates, which divide the image. To allow compact notation, assuming that:

$$I_{b_i} = \{I(x) : x \in b_i\}. \tag{1}$$

Each block b_i is transformed into the frequency domain $\widehat{I_{b_i}} = DCT(I_{b_i})$ by the DCT transformation. $U = (u, v)$ are the frequency coordinates of the DCT coefficients and $\widehat{I(u)}$ is the DCT transform for an image N x N represented by the pixels $I(x)$ such that $x, y = 1, ..., N$.

2.1. Shape-adaptive transform

In our approach, we need a background extraction for foreground detection [31]. After that, to extract the feature's regions, we propose to apply shape-adaptive DCT (SA-DCT) [32] to each segment $\mathcal{S}$ of the boundary blocks Figure 1(a) of the region of interest and the classical DCT to the interior blocks. The basic concept of the SA-DCT is to perform vertical 1-D DCTs on the active pixels first Figure 1(b), and then to apply horizontal DCTs to the vertical DCT coefficients with the same frequency index Figure 1(c). The most important benefit of SA-DCT is its capability to adapt to arbitrarily-shaped regions; the method falls back to standard DCT on rectangular image blocks.
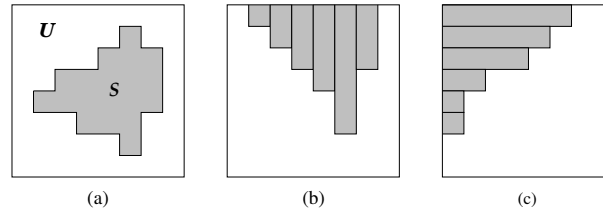

Figure 1. Illustration of SA-DCT for (a) arbitrarily-shaped region, (b) vertical alignment followed by vertical 1-D DCTs, and (c) horizontal alignment followed by horizontal 1-D DCTs [33]

Recall that $\widehat{I}_{b_i}$ are DCT-transformed blocks of intensities I_{b_i}. Let $\mathcal{P}_{b_i}^n$ be n-th the segment $\mathcal{S}_{b_i}^n$ after SA-DCT. Note that the shape of $\mathcal{P}_{b_i}^n$ is different from that of $\mathcal{S}_{b_i}^n$ due to the executed vertical and horizontal shifts, but that the number of pixels is unchanged. Also, let $\widehat{I}_{b_i}^n(\mathbf{u})$ be an SA-DCT coefficient in $\mathcal{P}_{b_i}^n$ at frequency $\mathbf{u}$. To construct the AC-pattern as shown in subsection 2.2, we will select at most 9 coefficients in each extrapolated segment $\widetilde{I}_{b_i}^n$. Where $\widetilde{I}_{b_i}^n$ is an extrapolated n-th segment of the SA-DCT-transformed block intensity:

$$\widetilde{I}_{b_i}^n(\mathbf{u}) = \begin{cases} \widehat{I}_{b_i}^n(\mathbf{u}) & \text{if } \mathbf{u} \in \mathcal{P}_{b_i}^n \\ \mathbf{v} & \text{otherwise.} \end{cases} \tag{2}$$

2.2. Histogram of the high frequency components ACs

In this study, we consider DCT blocks 4x4. This method takes into account 9 AC coefficients Figure 2(a) among the 15 coefficients used by [9]. Then, we use the statistical information to build the AC-pattern Figures 2(b) and 2(c). Finally, the statistical of 3 groups (horizontal: C_1, C_2, C_3, vertical: C_4, C_8, C_{12} and diagonal: C_5, C_{10}, C_{15}) is used to build the AC-pattern Figure 2(d). This selection is retained because it is able to represent the internal structure of the contents of the block and reduces the complexity of the feature vector [30]. To build the histogram H_{AC} of AC-patterns for an image, we calculate the number of appearances of this AC-pattern in this image.
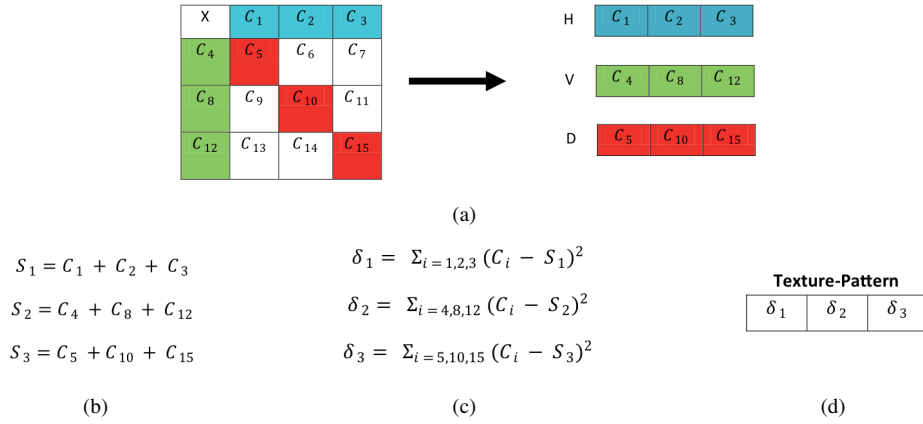
(a)

$$S_1 = C_1 + C_2 + C_3$$
$$S_2 = C_4 + C_8 + C_{12}$$
$$S_3 = C_5 + C_{10} + C_{15}$$

$$\delta_1 = \Sigma_{i=1,2,3} (C_i - S_1)^2$$
$$\delta_2 = \Sigma_{i=4,8,12} (C_i - S_2)^2$$
$$\delta_3 = \Sigma_{i=5,10,15} (C_i - S_3)^2$$

Texture-Pattern

δ_1	δ_2	δ_3

(b)   (c)   (d)

Figure 2. AC-pattern construction process of (a) three groups of AC coefficients extracted from DCT block, (b) sums of each group, (c) sums of squared-differences, and (d) texture-pattern

2.3. Histogram of the low frequency components DCs

The DC-pattern describes the global characteristics using the gradient between each block and its neighbors (Inter-block). DC-DirecVec [9] is defined and used as a characteristic for the DC-pattern. More precisely, DC-pattern is defined as a set of directions having the greatest differences between the DC value of the current block and the DC values of neighboring blocks. The absolute values of these differences are arranged in descending order and the first γ directions with the largest differences form the DC-pattern. Also, the number of appearances of DC-pattern gives us the histogram H_{DC}.

2.4. Feature descriptor

For each block, the AC-pattern is formed by 9 coefficients and the DC-pattern is built by the DC coefficient of the block itself and the difference between this value and its 8 other neighboring DCs. The concatenated histogram of the two histograms, H_{AC} and H_{DC}, represents the index of the dinosaur image Figure 3(a) and horse image Figure 3(b) and it is used to do the search. In this context, the descriptor is defined as (3):

$$H = [(1 - \alpha) \times H_{AC}, \alpha \times H_{DC}], \tag{3}$$

where α is the weight that controls the impact of the AC-pattern and DC-pattern histograms. This parameter can be set to improve search accuracy.
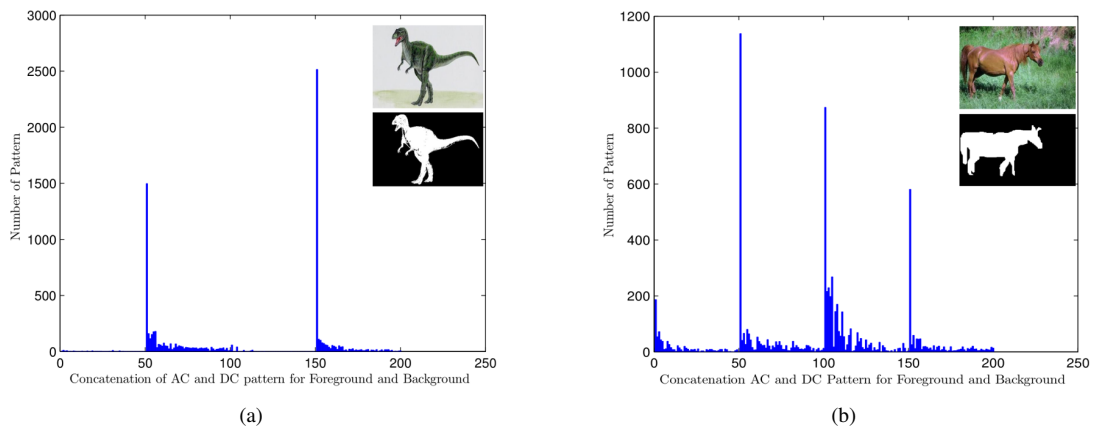

