Int J Multimed Info Retr (2015) 4:261–276
DOI 10.1007/s13735-015-0084-1

REGULAR PAPER

Region-based image retrieval using shape-adaptive DCT

Amina Belalia[1] · **Kamel Belloulata**[1] · **Kidiyo Kpalma**[2]

Received: 1 June 2015 / Revised: 1 August 2015 / Accepted: 11 August 2015 / Published online: 25 August 2015

Abstract Content-based image retrieval (CBIR) is the process of searching digital images in a large database based on features, such as color, texture and shape of a given query image. As many images are compressed by transforms, constructing the feature vector directly in transform domain is a very popular topic. Therefore, features can be extracted directly from images in compressed format by using, for example, discrete cosine transform (DCT) for JPEG compressed images. Also, region-based image retrieval (RBIR) has attracted great interest in recent years. This paper proposes a new RBIR approach using shape-adaptive discrete cosine transform (SA-DCT). In this retrieval system, an image has a prior segmentation alpha plane, which is defined exactly as in MPEG-4. Therefore, an image is represented by segmented regions, each of which is associated with a feature vector derived from DCT and SA-DCT coefficients. Users can select any region as the main theme of the query image. The similarity between a query image and any database image is ranked according to a same similarity measure computed from the selected regions between two images. For those images without distinctive objects and scenes, users can still select the whole image as the query condition. The experimental results show that the proposed approach is able to identify main objects and reduce the influence of background in the image, and thus improve the performance of image retrieval in comparison with a conventional CBIR based on DCT.

✉ Kamel Belloulata
kamel.belloulata@USherbrooke.ca

[1] Electronics Department, Faculty of Engineering, University of Djillali Liabes, BP 89, Sidi bel Abbès, Algeria

[2] UMR 6164, IETR, UEB INSA, 35708 Rennes, France

Keywords Content-based image retrieval (CBIR) · DCT · Segmentation · Region-based image retrieval (RBIR) · Semantic image retrieval · SA-DCT

1 Introduction

Widespread use of digital imaging devices in various areas of application has resulted in large volumes of images acquired and stored on computers. Effective and efficient retrieval of images stored in such large databases has become desirable. Comparing to the conventional retrieval using descriptive keywords, content-based image retrieval (CBIR) offers a way of retrieving images according to their visual content [1]. After two decades of research, it is acknowledged that the performance of CBIR systems is mainly limited by the gap between low-level features and high-level semantic concepts [2]. Unlike early CBIR approaches, which compute global features of images, the region-based image retrieval (RBIR) extract features of the segmented regions [3] and perform similarity comparisons at the granularity of the region [4]. The CBIR systems often used global features to represent the content of an image [5–8]. If the images contain no obvious objects, the global feature-based approaches proved to be successful in describing the image content. However, for images with salient objects, using global features leads to inaccurate content descriptions. In order to solve this problem, RBIR systems are researched and developed [9,10]. These systems applies image segmentation to decompose an image into regions to enhance the ability of capturing as well as representing the focus of the user's perceptions of image content [11,12]. The RBIR systems attempted to overcome the limitation of global-based retrieval systems by emphasizing the target objects and minimizing the influence of background. During retrieval, the user is pro-

vided with segmented regions of the query image and is required to assign several properties, such as the regions to be matched, the features of the regions, and even the weights of different features [13]. The matching of the regions is restricted to be one-to-one, i.e., one region of an image can only match one region of another image or image by image [14], i.e., many region to many region [15]. However, semantically precise image segmentation is still an open problem [16].

For compressed images, features can be extracted directly from images in their JPEG compressed format [17–20] by using, for example, Discrete Cosine Transform (DCT) which is a part of compression process [21,22]. Recently, there were numerous approaches based on using the DCT block processed images for information extraction [23–28]. The improvement in using DCT is explained by its decorrelation properties, feature preservation and reduction in complexity. It means that much of the energy lies in low frequency coefficients, so that high frequency can be discarded without visible distortion. It is evident that the use of DCT block transformation in combination with other techniques results in good retrieval performance. In most cases DCT can be seen as a preprocessing step followed by a more or less sophisticated method for extracting structural features. It has been shown in some cases that the DCT based processing gives significantly better retrieval results than direct pixel based processing [24,26,27,29].

In early CBIR which uses the DCT blocks [24,26,27,30], the boundary blocks contain pixels from an object and either from background or from another object (Fig. 1). Thus, region-based retrieval is not possible. Also, the retrieval quality may suffer since pixels on different sides of the boundary may have different characteristics; by applying the standard DCT to such a block, spectral properties of these pixels are mixed up making the search for a good object retrieval unreliable. In particular, a sharp intensity transition will cause significant spectral oscillations. To develop their RBIR system, Liu and Zhang [10,11] have used Projection On Convex Sets (POCS) in the wavelet domain by using Gabor filters [31]. Recently, Zhang et al. [32] have also proposed to apply Curvelet transform to an arbitrarily shaped object for RBIR. In this paper, we propose to apply SA-DCT [33], which is a part of the compression process in the MPEG-4 standards [34], that takes into account prior segmentation of the image into regions [35].

The paper is organized as follows. In Sect. 2, we briefly review the theory of CBIR in the DCT domain. In Sect. 3, we make a problem statement, the Shape-adaptive DCT transform is described and the proposed RBIR system is presented. In Sect. 4 numerous experimental results are shown and the robustness of our system is proved whereas in Sect. 5 we draw conclusions.

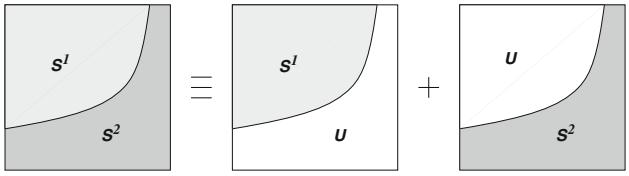


Fig. 1 Block with two segments $\mathcal{S}^1$ and $\mathcal{S}^2$, belonging to different regions, and its decomposition into two blocks with active pixels (*gray*), either in $\mathcal{S}^1$ or in $\mathcal{S}^2$, and inactive (*undefined*) pixels in $\mathcal{U}$ (*white*)

2 Content-based image retrieval using DCT transform and histogram of its patterns

Let I be the intensity of an image to be transformed and $\mathbf{x} = (x, y)$ be spatial coordinates of a pixel in this image. Let $\{b_1, \ldots, b_M\}$ be the set of M non-overlapping blocks, i.e., collections of pixel coordinates, partitioning the image. To permit compact notation, let I_{b_i} denote restrictions of image intensity I to the block b_i. In other words, $I_{b_i} = \{I(\mathbf{x}) : \mathbf{x} \in b_i\}$. Each block b_i is then transformed to the frequency domain by the DCT. This transform is originally defined in the 1-D form, and can be used to construct 2-D separable transform. It has been found useful for source coding, especially image and video coding.

In DCT-based methods, first, all blocks are transformed via DCT: $\widehat{I}_{b_i} = \mathrm{DCT}(I_{b_i})$. Let $\mathbf{u} = (u, v)$ be a 2-D frequency of a DCT coefficient. The DCT transform $\widehat{I}(\mathbf{u})$ for a block represented by NxN pixel values $I(\mathbf{x})$ for $x, y = 1, \ldots, N$ can be defined as

$$\widehat{I}(u, v) = \frac{2}{N} c(u)c(v) \sum_{x=1}^{N} \sum_{y=1}^{N} I(x, y) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \times \cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (1)$$

$c(u) = c(v) = \frac{1}{\sqrt{(2)}}$ for $u = 0$ $v = 0$ and $c(u) = c(v) = 1$ for $u, v > 0$.

2.1 AC-pattern and its histogram

The proposed approach, for 4×4 DCT block, selects 9 coefficients out of all 15 AC coefficients in each block and uses their statistical information to construct the AC-pattern. A new problem could occur from the fact that various DCT blocks sizes have been used in different compression standards like JPEG and MPEG. Jiang and Feng [36] have proposed the spatial relationship of DCT coefficients between a block and its sub-blocks, to deal with inter-transfer of DCT coefficients from different blocks with various sizes. Our selection gathers 9 coefficients into 3 groups: Horizontal (Group H), Vertical (Group V) and diagonal (Group D) as shown in Fig. 2. For each group, the sum of the coefficients is cal-