(a)   (b)

Figure 3. Histograms of the AC-patterns and Dc-patterns for the RBIR system for (a) combined AC-DC patterns for the dinosaur and (b) combined AC-DC patterns for the horse

3. SIMILARITY MEASURE

The feature vector can be seen as a normalized vector. This allows comparison of statistical information between images by calculating the distance between histograms of feature patterns [34]. In this section, we will detail the measures of similarities between the query histogram H_Q and the database histograms H_D. This similarity compares *coefficient to coefficient* of the histograms with the same index, i.e., they compare $H_Q(i)$ and $H_D(i)$ for i, but not $H_Q(i)$ and $H_D(j)$ for i $\neq$ j.

3.1. Coefficient to coefficient similarity measure

This measurement category compares the corresponding coefficients of the histograms H_Q and H_D. The similarity between two histograms is the combination of these coefficient-to-coefficient comparisons.

3.1.1. Minkowski distance

The Minkowski distance is defined as (4):

$$d_{L_p}(H_Q, H_D) = \left(\sum_{i=1}^{N} |H_Q(i) - H_D(i)|^p \right)^{\frac{1}{p}},$$ (4)

where p = 1, 2 or ∞, this distance is always referred by the Manhattan distance, the Euclidean distance or the Chebyshev distance respectively. These three distances are the most used in image search.

3.1.2. Histogram intersection

The histogram intersection is defined as (5):

$$d_{\bigcap}(H_Q, H_D) = \sum_{i=1}^{N} min(H_Q(i), H_D(i)),$$ (5)

Note that this distance is less complex because it consists only of minimums and additions.

3.1.3. Distance chi-square

Chi-square distance (χ^2) is used to compare two sets of data and to determine if they are taken from the same distribution function.

$$d_{\chi^2}(H_Q, H_D) = \sum_{i=1}^{N} \frac{(H_Q(i) - H_D(i))^2}{H_Q(i) + H_D(i)},$$ (6)

where i represents the component of the descriptor and H_Q, H_D represent the different descriptors, N is the dimension of the descriptor. Note that in our system, the similarity between the request image and an image in the database is estimated by the chi-squared distance (6).

4. PERFORMANCE EVALUATION

To evaluate and compare the performance of different indexing and image search systems, an evaluation of their performance is necessary. This allows researchers to fully understand the limitations of their algorithms and to compare their results with other objective algorithms. In this section, we discuss some performance evaluation measures for indexing and image search systems.

4.1. Precision and recall

The most commonly used performance measure for CBIR and RBIR is the precision-recall curve. The precision measures the capacity of the system to provide a maximum of relevant images on the set of images it provides and is defined as the ratio between the number of relevant images on the r first images and the number of images (r). The recall corresponds to the ability of a system to find the relevant images from the database in relation to a query and it is defined as the ratio between the number of relevant images among the r images and the total number of relevant images in the database. The system which gives better precision for the same recall is the most efficient system.

$$Precision = \frac{r}{r + s} \qquad (7)$$

$$Recall = \frac{r}{r + t} \qquad (8)$$

where r is the number of relevant images retrieved, s is the number of irrelevant images retrieved, t is the number of irrelevant images not retrieved from the database.

4.2. Mean average precision (MAP)

Images are considered to be similar if the distance (6) between their feature descriptors is less than a threshold. The performance of the systems can be evaluated using the precision-recall curves as shown in subsection 4.1. MAP is the way to turn the precision-recall graph into a single value. The MAP, for all requests, is defined by (9):

$$MAP = \frac{1}{Q} \sum_{q \in Q} AP(q), \qquad (9)$$

where Q is the set of requests and, $AP(q)$ is defined as (10):

$$AP(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P(R_n), \qquad (10)$$

where R_n is the recall after the relevant n^{th} image is selected. $P(R_n)$ is the precision when the recall is R_n. MAP contains, in addition to precision and recall, the position of the relevant image.

4.3. Average retrieval rate (ARR)

The retrieval rate (RR) for a query is defined as the percentage of the number of relevant images retrieved out of the total number of relevant images in the database, observed in the first K images retrieved:

$$RR = reminder = \frac{r}{r + t} \qquad (11)$$

ARR is defined as the average value of the set of retrieval rates (RR) of the first K images found on each request.

5. SIMILARITY MEASURE IN REGION-BASED SYSTEM

One of the problems of an image-by-object search system is how to compare between two images, that is to say the definition of similarity measure between the images. A simple solution adopted by the early systems [22], [23] is the use of the individual similarity measure region to region. To use such a system, the user is expected to select one or more regions from the query image to do the search. As discussed in [21], due to the uncontrollable nature of images, extracting objects from images automatically and precisely is still beyond the state of the art of new computer vision techniques [20], [35]. However, other systems tend to partition an object into multiple regions, none of which is representative of the semantic object [17]. Therefore, it is often difficult for users to determine which regions should be used for search. To provide users with a simpler interface and reduce inaccurate segmentation, image to image similarity measures that combine the properties of all regions have been proposed in [21], [22]. These systems only require users to choose the query, and therefore free them from enigmatic decisions about regions. For example, the SIMPLYcity system [21] uses integrated region matching (IRM) as a measure of similarity which is based on the region percentage to decide the region's importance. By allowing a many to many relationship between regions [24], the approach is robust against inaccurate segmentation. In our proposed system, two similarity measures are present: by region and by image, depending on whether the search is by region or by complete image.

5.1. Image to image similarity measure

Based on the assumption that any region could be useful when evaluating similarity [21], all regions of each image are considered. First of all, it should be noted that the similarity calculation, adopted by our system, between a region of a query image and another region from the database image is the χ^2 (6). The calculation of similarity between the images is based on the calculation of all the similarities between the regions after segmentation and then taking the sum of the minimums between the distances of the regions. The Table 1 and Figure 4 show an example of the process to calculate the similarity between two images. After the construction of the Table 1, we calculate the sum of the values of the last row to obtain the similarity between two images.