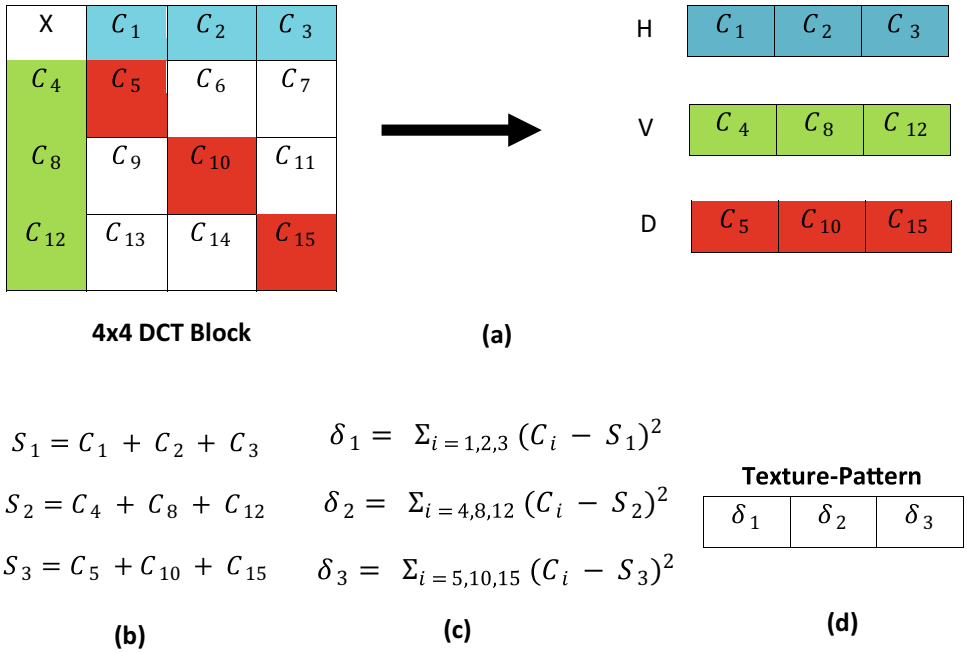
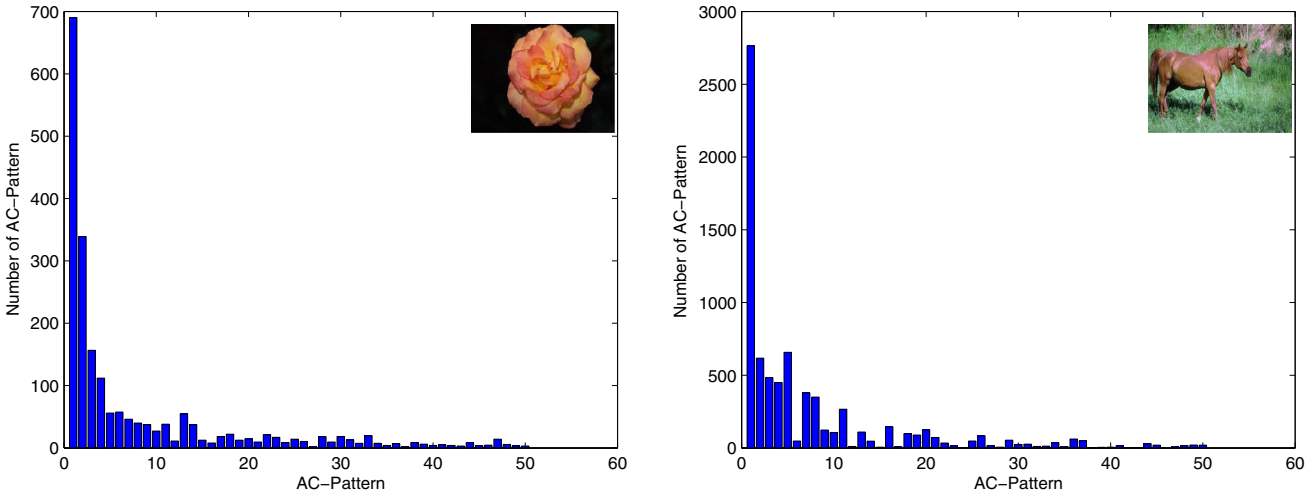


Fig. 2 The process of forming texture-pattern: **a** three groups of AC coefficients extracted from DCT block, **b** sums of each group, **c** sums of squared-differences and **d** texture pattern

culated firstly and then the squared differences between each coefficient and the sum of this group are calculated. Finally, these squared-differences of the three groups are used to construct the AC-Pattern. This selection is retained because of its ability to represent local structure of content block [37]. Compared with the method of [26] which uses 15 AC-Pattern, this selection can reduce the complexity of the feature vector obviously. The histograms of feature vectors based on the coefficients of quantized block transforms are used for the description of global statistical information. Histograms are relevant since they are based only on the overall statistics and no structural information about the location of features is considered [38]. Moreover, the variability in image sizes influences the number of occurrences of the patterns. Therefore, a normalization of the resulted histograms is mandatory. Since histograms are normalized, they can also be regarded as describing probability of features, and also as vectors with unit length. The idea of the proposed technique of histogram generation is based on the approach used in [26]. The histogram of the whole database is generated to produce a reference histogram from which we calculate decision levels that enable us to generate descriptors. Commonly and particularly in [26], histogram descriptors are designed by simply uniformly quantizing the feature of interest (the DCT coefficients) and then counting the number of elements having the same value and then taking the number of desired bins. This approach has proved to provide efficient results. The proposed technique is borrowed from image coding where histogram equalization is applied. In image coding, it is shown that this adaptive quantization can help to reduce the bit rate and to enhance signal-to-noise ratio. Starting from this observation, we propose to design the histogram-based descriptor by defining the histogram bins like in the case of adaptive quantization. First, the reference histogram of the database is generated and normalized to sum to the unity. Then the adaptive quantization is designed such a way that the "probability" to get a value between any two consecutive decision levels is equal. Knowing the desired number K of bins, this probability equals $PK = 1/K$. This enables to generate decision levels that are then used to define the descriptor for the query. Finally, for each individual histogram, use the generated decision levels and sum up the bins between two consecutive levels. This leads to the descriptor for the image under study. From the original histogram of AC-Pattern (Fig. 3), we can make two observations: the first is that there is only part of AC-Patterns that appear in large quantities and a large number of AC-Patterns that appear rarely [26]. So in consideration of time-consuming and efficiency, we just select some of AC-Patterns which have higher frequency to construct the histogram. For constructing the AC-Pattern histogram of an image, we just calculate the number of appearance of these AC-Patterns in this image, and then we get the AC-Pattern histogram H_{AC}. The second observation is that the first AC-Pattern inside the histogram is very dominant. This AC-Pattern mainly corresponds to uniform blocks in the image [37].

2.2 DC-pattern and its histogram

Differently from previous AC-Patterns that describe the local feature information inside each block (Intra-Block), DC-Patterns integrate more global features by using gradients between each block and its neighbors (Inter-Block). DC-DirecVec [26] is defined and used as feature for DC-Patterns.



(a) *AC-Pattern of Flower* **(b)** *AC-Pattern of Horse*

Fig. 3 Histogram of the first 50 AC-Patterns with highest frequency of occurrence (content-based image retrieval)

For a given DC value, the differences between its value and the eight neighboring DC values are calculated and arranged in 3x3 matrix as shown in Fig. 4. These differences are called direction values and represents orientation information. The center value of the matrix is calculated as the difference between the current DC value and the mean of the all the nine DC values of the matrix. Next the differences are ordered according to their absolute values, and the corresponding index of the direction is listed to be an 8-bin vector. Like the same observations can be done in AC-Pattern histogram H_{AC}, we select those dominant DC-Patterns to construct DC-Pattern histogram H_{DC}.

2.3 Feature descriptor and similarity measurement

For each block, AC-Pattern is formed by 9 coefficients and DC-Pattern is constructed by the DC coefficients of the block itself and those of its 8 neighboring blocks. So intra-block features and inter-block features can be represented by these two patterns. Feature histograms can be defined for different kinds of features and feature vectors based on the DC and AC coefficients which represent complementary types of local information. These histograms can be combined by the concatenation of AC-Pattern and DC-Pattern histograms to do image retrieval (Fig. 5). More general

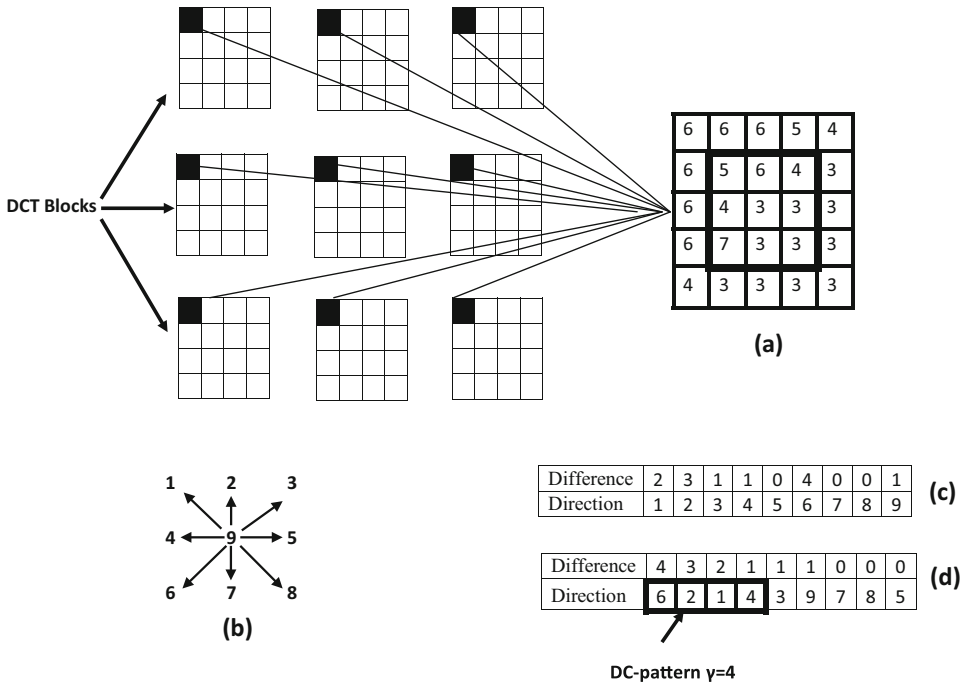
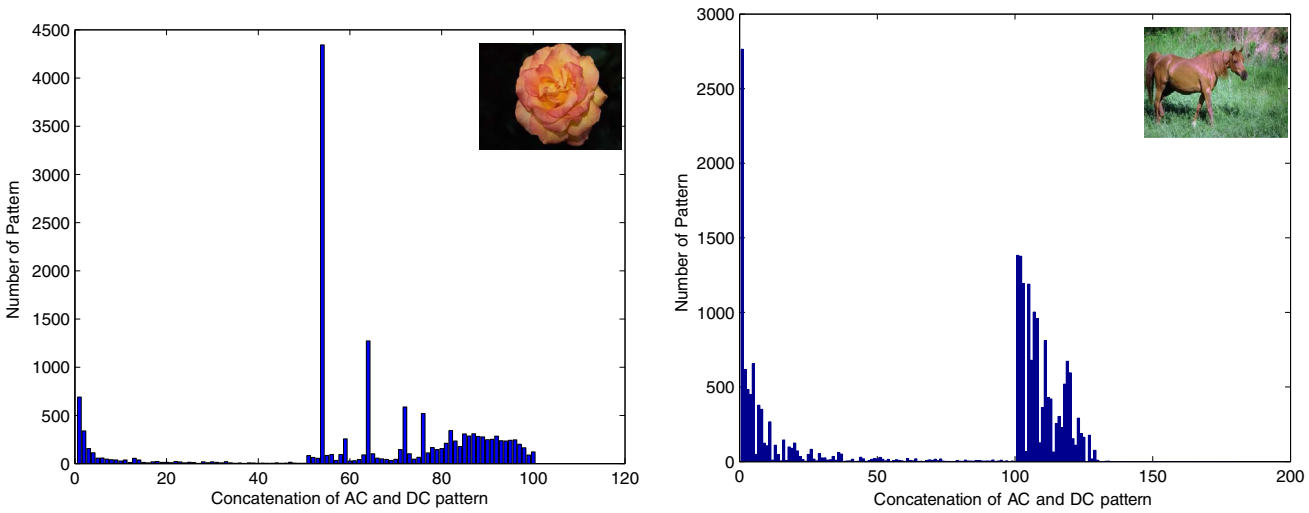


Fig. 4 The process of forming DC-Pattern [26]: **a** DC coefficients are extracted from neighboring blocks, **b** definition of directions, **c** differences for each direction and **d** directions with largest differences from DC-Patterns



(a) *AC and Pattern of Flower* **(b)** *AC and DC Pattern of Horse*

Fig. 5 Combined histogram of the first AC-Patterns and DC-Patterns with highest frequency of occurrence for content-based image retrieval

concatenation is defined by applying α [Eq. (2)], a weight parameter that controls the impact of AC-Patterns and DC-Patterns histogram. In this context, the descriptor is defined as follows:

$$H = [(1-\alpha) \times H_{\mathrm{AC}}, \alpha \times H_{\mathrm{DC}}] \tag{2}$$

The feature histograms can be seen as normalized vectors. This allow comparing statistical information of images in quantitative way by calculating the distance between the histograms under suitable norm. Many distances have been used to define the similarity of two color histogram representations [39]. In our system, the similarity between query and images in the database is assessed by the Chi-Squared distance [Eq. (3)] between these corresponding feature descriptor. The Chi-Square test is used to compare two binned data sets and to determine if they are drawn from the same distribution function.

$$\mathrm{Dis}_{(i,j)} = \sum_{k=1}^{M} \frac{(D_i(k) - D_j(k))^2}{D_i(k) + D_j(k)} \tag{3}$$

where k demonstrates the components of the descriptor and i and j demonstrate the different descriptors, M indicates the dimension of the descriptor. The earth movers distance (EMD) proposed by Rubner et al. [40] can also be used in our system.

3 Problem statement and related work

A number of image search engines have been developed. We cannot survey all related work in the allocated space. Instead, we try to emphasize some of the work that is most related to our work. Early CBIR systems in the compressed domain use global features, like DCT coefficients histograms [26,30] to represent images. They have used the histograms to provide information about the distribution of features, without any information about locations of features. These global features are extracted from entire images and often fail to capture local details that exist in most natural images. As a consequence, dissimilar images varying in local details have similar feature vector representations. This anomaly creates difficulties for global representations to have direct relations with image semantics. Also, Zhong and Defee [38] have proposed a framework for the unification of statistical and structural information for pattern retrieval. They incorporate structural information description for patterns by considering decomposition of patterns into *rectangular subareas*. The similarity between images is then determined by comparing the signatures of subimages. An obvious drawback of the system is the sharply increased computational complexity and increase of size of the search space due to exhaustive generation of subimages.

In our approach, to address this problem, we follow the techniques that divide images into semantics regions with a hope that a region level representation would capture local details. These techniques are formally known as RBIR techniques. Almost all region-based techniques work in the same way—divide images into regions and represent them in terms region features.

In the classical CBIR (Sect. 2), the blocks are defined independently of image content. The boundary blocks contain pixels from an object and either from background or from another object (Fig. 1). Thus, independent retrieving of objects is not possible. Also, the retrieval quality may suffer since pixels on different sides of the boundary may have

different characteristics; by applying the standard DCT to such a block, spectral properties of these pixels are mixed up making the search for a good region-to-region correspondence unreliable. In particular, a sharp intensity transition will cause significant spectral oscillations. To alleviate this problem, we propose a new approach to RBIR that takes into account prior segmentation of the image into regions so as to perform similarity comparisons at the granularity of the region independently of others regions (or from background). This allows independent retrieval of individual regions, thus permitting new functionalities.

3.1 Extrapolation-based method

To permit a texture feature extraction from arbitrary-shaped regions in RBIR system, Liu et al. [31,41] used an iterative approach based on the theory of successive Projections Onto Convex Sets (POCS) which they borrowed from [42]. POCS algorithm is the iterative technique and the iterative process is terminated when the pixels outside the boundary converge. The algorithm can be described by the following steps:

1. extrapolate the undetermined samples (Fig. 1) in $\mathcal{U}$ from $\mathcal{S}$ (to obtain a signal supported on a rectangle);
2. compute a 2-D DCT on the rectangle;
3. exit if the result is satisfactory;
4. set to zero some of the DCT coefficients (to limit signal bandwidth);
5. compute the inverse 2-D DCT for the bandwidth-limited signal;
6. within $\mathcal{S}$, replace the resulting samples with samples from the original signal (to assure unmodified signal inside $\mathcal{S}$);
7. return to (2).

According to the POCS theory, the method will converge only if the selection of retained DCT coefficients is fixed throughout iterations. This means that the selection of coefficients to be retained is highly dependent on the initial extrapolation; large high-frequency components may be produced by an unsuitable extrapolation, and thus may result in a suboptimal zone for coefficients to be retained. In [31], they use Discrete Wavelet Transform (DWT) instead of DCT. In Region-Based image compression, Stasinski and Konrad [43] have shown a great performances of the shape-adaptive DCT in comparison with POCS.

3.2 Region-based image retrieval using shape-adaptive DCT

In our system, all database images are processed off-line. Firstly, the system segments every image into two regions: Foreground (the main object) and Background, and then extracts low-level texture features of each region by using DCT and SA-DCT. Once features are extracted, we design two different systems to check the retrieval effectiveness. These two systems differ in query formulation and image distance calculation. The first system uses query by region of interest like Blobworld [44]. We name this *Single Region Query Based Technique*. The distance between the query and a database image is the minimum of distances between the query region and all regions of the database image. The second system uses global query, like SIMPLIcity [4], in which an entire query image is matched with an entire database image. We call this *Global Query Region-Based Technique*. The individual distances between regions of two images are combined into one distance measure.

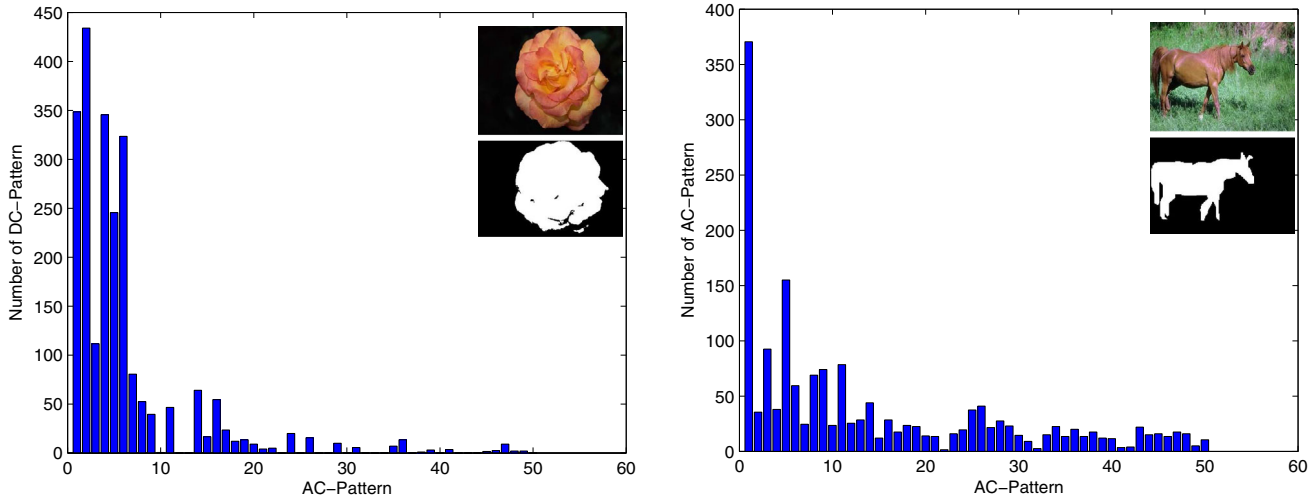
During retrieval, the user selects a region of interest from the query image as the query region. It is assumed that each image has at least one dominant region (Foreground) that expresses its semantics. The system calculates the low-level features of the query region. Then, a subset of N images is retrieved from the database. This set consists of those images which contain regions (objects) of same concept as that of the query region (object). These N images (objects) are ranked according to their distance to the query image. The low-level features are based on the combined Histogram [Eq. (2)] of AC and DC-Patterns (see Fig. 5), for the foreground only, the background only or the two regions together (full image). For the full image, the histograms can be combined by the concatenation of Foreground and Background histograms (see Fig. 7) to do image retrieval. More general concatenation is defined by applying β [Eq. (4)], a weight parameter that controls the impact of Foreground and Background histogram. In this context, the global descriptor is defined as follows:

$$H_T = [(1-\beta) \times H_{\text{Fore}}, \beta \times H_{\text{Back}}] \tag{4}$$

Histogram-based descriptor for images retrieval is basically motivated by the importance of the interrelation among DCT coefficients, thus regions are processed in such a way that each coefficient (SA-DCT) provides a piece of information about both the shape and color content of the objects. From the AC-Patterns histograms (Figs. 3, 6), we can see that there is AC-Pattern in which all the AC coefficients are zero (Fig. 6) which mainly corresponds to uniform block of image.