Table 1. Similarity between regions

B/R	Foreground	Background
Foreground	0.2	0.7
Background	0.3	0.5
textbfMin	0.2	0.5
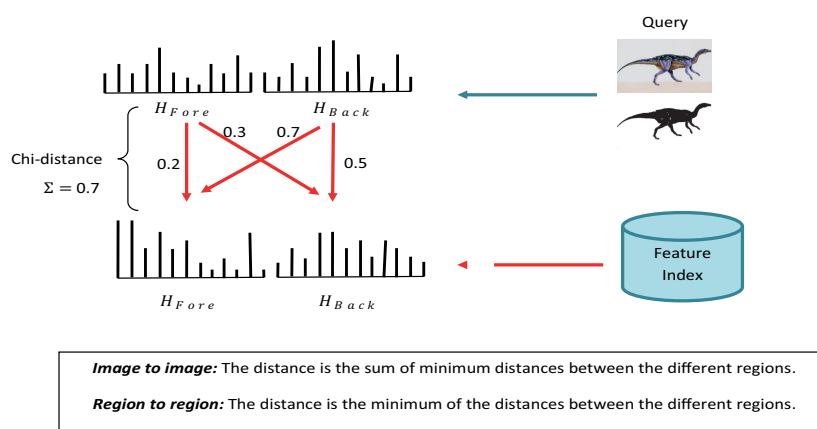
Figure 4. Similarity measure between images

5.2. Region to region similarity measure

The second possibility adopted by our system is the similarity computation *region to region*. The user selects a region from the two possible regions (foreground or background) from the query image to perform the search. The region to region similarity measure Figure 4 consists in calculating the similarity between the selected region and the regions of the images in the database then taking the minimum of the distances between the regions.

6. RESEARCH METHOD

In the proposed system Figure 5, after image segmentation [20], there are two types of parameters which can be adjusted. The parameters dedicated to indexing the images (regions) as shown in subsection 6.1 and the parameters dedicated to the retrieval as shown in subsection 6.2.

6.1. Indexing parameters

There are several indexing parameters used in the proposed system. These parameters must be optimized to improve the performance of the system.

− QP_{AC} and QP_{DC}: Quantization parameters for AC and DC respectively

If these parameters are small enough, the total number of different AC-patterns and DC-patterns becomes larger, this makes the histogram generation process more complicated and time consuming. On the other hand, if the parameters QP_{AC} and QP_{DC} are also large, then the rightmost AC-patterns and DC-patterns coefficients tend towards zero, this decreases the search performance. Therefore, a compromise must be found for the choice of these parameters between performance and consumption of time, see subsection 2.2.

− γ: first directions with the biggest DC-Pattern differences

This parameter can be adjusted to have better search performance, see subsection 2.3.

− ACBins and DCBins: Number of AC bins and DC bins needed to build the characteristic histogram

Histograms are constructed by the ordering of feature occurrences. The resulting histogram has bins ordered in descending order. The size of this histogram is a free parameter that can be adjusted to improve search performance. Eliminating a few bins that represent irrelevant characteristics improves search performance. On the other hand, the elimination of multiple bins result in degraded performance. So you have to choose the optimal length which induces better research.

− NC: The number of AC coefficients entering into the construction of the AC-pattern

NC represents the number of AC coefficients that form the diagonal, horizontal and vertical groups seen in subsection 2.2. This parameter can take the value $NC = 2$ or $NC = 3$.

The Figure 5 illustrates the learning process allowing to optimize the indexing parameters mentioned above.
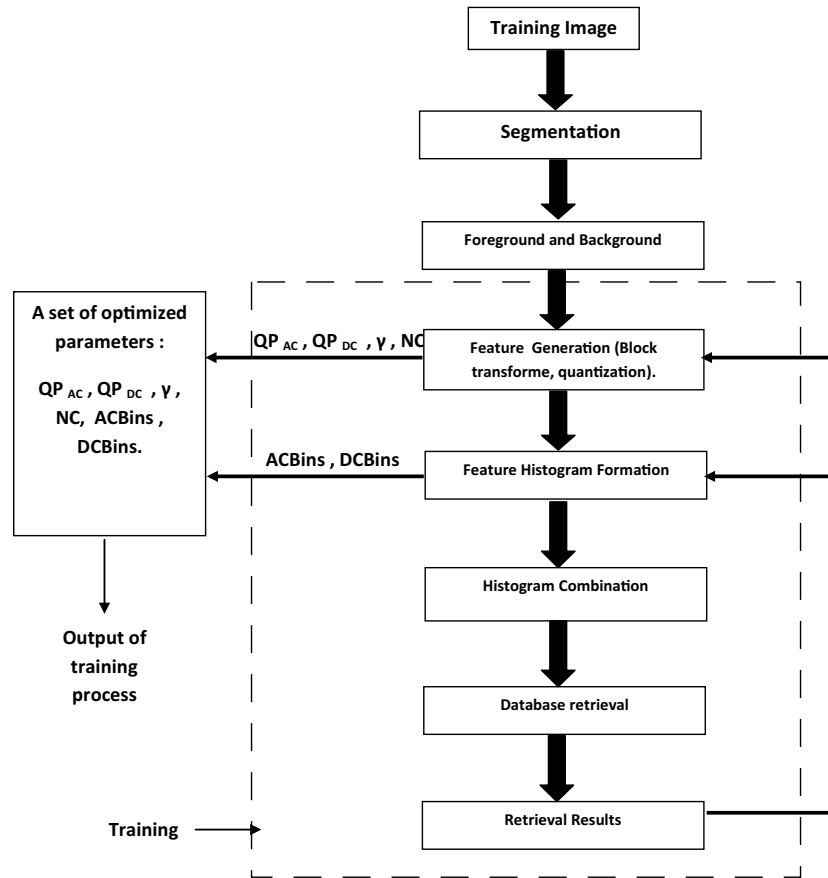
Figure 5. Learning process optimizing the indexing parameters QP_{AC}, QP_{DC}, γ, NC, $ACBins$, $DCBins$

6.2. Search parameters

There are several search parameters used in the proposed system. These parameters must also be optimized to improve the performance of the system.

− β: a weight that combines the foreground histogram with that of the background.

The β parameter can be adjusted to improve the performance of the proposed system.

− α_{Fore}: a weight that combines the AC-pattern histogram with the DC-pattern histogram, for the foreground. The (12) represents the combined histogram of the foreground H_{Fore}; where α_{Fore} is a weight parameter that can be adjusted to improve the quality of the search.

$$H_{Fore} = [(1 - \alpha_{Fore}) \times H_{AC_{Fore}}, \alpha_{Fore} \times H_{DC_{Fore}}], \tag{12}$$

− α_{Back}: a weight that combines the AC-pattern histogram with the DC-pattern histogram, for the background.

The (13) represents the combined histogram of the background H_{Back}; where α_{Back} is a weight parameter that can be, also, adjusted to improve the quality of the search.

$$H_{Back} = [(1 - \alpha_{Back}) \times H_{AC_{Back}}, \alpha_{Back} \times H_{DC_{Back}}], \tag{13}$$

Demonstrations: The substitution of (12) and (13) in:

$$H_{image} = [(1 - \beta) \times H_{Fore}, \beta \times H_{Back}], \tag{14}$$

gives the following (15):

$$H_{image} = [(1 - \alpha_{Fore} - \beta + \alpha_{Fore}\beta) \times H_{AC_{Fore}}, (\alpha_{Fore} - \alpha_{Fore}\beta)H_{DC_{Fore}},$$
$$(\beta - \beta\alpha_{Back}) \times H_{AC_{Back}}, \beta\alpha_{Back} \times H_{DC_{Back}}] \tag{15}$$

Several cases may appear:
1. If $\alpha_{Fore} = 1$ and $\alpha_{Back} = 1 \Rightarrow H_{Full} = [(1 - \beta) \times H_{DC_{Fore}}, \beta H_{DC_{Back}}]$. We will have a search with the DC in the foreground and the background.
   If $\beta = 1 \Rightarrow H_{Full} = H_{DC_{Back}}$, a search with the background DC.
   If $\beta = 0 \Rightarrow H_{Full} = H_{DC_{Fore}}$, a search with the foreground DC.
2. If $\alpha_{Fore} = 0$ and $\alpha_{Back} = 0 \Rightarrow H_{Full} = [(1 - \beta) \times H_{AC_{Fore}}\beta H_{AC_{Back}}]$. We will have a search with the AC in the foreground and the background.
   If $\beta = 1 \Rightarrow H_{Full} = H_{AC_{Back}}$, a search with the background AC.
   If $\beta = 0 \Rightarrow H_{Full} = H_{AC_{Fore}}$, a search with the foreground AC.
3. If $\alpha_{Fore} = \alpha_{Back} = \beta = \frac{1}{2} \Rightarrow H_{Full} = [\frac{1}{4}H_{AC_{Fore}}, \frac{1}{4}H_{DC_{Fore}}, \frac{1}{4}H_{AC_{Back}}, \frac{1}{4}H_{DC_{Back}}]$

The Table 2 shows an example of the combination of the parameters $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$ when these are binary values. From the results found, we notice that there are combinations of $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$ which give the same H_{Full}. These are the combinations 0 and 2, 1 and 5, 3 and 7 and finally 4 and 6.