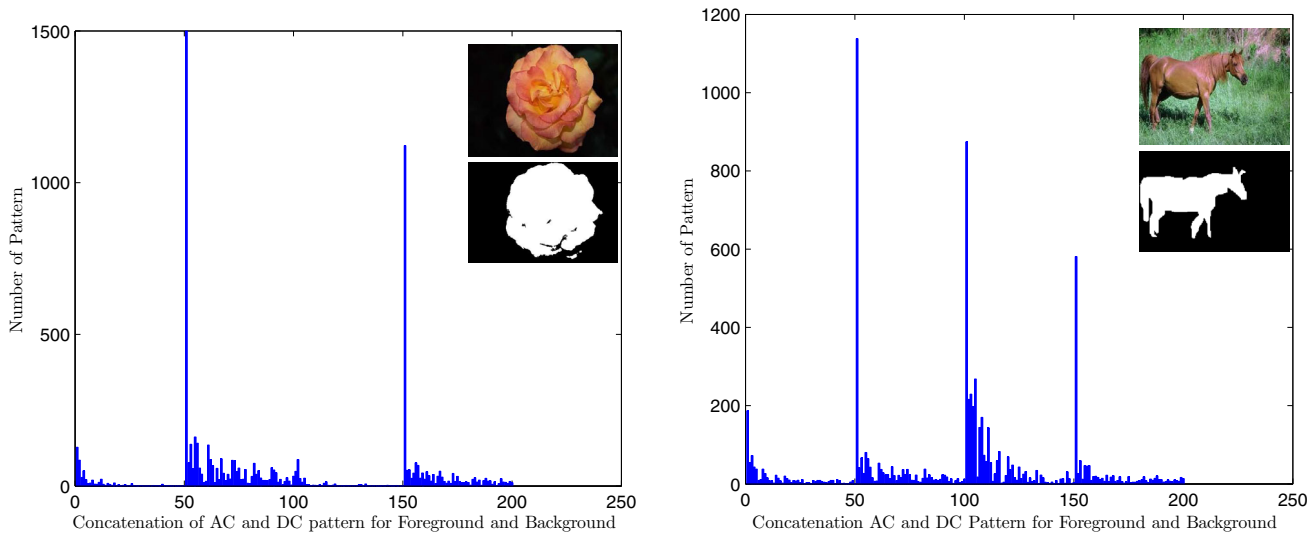
3.3 Image segmentation

In this paper, we use the supervised Total Variation segmentation method (TV-Seg) [45]. We use this method to segment images by taking into account both segmentation quality and computational complexity. It is noted that there are also other meaningful segmentation works, such as [16,46,47], which could be used to retrieve images by combining with our pro-



(a) *AC-Pattern of Flower* **(b)** *AC-Pattern of Horse*

Fig. 6 Histogram of the first 50 AC-Patterns with highest frequency of occurrence (region-based image retrieval)



(a) *Combined histogram for Flower image* **(b)** *Combined histogram for Horse image*

Fig. 7 Combined histogram of the AC-Patterns and DC-Patterns, with highest frequency of occurrence, for the foreground and background (region-based image retrieval)

posed RBIR method. Note that our objective is to check the retrieval effectiveness of different representations but not the segmentation algorithms.

3.4 Shape-Adaptive Transform

After image segmentation, we need to extract regions features. We propose to apply shape-adaptive DCT (SA-DCT) [33,48] to each segment $\mathcal{S}$ of the boundary blocks 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, and then to apply horizontal DCTs to the vertical DCT coefficients with the same frequency index. Figure 8 illustrates the procedure. The final coefficients of the SA-DCT are located in the upper-left corner of each block. The number of the SA-DCT coefficients is identical to the number of active pixels. Since the shape of each segment is stored, the retrieval system can perform the inverse SA-DCT. 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. The performance of SA-DCT is surprisingly high and closely approaches that of advanced methods based on basis orthogonalization [49]. Moreover,

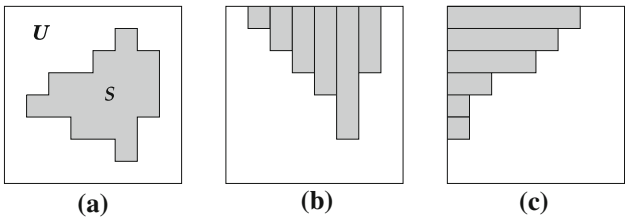


Fig. 8 Illustration of SA-DCT: **a** arbitrarily-shaped region; **b** *vertical* alignment followed by vertical 1-D DCTs; **c** *horizontal* alignment followed by horizontal 1-D DCTs

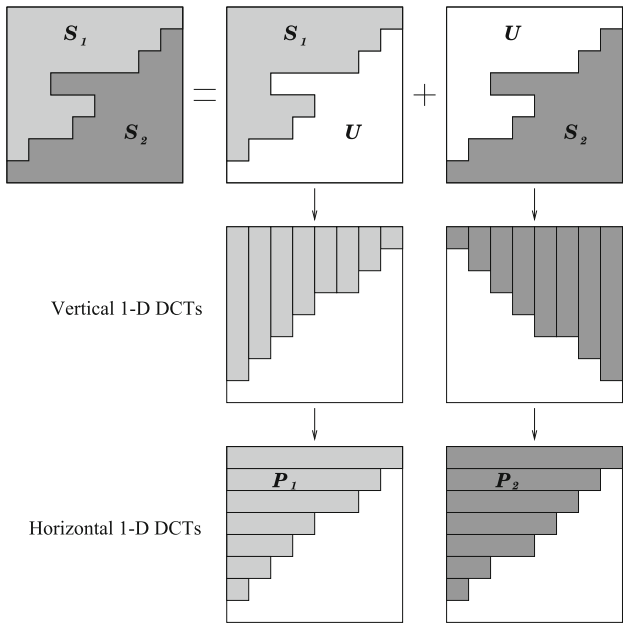


Fig. 9 Example of application of the SA-DCT to a boundary block with two segments. Note that due to *horizontal* and *vertical* shifts the shape $\mathcal{P}_i$ is different from that of $\mathcal{S}_i$

SA-DCT can be implemented in real-time [50] whereas the orthogonalization-based approaches are very demanding memory- and CPU-wise. Due to these properties, the SA-DCT algorithm has become a common tool for coding of arbitrarily shaped image regions [51,52] and, in particular, has been incorporated into MPEG-4 [34]. In this paper, we use a variant of SA-DCT, called ΔDC-SA-DCT [53]. It improves the performance of the SA-DCT by means of two additional processing steps: extraction of the DC component from the segment $\mathcal{S}$ before performing forward SA-DCT and ΔDC correction carried out during the inverse SA-DCT.

Recall that $\widehat{I}_{b_i}$ are DCT-transformed blocks of intensities I_{b_i}. Let $\mathcal{P}^n_{b_i}$ (Fig. 9) be the segment $\mathcal{S}^n_{b_i}$ (n-th segment of boundary block b_i) after SA-DCT. Note that the shape of $\mathcal{P}^n_{b_i}$ is different from that of $\mathcal{S}^n_{b_i}$ due to the executed vertical and horizontal shifts, but that the number of pixels is unchanged. Also, let $\widehat{I}^n_{b_i}(\mathbf{u})$ be an SA-DCT coefficient in $\mathcal{P}^n_{b_i}$ at frequency $\mathbf{u}$. To construct the AC-Pattern (Sect. 2.1), we will select at most 9 coefficients in each extrapolated segment $\widetilde{I}^n_{b_i}$.

where $\widetilde{I}^n_{b_i}$ is an extrapolated n-th segment of the SA-DCT-transformed block intensity:

$$\widetilde{I}^n_{b_i}(\mathbf{u}) = \begin{cases} \widehat{I}^n_{b_i}(\mathbf{u}) & \text{if } \mathbf{u} \in \mathcal{P}^n_{b_i}, \\ \mathbf{v} & \text{otherwise.} \end{cases} \tag{5}$$

Although various $\mathbf{v}$ values could be used, the to-be-padded coefficients are at higher frequencies and therefore a logical choice, that we adopt here, is to set $\mathbf{v}$ to zero (ZERO padding). Note that the same approach of padding was used with POCS in [31,41].

4 Experimental results

4.1 Experimental databases

To assess the effectiveness of the proposed system, we use three publicly available and widely used databases, Corel-1000 database [54], Caltech-256 database [55], and GTF database, a commonly used database for face recognition [56].