Table 2. H_{Full} for binary combinations of parameters $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$

Combination	α_{Fore}	α_{Back}	β	H_{Full}
0	0	0	0	$H_{AC_{Fore}}$
1	0	0	1	$H_{AC_{Back}}$
2	0	1	0	$H_{AC_{Fore}}$
3	0	1	1	$H_{DC_{Back}}$
4	1	0	0	$H_{DC_{Fore}}$
5	1	0	1	$H_{AC_{Back}}$
6	1	1	0	$H_{DC_{Fore}}$
7	1	1	1	$H_{DC_{Back}}$

7. HISTOGRAM OPTIMIZATION FOR THE BORDER BLOCKS

In this section, we present a way to optimize our search and the content of the full histogram. It consists in separating the AC-patterns and DC-patterns at the border level of the object. So the construction process of the combined histogram by the proposed method is shown in Figure 6. Figure 6(a) presents the combined histogram taking into account the parameters $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$ and $\{ \gamma_{CF}, \gamma_{FF}, \gamma_{CB}, \gamma_{FB} \}$ and Figure 6(b) presents the Illustrative diagram showing an image composed of two objects, foreground and background. The foreground histogram, HFore, is combined from the histogram from the interior (full) blocks, HFF, and the histogram from the border (contour) block segments, HCF, from the foreground. While the background histogram, HBack, is combined from the histogram from the interior (full) blocks, HFB, and the histogram from the border (contour) block segments, HCB, from the background. The overall combined histogram is calculated from the foreground, HFore, and background, HBack histograms. The following equations illustrate the use of these parameters for the construction of the image descriptor.
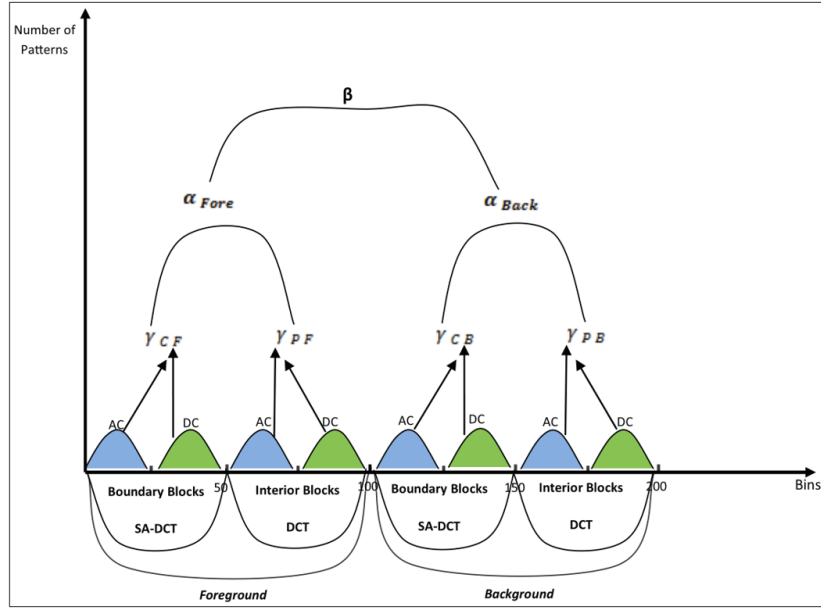
(a)
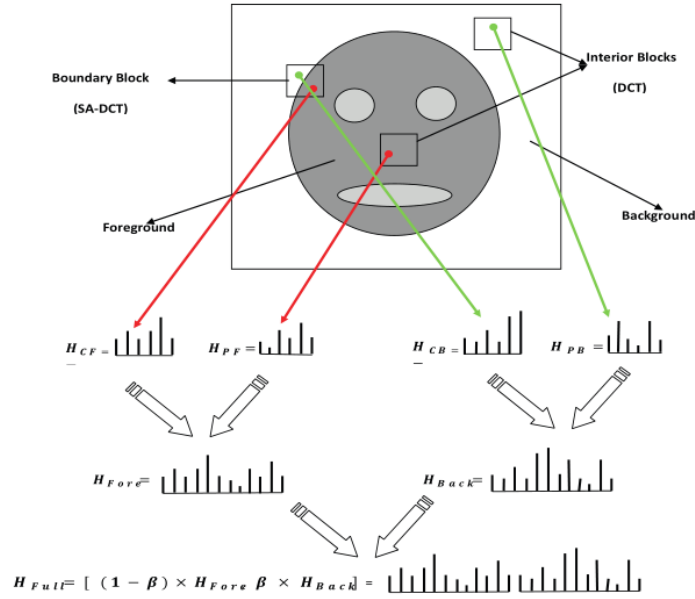
(b)

Figure 6. Process of construction of the combined histogram by the proposed method (a) combined histogram taking into account the proposed parameters and (b) illustrative diagram showing an image composed of twos objects, foreground and background

− For the foreground
  (a) Foreground border (contour) blocks

$$H_{CF} = [(1 - \gamma_{CF}) \times H_{AC_{CF}}, \gamma_{CF} \times H_{DC_{CF}}], \tag{16}$$

H_{CF}: Combined histogram of patterns (AC-DC) from the segments of the border blocks.

$H_{AC_{CF}}$: AC-pattern histogram from the segments of the border blocks.

$H_{DC_{CF}}$: DC-pattern histogram resulting from the segments of the border blocks.

γ_{CF}: A weight to combine H_{CF}.

(b) Foreground interior (full) blocks

$$H_{FF} = [(1 - \gamma_{FF}) \times H_{AC_{FF}}, \gamma_{FF} \times H_{DC_{FF}}], \tag{17}$$

H_{FF}: Combined histogram of patterns (AC-DC) from interior (full) blocks.

$H_{AC_{FF}}$: AC-pattern histogram from the interior (full) blocks.

$H_{DC_{FF}}$: DC-pattern histogram from interior (full) blocks.

γ_{FF}: A weight to combine H_{FF}.

(c) The foreground blocks (Interior + Border)

$$H_{Fore} = [(1 - \alpha_{Fore}) \times H_{FF}, \alpha_{Fore} \times H_{CF}], \tag{18}$$

− For the background

    (a) Background border (contour) blocks

$$H_{CB} = [(1 - \gamma_{CB}) \times H_{AC_{CB}}, \gamma_{CB} \times H_{DC_{CB}}], \tag{19}$$

H_{CB}: Combined histogram of patterns (AC-DC) from the segments of the border blocks.

$H_{AC_{CB}}$: AC-pattern histogram resulting from the segments of the border blocks.

$H_{DC_{CB}}$: DC-pattern histogram resulting from the segments of the border blocks.

γ_{CB}: A weight to combine H_{CB}.

    (b) Background interior (full) blocks

$$H_{FB} = [(1 - \gamma_{FB}) \times H_{AC_{FB}}, \gamma_{FB} \times H_{DC_{FB}}], \tag{20}$$

H_{FB}: Combined pattern histogram (AC-DC) from interior blocks.

$H_{AC_{FB}}$: AC-pattern histogram from the interior (full) blocks.

$H_{DC_{FB}}$: DC-pattern histogram from interior (full) blocks.

γ_{FB}: A weight to combine H_{FB}.

    (c) The background blocks (interior + border)

$$H_{Back} = [(1 - \alpha_{Back}) \times H_{FB}, \alpha_{Back} \times H_{CB}], \tag{21}$$

− For the full image

$$H_{image} = [(1 - \beta) \times H_{Fore}, \beta \times H_{Back}]. \tag{22}$$

The global descriptor can be obtained by combining the foreground and background histograms. The Figures 7(a) and 7(b) represent the descriptors of the flower and horse images respectively, calculated by the method mentioned above.

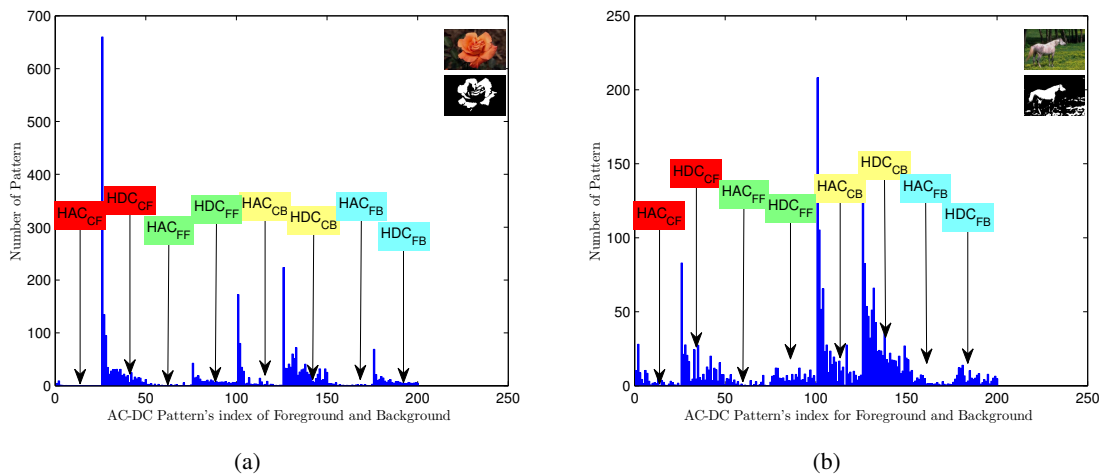(a)   (b)

Figure 7. Combined histograms of the first high frequency occurrences of AC-Patterns and DC-Patterns for RBIR principle of optimization: (a) global descriptor of the flower image and (b) global descriptor of horse image

8.   RESULT AND DISCUSSION

8.1.   Standard benchmark data-sets

To show the efficiency of the proposed systems, several tests were carried out on two data-sets of image: Corel-1000 and Caltech-256. The Corel-1000 database was collected by Wang *et al.* [36], contains 1000 256 x 384 or 384 x 256 images, these images are classified into 10 semantic categories: Africans, buildings, beach, buses, dinosaurs, elephant, flowers, horse, mountains and foods. The Caltech-256 database was collected by Griffin *et al.* [37], contains 30,607 images from 256 categories, 80 to 827 images per category. In our experiments, we randomly selected 10 categories from Caltech-256, which contain 1,299 images. The 10 categories are: AK-47, American-flags, backpacks, baseball-bats, baseball-gloves, basketball hoops, bats, bathtub, beer-mug and blimp.

8.2.   The average precision-recall results

Figure 8(a) presents the improvement of the average precision-recall results, on the Corel-1000 database, for the proposed approach (SA-DCT + histogram optimization) over the region-based (SA-DCT) [30] and the conventional content-based approach (DCT) [9] approaches. The same curves, for Caltech-256, are shown in Figure 8(b). The best performance is obtained by using the global combined and optimized descriptor (22) of the foreground and the background together. The proposed system attempts to overcome the limitation of global-based retrieval (CBIR/RBIR with DCT/SA-DCT) systems by emphasizing the target objects only and minimizing the influence of background.

8.3.   The mean average precision results (MAP)

We have made a comparison between our proposed method (PM) and many benchmark existing RBIR methods like: regional convolution mapping feature with integrated category matching (RCMF+ICM) [17], adaptive region matching (ARM) [25], SIMPLIcity [21], integrated region matching (IRM) [25], MN-MIN [23], Shape-adaptive-DCT [30]. We also compare our approach with CBIR approaches like DCT+SVD [10] and DCT [9]. The specific settings of each method can be found in the related references.

From Tables 3, we can see that our Proposed Method (PM in bold) performs better than all others 8 methods on 10 categories of Corel-1000 Figure 9(a) and Caltech-256 Figure 9(b) databases. Our approach outperforms all the others methods (RBIR and CBIR). According to the Table 3, we notice that in general the proposed RBIR system works better than other existing RBIR systems, except for four classes from the Corel-1000 database, this is due to poor segmentation of images from these classes which have not a clear semantic object also. It can be concluded from the experimental results that the proposed system improves performance

on Corel-1000 and Caltech-256 databases. In addition, a minimum number of AC coefficients and a small number of histogram coefficients were used to reduce computation time.

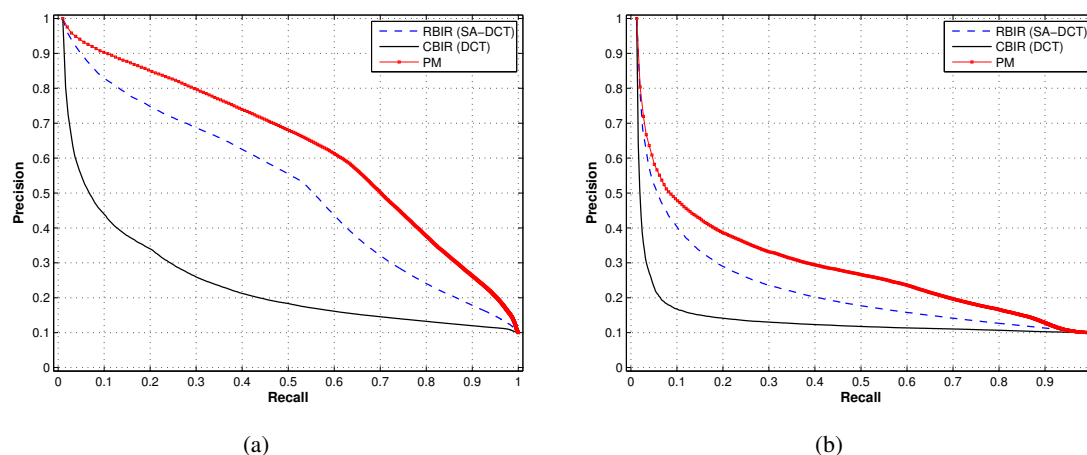| (a) | (b) |

Figure 8. The average precision-recall results: (a) Corel-1000 database and (b) Caltech-256 database, between the proposed method (PM), the region-based approach [30] and the content-based approach [9]