Corel-1000 is a real-world image database collected by Wang et al. [54], including 1000 256 × 384 or 384 × 256 images, and it is a subset of Corel database. These images are classified into 10 semantic categories (see Table 1): African, buildings, beach, buses, dinosaurs, elephants, flowers, horses, mountains and food.

Caltech-256 database collected by Griffin et al. [57], consists of 30 607 images from 256 object categories and contains from 80 to 827 images per category. In our experiments, we select 10 categories (see Table 2) randomly from Caltech-256, which contains 1299 images. The 10 object categories are Ak-47, American-flag, Backpack, Baseball-bat, Baseball-glove, Basketball-hoop, Bat, Bathtub, Beer-mug and Blimp.

Table 1 Corel-1000 categories of images tested

ID	Category name
1	African people and villages
2	Beach
3	Buildings
4	Buses
5	Dinosaurs
6	Elephants
7	Flowers
8	Horses
9	Mountains and glaciers
10	Food

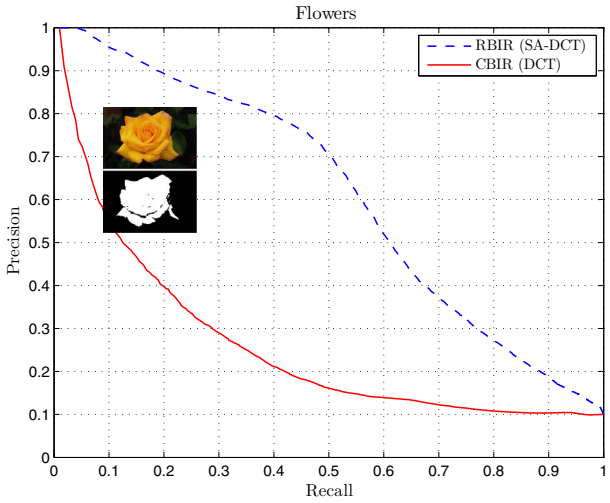
Table 2 Caltech-256 categories of images tested

ID	Category name
1	Ak-47
2	American-flag
3	Backpack
4	Baseball-bat
5	Baseball-glove
6	Basketball-hoop
7	Bat
8	Bathtup
9	Beer-mug
10	Blimp

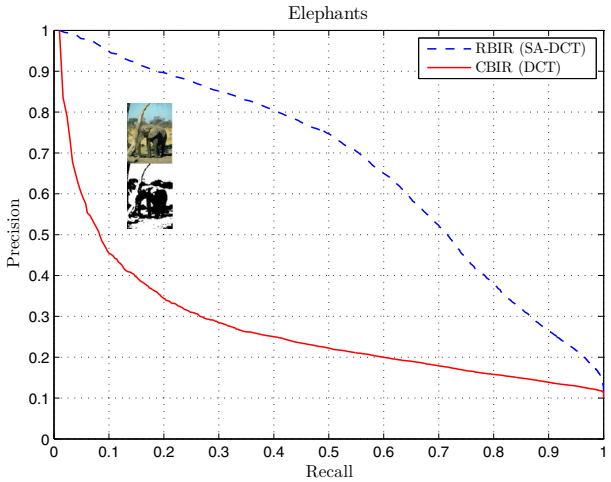
The GTF database contains 15 images of 50 people. The images are at the resolution 640 × 640 pixels in which the size of face is 150 × 150 pixels.

4.2 Performance evaluation

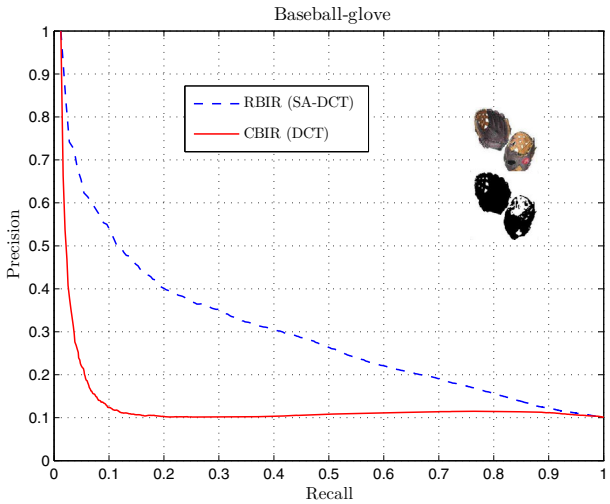
Images are considered as similar if the distance [Eq. (3)] between their features descriptors is under a given threshold. Then the performance can be evaluated by precision and recall. Precision indicates the retrieval accuracy and is defined as the ratio of the number of relevant retrieved images over the number of total retrieved images. Recall indicates the



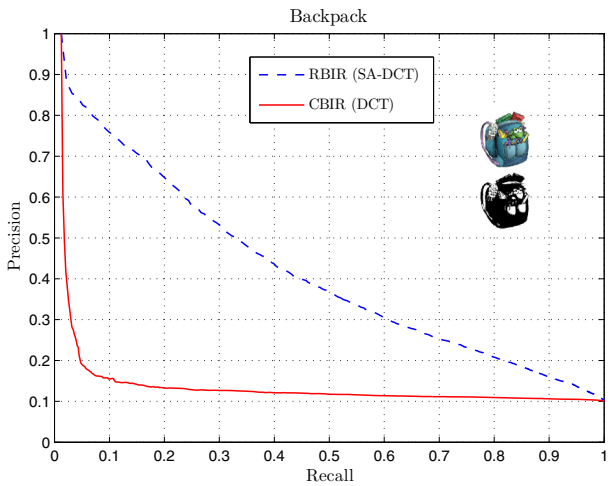
(a) *The P-R curves using "Flowers" from Corel-1000*



(b) *The P-R curves using "Elephants" from Corel-1000*

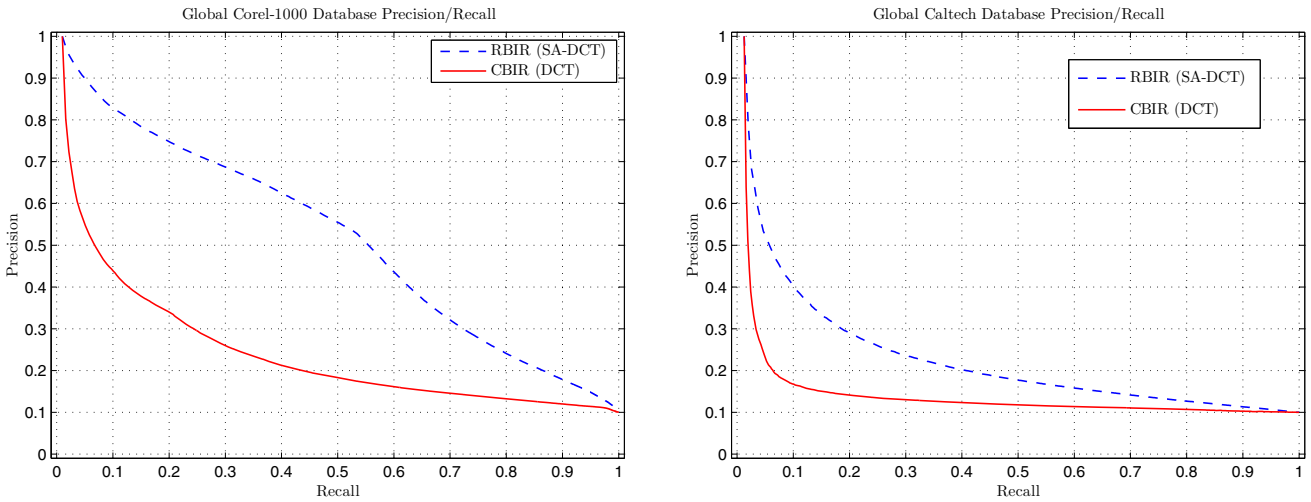


(c) *The P-R curves using "Baseball-glove" from Caltech-256*



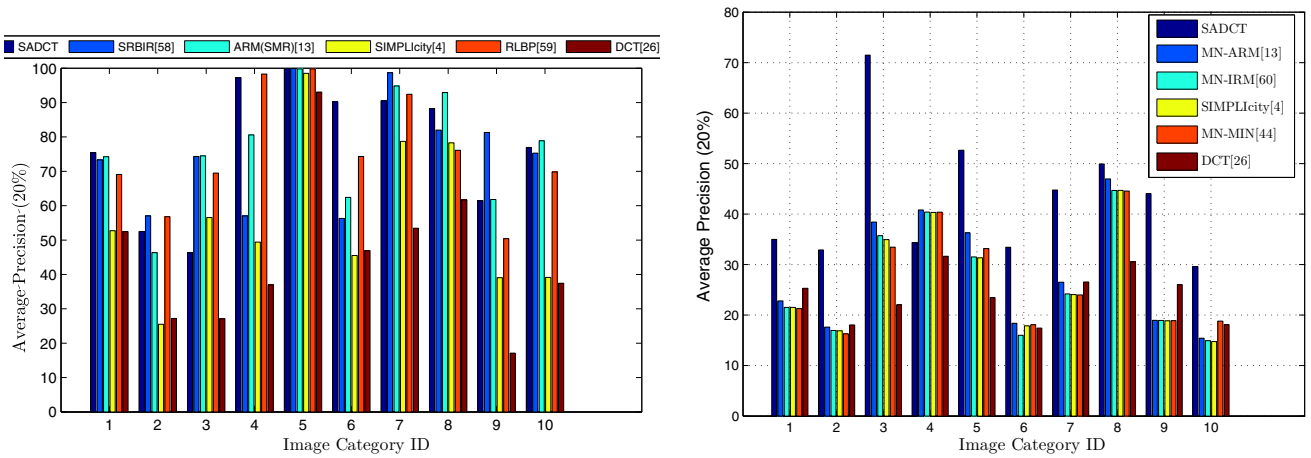
(d) *The P-R curves using "Backpack" from Caltech-256*