Table 3. The mean average precision (MAP) comparison between different methods on the Corel-1000 database

| Category	African	Beach	Buildings	Buses	Dinosaurs	Elephants	Flowers	Horses	Mountains	Foods	Average
PM	79.45	65.36	**78.94**	**99.22**	**100**	**92.60**	97.17	**96.28**	63.87	**88.99**	**86.19**
SA-DCT [30]	75.45	52.50	76.30	97.25	**100**	90.20	95.50	94.25	61.85	78.95	82.18
RCMF + ICM [17]	66.3	52.7	70.6	99.1	**100**	90.8	**99.2**	95.1	71.5	85.3	83.06
DCT + SVD [10]	**88.98**	**74.85**	75.92	78.75	99.90	91.89	83.78	95.65	**73.98**	78.97	84.27
MN-ARM [25]	74.23	46.35	74.51	80.60	99.80	62.42	94.84	92.91	61.80	78.88	76.63
MN-IRM [38]	73.80	42.86	74.32	74.24	99.80	61.00	95.23	93.49	51.33	66.54	73.26
SIMPLIcity [21]	52.73	25.52	56.54	49.40	98.50	45.51	78.73	78.28	39.00	39.13	56.33
MN-MIN [23]	43.40	29.43	31.82	46.63	90.11	33.07	64.94	63.62	27.90	35.21	46.61
DCT [9]	52.45	27.20	27.10	36.95	93.05	46.90	53.45	61.65	17.10	37.4	45.33

8.4.   Semantic retrieval

As stated previously, the proposed system allows a similarity measurement region to region or image to image according to the content of the query image. This helps reduce the impact of negative interference regions and the loss of important information. In what follows, we will evaluate the system with optimized global parameters by considering the previous aspects.

8.4.1. Decreasing negative influence of interference regions

Since the image to image similarity measure takes into account the properties of all the regions, this requires interfering regions. However, the region to region similarity measure can solve this problem. Considering the beach and building categories of the Corel-1000 database which have interfering regions. The Figures 10(a) and 10(b) give a comparison between the results of our system with optimized parameters depending on whether the search is region to region or image to image on the beach and building categories respectively. The comparison in the Figure 10 shows that by searching for the most similar images (average retrieval rate P(N), (11)), the similarity measure region to region have a consistent superiority over image to image.
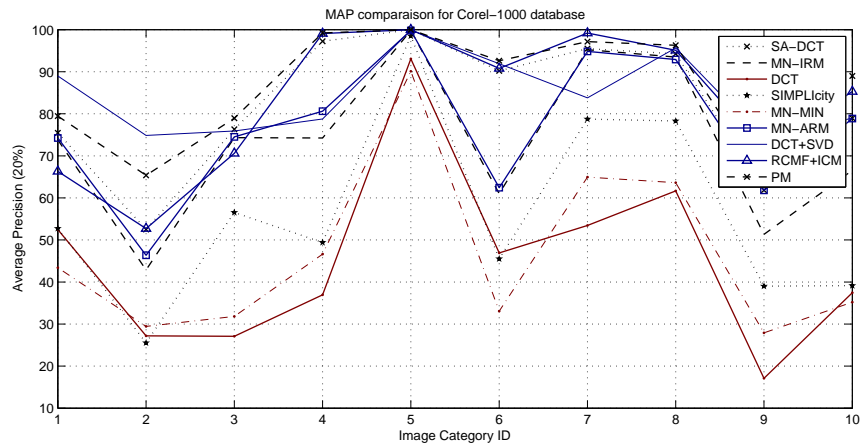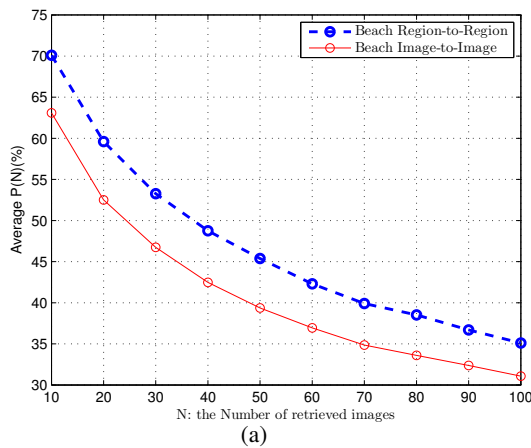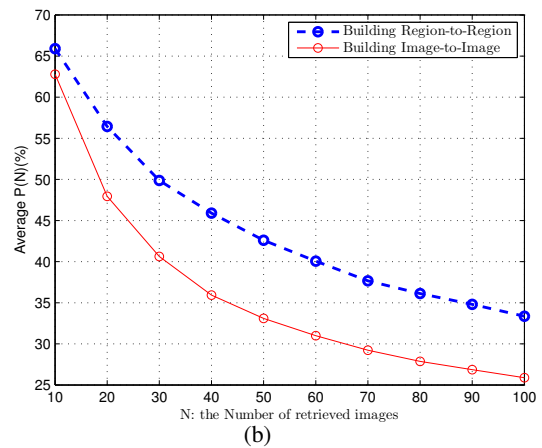
(a)
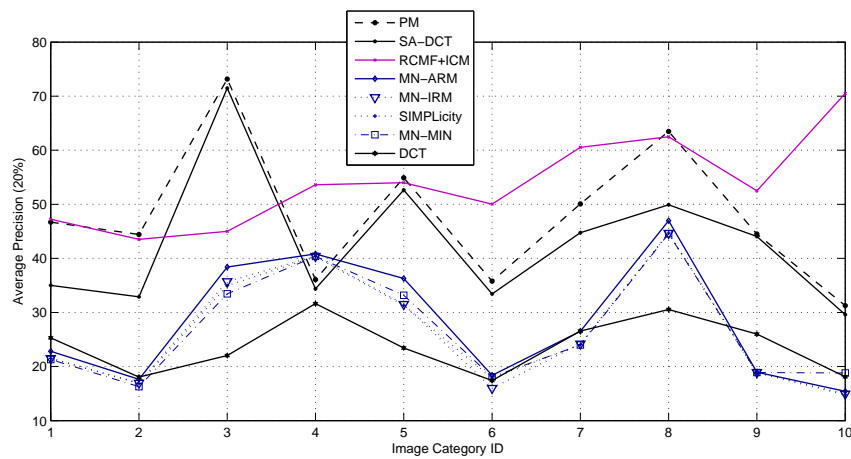
(b)

Figure 9. Mean average precision (MAP) for (a) Corel-1000 database images and (b) Caltech-256 database images

Figure 10. The average retrieval rate P(N) of the region-to-region and image-to-image search: (a) beach and (b) building from Corel-1000 database

8.4.2. The negative influence of important information loss

For images without a main region, if we adopt the similarity measure region to region to search for relevant images, this will result in unsatisfactory search results due to the significant loss of information. Considering the African and food categories from the Corel-1000 database which are images without main object, we will evaluate our system on both categories. In the same vein, the Figures 11(a) and 11(b) give a comparison between the results of our system with optimized parameters depending on whether the search is region to region or image to image on the African and food categories respectively. The comparison results are denoted by: African (region-to-region), African (image-to-image), food (region-to-region), and food (image-to-image). The comparison in the Figure 11 shows that by finding the most similar images (average retrieval rate P(N), (11)), the similarity measure image to image has considerable superiority over region to region.