Fig. 10 The average Precision-Recall curves between our Region-Based approach (*with mask*) and the conventional Content-Based approach (*without mask*). The image categories **a** "Flowers", **b** "Elephants" are from Corel-1000 database, and the image categories, **c** "Baseball-glove" and **d** "Backpack" are from Caltech-256 database



(a) *Precision-Recall on Corel-1000 Database* **(b)** *Precision-Recall on Caltech-256 Database*

Fig. 11 The average Precision-Recall results: **a** Corel-1000 database, **b** Caltech-256 database, between our Region-Based approach (*with mask*) and the conventional Content-Based approach (*without mask*) [26]



(a) *Precision-Recall on Corel-1000 Database* **(b)** *Precision-Recall on Caltech-256 Database*

Fig. 12 The mean average precision (MAP) **a** Corel-1000 database, **b** Caltech-256 database

ability of retrieving relevant images from the database. It is defined as the ratio of the number of relevant retrieved images over the total number of relevant images in the database. Relevant images are referred to images in the same category. We have also used the mean average precision (MAP). It is the common way to transform P-R graph into one value. The Mean Average Precision over all queries is defined as

$$\text{MAP} = \frac{1}{Q} \sum_{q \in Q} \text{AP}(q) \tag{6}$$

where Q is the set of queries and average precision of query q, $\text{AP}(q)$ is defined as

$$\text{AP}(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P(R_n) \tag{7}$$

where R_n is recall after the nth relevant image is retrieved. $P(R_n)$ is precision when recall is R_n. MAP contains not only precision and recall, but also rank of relevant images.

4.3 Results with the Corel-1000 and Caltech-256 databases

In evaluating the effectiveness of the RBIR (DCT + SA-DCT) system in comparison to the CBIR (DCT) system, the one which gives higher precision value at the same recall value is the more effective system. Ten categories of images are tested. The average Precision-Recall curves between our region-based approach (with mask) and the conventional content-based approach (without mask) are shown in Fig. 10. We have shown the categories "Flowers" Fig. 10a. and "Elephants" Fig. 10b, from Corel-1000 database, and

Table 3 Average precision of different methods on Corel-1000 database

ID category	1	2	3	4	5	6	7	8	9	10
SA-DCT	**75.45**	**52.5**	**76.3**	**97.25**	**100**	**90.20**	**95.50**	**94.25**	**61.85**	**78.95**
MN-ARM [13]	74.23	46.35	74.51	80.60	99.80	62.42	94.84	92.91	61.80	78.88
MN-IRM [60]	73.80	42.86	74.32	74.24	99.80	61.00	95.23	93.49	51.33	66.54
SIMPLIcity [4]	52.73	25.52	56.54	49.40	98.50	45.51	78.73	78.28	39.00	39.13
MN-MIN [44]	43.40	29.43	31.82	46.63	90.11	33.07	64.94	63.62	27.90	35.21
DCT [26]	52.45	27.20	27.1	36.95	93.05	46.90	53.45	61.65	17.10	37.45

Table 4 Average precision of different methods on Caltech-256 database

ID category	1	2	3	4	5	6	7	8	9	10
SA-DCT	**35.0**	**32.89**	**71.48**	**34.37**	**52.65**	**33.43**	**44.76**	**49.92**	**44.06**	**29.60**
MN-ARM [13]	22.81	17.63	38.41	40.83	36.28	18.37	26.51	46.98	18.95	15.41
MN-IRM [60]	21.53	16.96	35.73	40.39	31.52	16.00	24.20	44.68	18.94	14.94
SIMPLIcity [4]	21.51	16.88	34.95	40.33	31.33	17.87	24.06	44.69	18.86	14.76
MN-MIN [44]	21.30	16.28	33.45	40.38	33.17	18.09	24.00	44.57	18.90	18.80
DCT [26]	25.31	18.04	22.03	31.64	23.43	17.42	26.56	30.54	26.01	18.12

the categories "Baseball-glove" Fig. 10c and "Backpack" Fig. 10d, from Caltech-256 database as the reference query images. The comparison of the average Precision-Recall results, on the Corel-1000 database, between the proposed Region-Based approach and the conventional Content-Based approach [26] is shown in Fig. 11a. The same curves, for Caltech-256, are shown in Fig. 11b. The best performance is obtained by using the global combined descriptor [Eq. (4)] of the foreground and the background together. So, the proposed system attempts to overcome the limitation of global-based retrieval (CIBR with DCT only) systems by emphasizing the target objects and minimizing the influence of background.

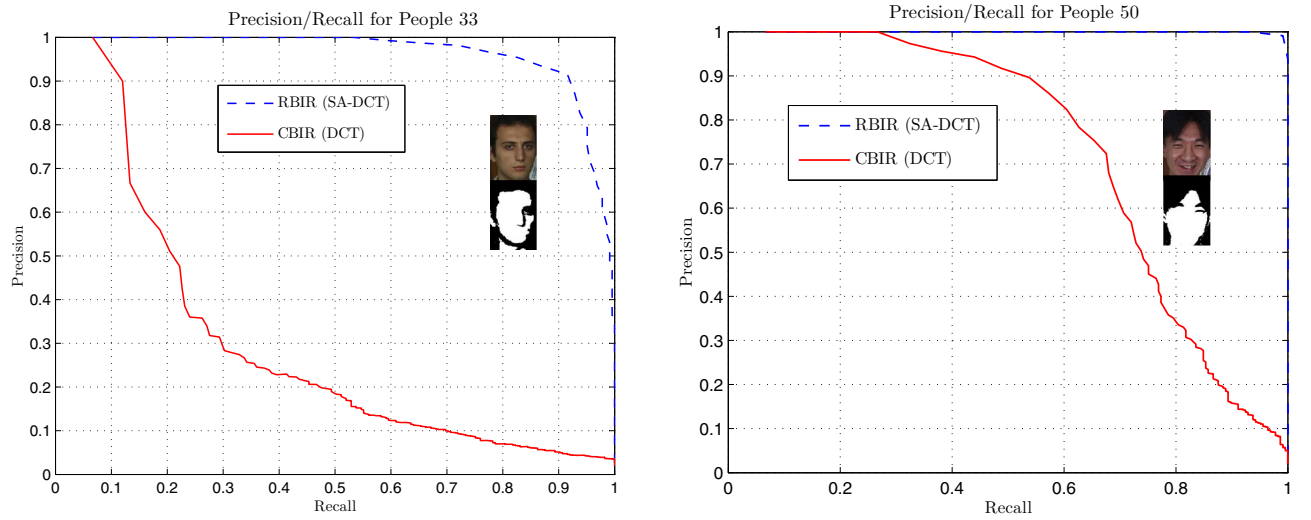
In Fig. 12, we have made a comparison between our approach and five existing RBIR methods: significant RBIR [58], adaptive region matching (ARM) [13], SIMPLIcity [4], robust local binary pattern (RLBP) [59] and MN-MIN [44]. The specific settings for each method can be found in the related references.

From Tables 3 and 4 (the data in bold), and Fig. 12 we can see that our RBIR performs better than all other 6 methods on 10 categories of Corel-1000 and Caltech-256 databases. Our RBIR(SADCT) outperforms all the others methods (RBIR and CBIR), also the RBIR(SADCT) works better than region-by-region matching.

We can conclude, from the experimental results, that the proposal improves the performance on Corel-1000 database. Furthermore, fewer AC coefficients and fewer number of bins of histogram are used leading to lower computation time.

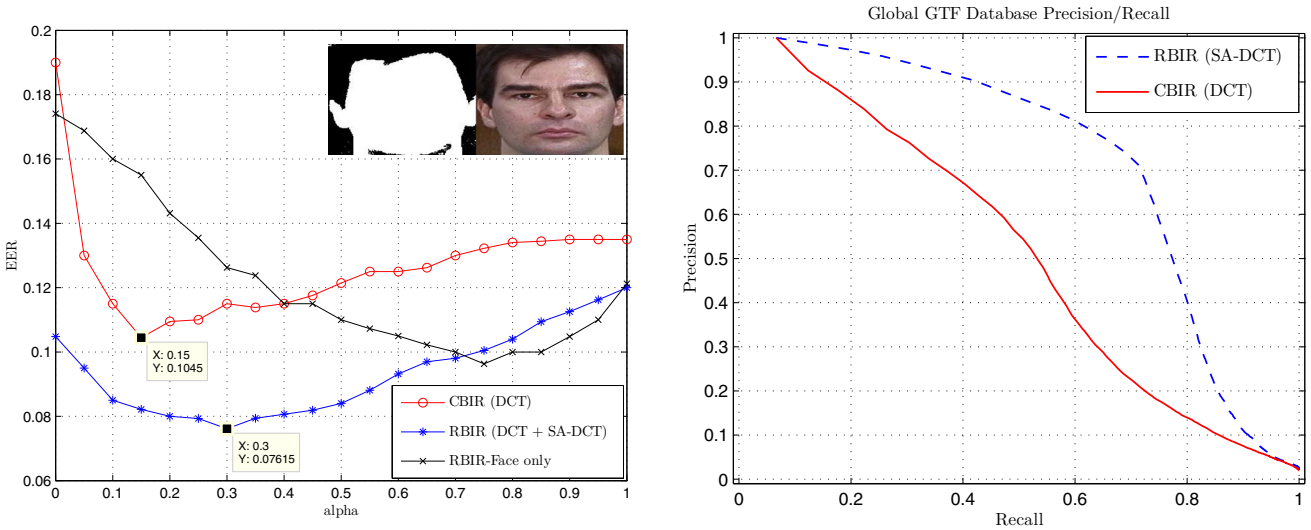
4.4 Results with GTF face database

Equal error rate (EER) [61] is often used to evaluate the performance of face recognition algorithm. So considering a query image belonging to class A, two things could occur: on one hand, it could be recognized rightly; on the other hand, it could be falsely rejected from class A, then the ratio of how many images of class A are in this situation is called false rejected rate (FRR). In contrast, considering a query image out of class A, when it is compared with the images of class A, it could be rejected rightly or it could be falsely accepted as class A, so the ratio of how many images of other classes are in this situation is defined as false accept rate (FAR). When FRR and FAR take equal values, an equal error rate (EER) is got. The lower the EER is, the better is the performance of the system, as the total error rate is the sum of FAR and FRR. Finally, we use the descriptor of the AC-Patterns and DC-Patterns together to do face recognition. The weight parameter α [see Eq. (2)] is changed to see the global comparison of the performance, as shown in Fig. 14a for GTF database. The best performance, with EER $= 0.076$, is also obtained by using the global combined descriptor [Eq. (4)] of the foreground (face) and the background together. The precision-Recall curves in Fig. 13 and Fig. 14b confirms the results obtained by EER. So, the proposed approach improves the performance on GTF database also. Furthermore, fewer AC coefficients and fewer number of bins of histogram are used.



(a) *The P-R curves using "People33" from GTF Database* **(b)** *The P-R curves using "People50" from GTF Database*

Fig. 13 The average Precision–Recall curves between our region-based approach (*with mask*) and the conventional content-based approach (*without mask*). The image categories **a** "People33" and **b** "People50" are from GTF database



(a) *The Equal Error Rate (EER)* **(b)** *The Precision-Recall curves*

Fig. 14 The comparison of the results on GTF database: **a** EER and **b** the average Precision–Recall curves

4.5 Robustness evaluation

In this section, we analyze the robustness of our RBIR system based on SA-DCT transform. Figure 15 illustrates the results on Corel-1000 and Caltech-256 databases. These subfigures in the first and the third colones of Fig. 15 show the changing rank of the target image in the retrieved images, and in the second and fourth colones the Chi-Distance Eq. (3) between altered images and target images as we increase the significance of image variations. It is illustrated in Fig. 15 that as we increase brighten/darken percentage, size of Gaussian filter window and rotated angle to certain degree, the rank of the target image still be in a strong position and the distance between altered image and target image is still small. Therefore our system is robust with respect to different intensity variation, blur variation and rotation. Experimental results in Fig. 16 demonstrate some query example.

4.6 Computational cost

All experiments in this paper have been implemented on an Intel (R) Xeon (R) Quad-Core X5650 (2.67 and 2.66 GHz, two processors, 12 GB RAM), 64 bit Windows 7 operating system in MATLAB R2009 environment. We consider the classical retrieval scenario, in which the feature rep-

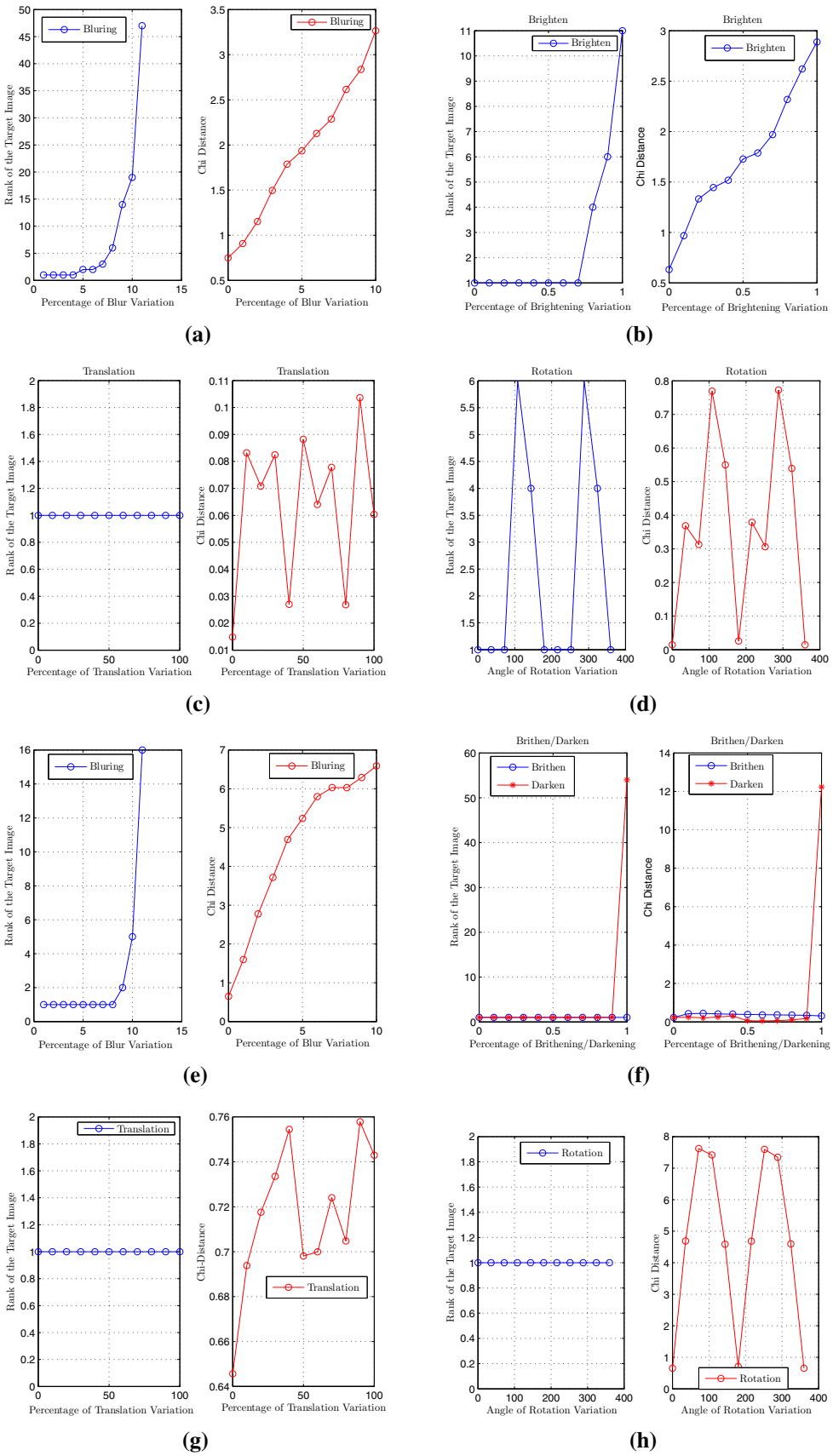


Fig. 15 Robustness of the system to image alteration: the curves **a**–**d** are for Corel-1000 database, and the curves **e**–**h** are for Caltech-256 database

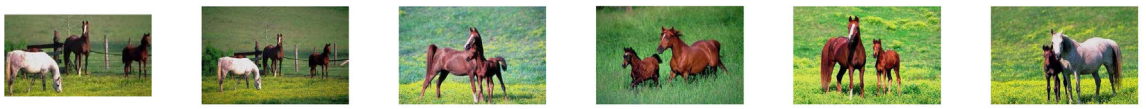
(a) *Darkening*

(b) *Brightening*

(c) *Gaussian Filtering (Blurring)*

(d) *Sharpening*

(e) *Rotation with 45 degrees*



(f) *Cropping*

(g) *Translation*

Fig. 16 Some examples of robustness evaluation. The first one is query image, top five matches followed

resentation of all images in the database are calculated off-line and the feature representation of query image is calculated on the fly when the retrieval is executed. We measure the average running time of each query. In general, one second is needed to segment an image. Note that the running time of feature representation includes DCT and SA-DCT transformations, feature vector construction and histogram construction is about 36 s per image and it remains nearly stable. The matching speed is very fast. When the query image is in the database, it takes about 1 s of CPU time on average to sort all the images in the 1000 image database using the Chi-Distance similarity measure.

5 Conclusion

Based on the conventional CBIR techniques in DCT domain, a new RBIR system based on SA-DCT is proposed to improve the retrieval accuracy. Consequently, one can retrieve a region without reference to information about other regions of the image. Clearly, this permits interesting operations such as object-based querying. Since an automatic computation of semantically meaningful object is extremely difficult, our approach, by exploiting a prior segmentation, can delegate the segmentation to sophisticated, high performance, and necessarily CPU-intensive algorithms that can be executed off-line. The proposed method has the power of capturing both local features and global features, and making use of both semantic features and low level features. The experimental results indicate its correctness and efficiency, and a compromise between high retrieval ratio and less complexity. In the future, we will enhance our technique for images which have more than two regions (objects). By constructing region importance index [13], this will reduces the adverse effect of inaccurate segmentation.

Acknowledgments This work is currently supported by the Partenariat Hubert Curien PHC-TASSILI under Grant No. 12MDU864. The authors thank for their financial supports. We would like to thank the editor and anonymous reviewers for insightful comments and helpful suggestions to improve the quality of the paper.