8.5. Some retrieval examples

Moreover, the retrieval accuracy for Corel-1000 and Caltech-256 dataset are shown for different categories of images as illustrated in Figures 12(a) and 12(b) (see in appendix). For these two examples, we have proved that we can remove the irrelevant images by using our optimization histogram approach. For these two examples, we have proved that we can remove the irrelevant images by using our optimization histogram approach.
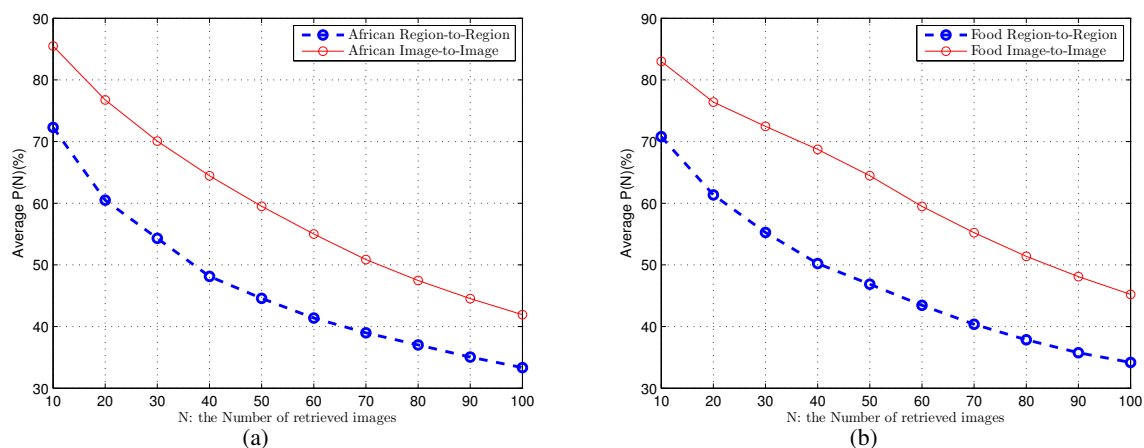

Figure 11. The average precision P(N) of region-to-region and image-to-image search: (a) African and (b) food, from Corel-1000 database

9.   CONCLUSION

Firstly, the concept of RBIR was introduced. Second, the principle of the proposed method based on SA-DCT has been detailed. Finally, we mentioned the experimental results. The proposed method is based on the DCT + SA-DCT coefficients histogram depending on whether the parameters $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$ are optimized. An optimization method has been proposed to improve search performance. It consists in separating the histogram of the object (foreground and background), into a histogram resulting from the patterns (AC and DC) of the interior blocks and a histogram resulting from the patterns (AC and DC) of the border blocks segments by using other parameters $\{ \gamma_{CF}, \gamma_{FF}, \gamma_{CB}, \gamma_{FB} \}$. The proposed system was able to correct the defects of the classic system (CBIR with DCT only). Two types of similarity measure are adopted by our RBIR system, *image to image* and *region to region*. The experimental results show that these two measures make it possible to reduce the negative influence of the interfering regions and reduce the harmful influence of the significant loss of information. Experimental results on Corel-1000 and Caltech-256 databases show that the proposed method is more efficient than some well established RBIR methods. Like a perspective, we'll use the deep learning (CNN) to optimize, at the first stage: a) the indexation parameters: QP_{AC}, QP_{DC}, γ, ACBins, DCBins and NC; at the second stage: b) the search parameters: $\{ \alpha_{Fore}, \alpha_{Back}, \beta \}$, and the histogram optimization parameters: $\{ \gamma_{CF}, \gamma_{FF}, \gamma_{CB}, \gamma_{FB} \}$.

APPENDIX
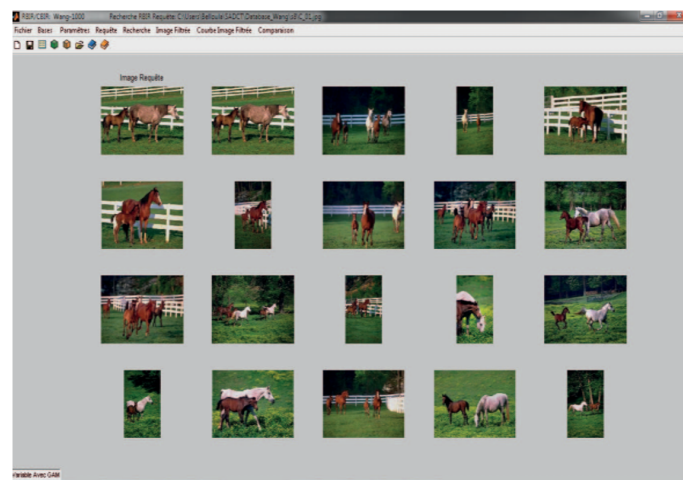
(a)
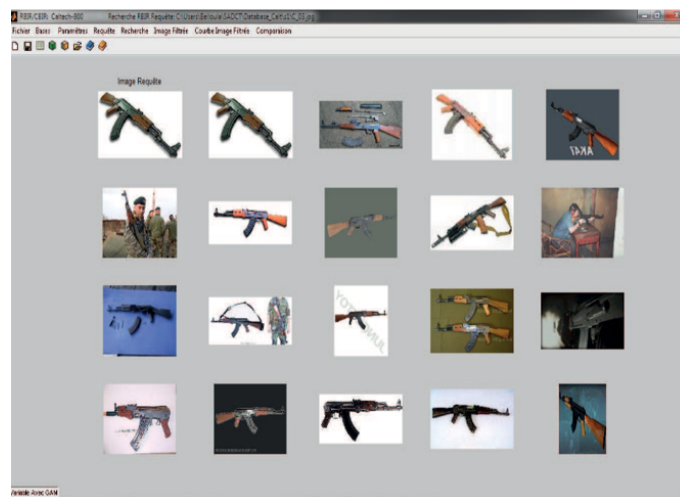
(b)

Figure 12. Example of image retrieval with histogram optimization on Corel-1000 and the Caltech-256 databases. The first image is the query image, the 19 other images are the retrieved images with the proposed method (PM) for (a) horse image and (b) image AK-47

ACKNOWLEDGEMENT

REFERENCES

[1]   A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, 2000, doi: 10.1109/34.895972.

[2]   R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image rerieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (Csur)*, vol. 40, no. 2, pp. 5:1–60, 2008.

[3]   M. M. Jan, N. Zainal, and S. Jamaludin, "Region of interest-based image retrieval techniques: A review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, 2020, doi: 10.11591/ijai.v9.i3.pp520-528.

[4]   A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 32, 2015, doi: 10.1016/j.jvcir.2015.07.012.

[5]   G. Feng and J. Jiang, "Jpeg compressed image retrieval via statistical festures," *Pattern Recognition*, vol. 36, 2003, doi: 10.1016/S0031-3203(02)00114-0.

[6] J. Jiang and G. Feng, "Robustness analysis on facial image description in dct domain," *Electronics Letters*, vol. 43, 2007, doi: 10.1049/el:20072735.

[7] C. C. Chang, J. C. Chuang, and Y. S. Hu, "Retrieving digital images from a jpeg compressed image database," *Image and Vision Computing*, vol. 22, 2004, doi: 10.1016/j.imavis.2003.11.008.

[8] D. Edmundson, G. Schaefer, and M. E. Celebi, "Robust texture retrieval of compressed images," *Proceedings - International Conference on Image Processing, ICIP*, 2012, doi: 10.1109/ICIP.2012.6467386.

[9] D. Zhong and I. Defée, "Dct histogram optimization for image database retrieval," *Pattern Recognition Letters*, vol. 26, 2005, doi: 10.1016/j.patrec.2005.04.012.

[10] M. Majhi and A. K. Pal, "An image retrieval scheme based on block level hybrid dct-svd fused features," *Multimedia Tools and Applications*, vol. 80, 2021, doi: 10.1007/s11042-020-10005-5.

[11] K. O. Cheng, N. F. Law, and W. C. Siu, "Fast extraction of wavelet-based features from jpeg images for joint retrieval with jpeg2000 images," *Pattern Recognition*, vol. 43, 2010, doi: 10.1016/j.patcog.2010.05.012.