References

1. Thomee B, Lew M (2012) Interactive search in image retrieval: a survey. Int J Multimed Inf Retr 1:71–86
2. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv 40(2):5:1–60
3. Huang W, Gao Y, Chan KL (2010) A review of region-based image retrieval. J Signal Process Syst 59:143–161
4. Wang JZ, Li J, Wiederhold G (2001) Simplicity: semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 23(9):947–963
5. Agarwal M, Maheshwari R (2012) A trous gradient structure descriptor for content based image retrieval. Int J Multimed Inf Retr 1(2):129–138
6. Murala S, Maheshwari RP, Balasubramanian R (2012) Directional local extrema patterns: a new descriptor for content based image retrieval. Int J Multimed Inf Retr 1(3):191–203
7. Bai C, Zhang J, Liu Z, Zhao WL (2014) K-means based histogram using multiresolution feature vectors for color texture database retrieval. Multimed Tools Appl 69(5):1–20
8. Wang XY, Zhang BB, Yang HY (2014) Content-based image retrieval by integrating color and texture features. Multimed Tools Appl 68(4):545–569
9. Sun Y, Ozawa S (2005) Hirbir: a hierarchical approach to region-based image retrieval. Multimed Syst 10(6):559–569
10. Liu Y, Zhang DS, Lu G, Ma WY (2007) A survey of content-based image retrieval with high-level semantics. Pattern Recogn 40(1):262–282
11. Liu Y, Zhang DS, Lu G (2008) Region-based image retrieval with high-level semantics using decision tree learning. Pattern Recogn 41(1):2554–2570
12. Liu Y, Chen X, Zhang C, Sprague A (2009) Semantic clustering for region-based image retrieval. J Vis Commun Image Represent 20:157–166
13. Yang X, Cai L (2014) Adaptive region matching for region-based image retrieval by constructing region importance index. IET Comput Vis 8(2):141–151
14. Jing F, Li M, Zhang H, Zhang B (2004) An efficient and effective region-based image retrieval framework. IEEE Trans Image Process 13(5):699–709
15. Shokoufandeh A, Keselman Y, Demirci MF, Macrini D, Dickinson S (2012) Many-to-many feature matching in object recognition: a review of three approaches. IET Comput Vis 6(6):500–513
16. Zou W, Kpalma K, Ronsin J (2012) Semantic image segmentation using region bank. In: Proceedings ICPR'12 (international conference on pattern recognition), pp 922–925
17. Schneier M, Abdel-Mottaleb M (1996) Exploiting the jpeg compression scheme for image retrieval. IEEE Trans Pattern Anal Mach Intell 18(8):849–853
18. Eickeler S, Muller S, Rigoll G (2000) Recognition of jpeg compressed face images based on statistical methods. Image Vis Comput 18(4):279–287
19. Ngo C, Pong T, Chin R (2001) Exploiting image indexing techniques in dct domain. Pattern Recogn 34(9):1841–1845
20. Climer S, Bhatia SK (2002) Image database indexing using jpeg coefficients. Pattern Recogn 35(11):2479–2488
21. Dabbaghchian S, Ghaemmaghami M, Aghagolzadeh A (2010) Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology. Pattern Recogn 43:1431–1440
22. Cheng K, Law N, Siu W (2010) Fast extraction of wavelet-based features from jpeg images for joint retrieval with jpeg2000 images. Pattern Recogn 43:3314–3323
23. Jiang J, Amstrong A, Feng G (2002) Direct content access and extraction from jpeg compressed images. Pattern Recogn 35(11):2511–2519
24. Feng G, Jiang J (2003) Jpeg compressed image retrieval via statistical features. Pattern Recogn 36(4):977–985
25. Chang C, Chuang J, Hu Y (2004) Retrieving digital images from a jpeg compressed image database. Image and Vis Comput 22(6):471–484
26. Zhong D, Defee I (2005) Dct histogram optimization for image database retrieval. Pattern Recogn Lett 26(14):2272–2281
27. Bai C, Kpalma K, Ronsin J (2012) Color textured image retrieval by combining texture and color features. In: Proceedings EUSIPCO'12 (European signal processing conference), pp 170–174

28. Edmundson D, Schaefer G (2012) Fast jpeg image retrieval using optimised huffman tables. In: Proceedings ICPR'12 (international conference on pattern recognition), vol. IV, pp 3188–3191
29. Edmundson D, Schaefer G, Celebi M (oct 2012) Robust texture retrieval of compressed images. In: Proceedings ICIP-12 (IEEE international conference on image processing), vol. IV, pp 2421–2424
30. Zhong D, Defee I (2007) Performance of similarity measures based on histograms of local image feature vectors. Pattern Recogn Lett 28(15):2003–2010
31. Liu Y, Zhou X, Ma WY (2004) Extraction of texture features from arbitrary-shaped regions for image retrieval. In: Proceedings ICME'04 (international conference on Multimedia and Expo), pp 1891–1894
32. Zhang D, Islam M, Lu G, Sumana I (2012) Rotation invariant curvelet features for region based image retrieval. Int J Comput Vis 98(2):187–201
33. Sikora T, Makai B (1995) Shape-adaptive DCT for generic coding of video. IEEE Trans Circ Syst Video Technol 5(1):59–62
34. ISO/IEC JTC1/SC29/WG11 (1997) MPEG-4 video verification model version 8.0. MPEG97/N1796
35. Belloulata K, Belhallouche L, Belalia A, Kpalma K (2014) Region based image retrieval using shape-adaptive dct. In: Proceedings ChinaSIP-14 (2nd IEEE China summit and international conference on signal and information processing), pp 470–474
36. Jiang J, Feng G (2002) The spatial relationship of dct coefficients between a block and its sub-blocks. IEEE Trans Signal Process 5(11):1160–1169
37. Bai C, Kpalma K, Ronsin J (2012) A new descriptor based on 2d dct for image retrieval. In: Proceedings VISAPP'12 (international conference on computer vision theory and applications), pp 714–717
38. Zhong D, Defee I (2008) Face retrieval based on robust local features and statistical-structural learning approach. In: EURASIP journal on advances in signal processing, vol 2008, no. ID 631297, p 12
39. Ferman A, Tekalp M, Mehrotra R (2002) Robust color histogram descriptors for video segment retrieval and identification. IEEE Trans Circ Syst Video Technol 11(5):497–508
40. Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric for image retrieval. Int J Comput Vis 40:99–121
41. Liu Y, Zhang DS, Lu G, Ma WY (2006) Study on texture feature extraction in region-based image retrieval system. In: Proceedings MMM'06 (international multimedia modeling conference), pp 264–271
42. Chen H, Civanlar M, Haskell B (1994) A block transform coder for arbitrarily-shaped image segments. In: Proceedings ICIP-94 (iEEE international conference on image processing), pp 85–89
43. Stasinski R, Konrad J (1999) A new class of fast shape-adaptive orthogonal transforms and their application to region-based image compression. IEEE Trans Circ Syst Video Technol 9(1):16–34
44. Carson C, Belongie S, Greenspan H, Malik J (2002) Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Trans Pattern Anal Mach Intell 24(8):1026–1038
45. Bresson X, Esedoglu S, Vandergheynst P, Thiran J, Osher S (2007) Fast global minimization of the active contour/snake model. J Math Imag Vis 28(2):151–167
46. Zou W, Kpalma K, Ronsin J (2012) Semantic segmentation via sparse coding over hierarchical regions. In: Proceedings ICIP-12 (IEEE international conference on image processing), pp 2577–2580
47. Zou W, Kpalma K, Ronsin J (2013) Automatic foreground extraction via joint crf and online learning. Electron Lett 49(18):1140–1142
48. Sikora T (1995) Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments. Signal Process Image Commun 7(4–6):381–395
49. Gilge M, Engelhardt T, Mehlan R (1989) Coding of arbitrarily shaped image segments based on a generalized orthogonal transform. Signal Process Image Commun 1(2):153–180
50. Hsu H, Lee K, Chang N, Chang T (2008) Architecture design of shape-adaptive discrete cosine transform and its inverse for mpeg-4 video coding. IEEE Trans Circ Syst Video Technol 18(3):375–386
51. Belloulata K, Konrad J (2002) Fractal image compression with region-based functionality. IEEE Trans Image Process 11(4):351–362
52. Belloulata K, Belalia A, Zhu S (2014) Object-based stereo video compression using fractals and shape-adaptive dct. Int J Electron Commun 68(7):687–697
53. Kauff P, Schüür K (1998) Shape-adaptive DCT with block-based DC separation and ΔDC correction. IEEE Trans Circ Syst Video Technol 8(3):237–242
54. http://wang.ist.psu.edu/docs/related.shtml/test1.tar. Accessed Jan 2013
55. http://www.vision.caltech.edu/ImDatasets/Caltech256/. Accessed March 2014
56. http://www.anefian.com/research/face_reco.htm. Georgia tech, GTF database. Accessed March 2012
57. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset pp 1–20. http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001. Accessed Apr 2007
58. Manipoonchelvi P, Muneeswaran K (2014) Significant region-based image retrieval. In: Signal image and video processing, no. 6, Springer, New York, pp 1–8
59. Murala S, Wu QM (2014) Expert content-based image retrieval system using robust local patterns. J Vis Commun Image Represent 25:1324–1334
60. Yanping D, Wang JZ (2001) A scalable integrated region-based image retrieval system. In: Proceedings ICIP-01 (IEEE international conference on image processing), vol. I, pp 22–25
61. Bolle RM, Pankanti S, Ratha NK (2000) Evaluation techniques for biometrics-based authentication systems (frr). In: Proceedings ICPR'00 (international conference on pattern recognition, vol. II, pp 831–837