[12] A. Chaker, M. Kaaniche, A. Benazza-Benyahia, and M. Antonini, "Efficient transform-based texture image retrieval techniques under quantization effects," *Multimedia Tools and Applications*, vol. 77, 2018, doi: 10.1007/s11042-016-4205-5.

[13] L. Belhallouche, K. Belloulata, and K. Kpalma, "A new approach to region based image retrieval using shape adaptive discrete wavelet transform," vol. 8, no. 1, pp. 1–14, 2016.

[14] D. Zhong and I. Defée, "Performance of similarity measures based on histograms of local image feature vectors," *Pattern Recognition Letters*, vol. 28, 2007, doi: 10.1016/j.patrec.2007.05.019.

[15] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning," *Pattern Recognition*, vol. 41, 2008, doi: 10.1016/j.patcog.2007.12.003.

[16] Y. Liu, X. Chen, C. Zhang, and A. Sprague, "Semantic clustering for region-based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 20, 2009, doi: 10.1016/j.jvcir.2008.11.006.

[17] F. Meng, D. Shan, R. Shi, Y. Song, B. Guo, and W. Cai, "Merged region based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 55, 2018, doi: 10.1016/j.jvcir.2018.07.003.

[18] Y. T. Wu, F. Y. Shih, J. Shi, and Y. T. Wu, "A top-down region dividing approach for image segmentation," *Pattern Recognition*, vol. 41, 2008, doi: 10.1016/j.patcog.2007.11.020.

[19] E. Navon, O. Miller, and A. Averbuch, "Color image segmentation based on adaptive local thresholds," *Image and Vision Computing*, vol. 23, 2005, doi: 10.1016/j.imavis.2004.05.011.

[20] S. Zhu, X. Xia, Q. Zhang, and K. Belloulata, "An image segmentation algorithm in image processing based on threshold segmentation," *Proceedings - International Conference on Signal Image Technologies and Internet Based Systems, SITIS 2007*, 2007, doi: 10.1109/SITIS.2007.116.

[21] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, 2001, doi: 10.1109/34.955109.

[22] W. Y. Ma and B. S. Manjunath, "Netra: A toolbox for navigating large image databases," *Multimedia Systems*, vol. 7, 1999, doi: 10.1007/s005300050121.

[23] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, 2002, doi: 10.1109/TPAMI.2002.1023800.

[24] A. Shokoufandeh, Y. Keselman, M. F. Demirci, D. Macrini, and S. Dickinson, "Many-to-many feature matching in object recognition: A review of three approaches," *IET Computer Vision*, vol. 6, 2012, doi: 10.1049/iet-cvi.2012.0030.

[25] X. Yang and L. Cai, "Adaptive region matching for region-based image retrieval by constructing region importance index," *IET Computer Vision*, vol. 8, 2014, doi: 10.1049/iet-cvi.2012.0157.

[26] P. Manipoonchelvi and K. Muneeswaran, "Significant region-based image retrieval," *Signal, Image and Video Processing*, vol. 9, 2015, doi: 10.1007/s11760-014-0657-0.

[27] W. Yu, K. Yang, H. Yao, X. Sun, and P. Xu, "Exploiting the complementary strengths of multi-layer cnn features for image," *Neurocomputing*, vol. 237, 2017, doi: 10.1016/j.neucom.2016.12.002.

[28] I. Defée and D. Zhong, "Face retrieval based on robust local features and statistical-structural learning approach," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, 2008, doi: 10.1155/2008/631297.

[29] Y. Liu, X. Zhou, and W. Y. Ma, "Extracting texture features from arbitrary-shaped regions for image retrieval," *2004 IEEE International Conference on Multimedia and Expo (ICME)*, vol. 3, 2004, doi: 10.1109/icme.2004.1394628.

[30] A. Belalia, K. Belloulata, and K. Kpalma, "Region-based image retrieval in the compressed domain using shape-adaptive dct," *Multimedia Tools and Applications*, vol. 75, 2016, doi: 10.1007/s11042-015-3026-2.

[31] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E. S. Kim, and T. Le-Tien, "Nic: A robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, 2017, doi: 10.1109/TCSVT.2016.2543118.

[32] T. Sikora and B. Makai, "Shape -adaptive dct for generic coding of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, 1995, doi: 10.1109/76.350781.

[33] K. Belloulata and J. Konrad, "Fractal image compression with region-based functionality," *IEEE Transactions on Image Processing*, vol. 11, 2002, doi: 10.1109/TIP.2002.999669.

[34] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Transactions on Image Processing*, vol. 11, 2002, doi: 10.1109/TIP.2002.1006397.

[35] S. Zhu, X. Xia, Q. Zhang, and K. Belloulata, "A novel spatio-temporel video-object segmentation algorithm," *IEEE Iternational on Industrial Technology*, pp. 673–678, 2008.

[36] "http://wang.ist.psu.edu/ jwang/test1.tar," Jan. Last accessed- Jan. 2018.

[37] "http://www.vision.caltech.edu/image-databases/caltech256/," Jan. Last accessed- March 2014.

[38] Y. Du and J. Z. Wang, "A scalable integrated region-based image retrieval system," *IEEE International Conference on Image Processing*, vol. 1, 2001, doi: 10.1109/icip.2001.958943.

BIOGRAPHIES OF AUTHORS

Amina Belalia 🆔 Ꭶ SC P received her BEng in Computer Science from Mascara University, Algeria, in 2003 and MEng in Computer Engineering from Sherbrooke University, Quebec, Canada, in 2004. She has obtained her PhD in Image Processing and Multimedia from Sidi Bel Abbes University, Algeria, in 2015. Her main research interests include computer engineering, image processing and pattern recognition. Email: a.belalia@esi-sba.dz.

Kamel Belloulata 🆔 Ꭶ SC P received the D.E.A. and Ph.D. degrees in signal, image, and speech processing from the Institut National des Sciences Appliquée (INSA) de Lyon, France, in 1994 and 1998, respectively. From September 1998 to august 1999, he was a Postdoctoral Fellow at INRS-Télécommunications, Montréal, Qc, Canada. From September 1999 to December 2001, he has been an Assistant Professor with the Department of Electrical Engineering of University of Moncton, NB, Canada. Since January 2002, he has been an Associate Professor with the Electrical and Computer Engineering Department of University of Sherbrooke, Qc, Canada. Currently he is a full professor at the University of Sidi Bel Abbès, Algeria. His current research interests are in the areas of image and video compression, pattern recognition and semantic segmentation. He is a Member of IEEE. Web of Science ResearcherID: V-9150-2019. Email: kamel.belloulata@univ-sba.dz.

Shiping Zhu 🆔 Ꭶ SC P was born in Baoji city, Shaanxi Province, China in 1970. He received the B. Eng. and M. Eng. degrees in Measurement Technology and Instruments from Xian University of Technology, Xian, China, in 1991 and 1994, respectively. He received the Ph.D. degree in Precision Instruments and Machineries from Harbin Institute of Technology, Harbin, China, in 1997. From 2000 to 2002, he was a Postdoctoral Fellow at The Brain and Cognition Research Center, Universite Sabatier, Toulouse, France. From 2002 to 2004, he was a Postdoctoral Fellow at the Department of Computer Science and Department of Electrical and Computer Engineering, Universite de Sherbrooke, Qc, Canada. He is currently an Associate Professor with the Department of Measurement Control and Information Technology, School of Instrumentation Science and Optoelectronics Engineering, BeiHang University, Beijing China. His research interests include image/video processing and compression, computer vision, machine vision for 3D measurement, 3D visual. Email: shiping.zhu@buaa-edu.cn.