ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 13, No. 3, 2018, *pp*.201-211

A Salient Object-Based Image Retrieval Using Shape and Color Features

Shuxian Huang, Wenbing Chen
Nanjing University of Information Science and Technology, Nanjing 210044, China

(Received January 17 2018, accepted July 05 2018)

Abstract. In this paper, a salient object-based image retrieval method (SOBIR) is presented, which linearly combines the shape and colour features of the salient objects contained in target and candidate images respectively to carry out content-based image retrieval (CBIR). The framework of the proposed method is carried out as follows: first, the mean shift and region growing algorithms are used to segment an input image into many regions; secondly, based on these regional contrasts the saliency map, the binary image, and the salient object image are extracted respectively; thirdly, the shape representation of the salient object is extracted from the binary image using an improved polar Fourier Descriptor method, meanwhile the salient object contained in the input image is converted into a representation of its histogram in the L∗a∗b∗ colour space; Finally, the similarity between the two salient objects contained in the target and candidate images is defined by linearly combining both the shape and colour representations. Experimental results show that, compared to the latest two CBIR methods, the proposed SOBIR method exhibits an excellent performance in precision, recall, flexibility and efficiency.

Keywords: Image retrieval; salient object; region; shape; object detection; similarity measure.

1 Introduction

Colour is intuitive, stable, simple and popular feature to represent, analyze and recognize an image. Since Swain and Ballard [1] proposed the histogram intersection method used for image retrieval, a variety of histogram-based image retrieval methods have been continuously proposed [2, 3, 4, 5, 6, 7]. However, Figure 1 shows that a white cup and a white dish can not be correctly identified using single colour feature. Therefore, with using the histogram-based method, it seems to be difficult to correctly distinguish the two salient objects like visual perception. To solve the issue, we present a SOBIR method combining colour and shape features, intuitively which should be more exact and efficient than simply using colour feature. As Figure 2 shows the framework of the proposed method.

A simple iterative procedure referred to as the mean shift (MS) method [11, 12], which shifts each data point to the average of data points in its neighborhood, is generalized and used to smooth and cluster or segment an image. Subsequently the MS method has been widely used to smooth and segment an image [13, 14, 15, 16, 17] since it is simple, highly efficient and has no parameters. In more recent years, image segmentation methods are still being developed rapidly. Gong et.al.[18] proposes a fuzzy c-means clustering and kernel metric-based image segmentation method, which uses an improved fuzzy C-means algorithm with a tradeoff weighted fuzzy factor and kernel metric to segment an image. Although this method exhibits a better performance, one main drawback is that an input image has to be segmented into $c(c \geq 2)$ classes beforehand.

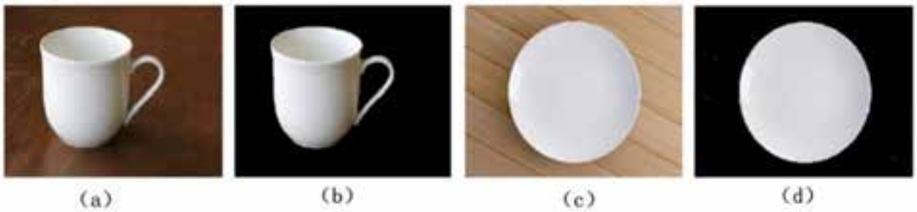

Figure 1: Two salient objects contained in the two images seem to be difficult to be identified by simply using the histogrambased CBIR. (a) and (c) are two original images; (b) and (d) are the two salient objects extracted from the corresponding original image respectively.

For salient object detection and extraction, in recent years a few popular methods have been proposed, such as methods based on low-level features of luminance and colour [8] and based on frequency-tuned salient region [9]. The regional contrast-based saliency extraction method [10], simultaneously evaluates global contrast differences and spatial coherence and yields full resolution saliency maps. However, within this prominent method, there exist two shortages: the first is a slow speed due to computing global contrast between regions; the second is it's effectiveness, which is suppressed for saliency detection and extraction since the contribution of colour similarity between involved two regions to saliency has not been highlighted. Therefore, We investigate deeply the method proposed in [9] and use MS and region growing (MSRG) to partition an image into many regions rapidly instead of Graph-cut.

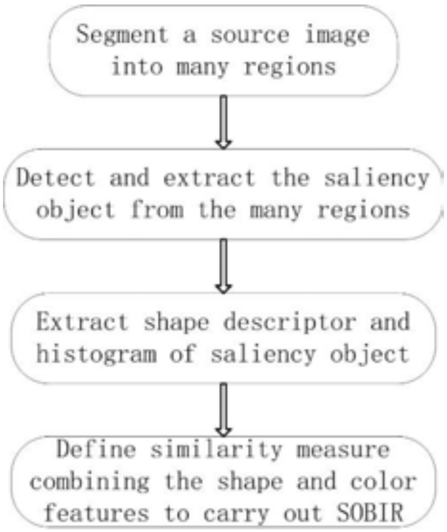
Segment a source image into many regions
Detect and extract the saliency object from the many regions
Extract shape descriptor and histogram of saliency object
Define similarity measure combining the shape and color features to carry out SOBIR

Figure 2: A salient object-based image retrieval framework.

Various shape descriptors exist in many literatures[21, 22], these descriptors are broadly categorized into two groups: contour-based and region-based shape descriptors. Zhang and Lu [21] proposed a region-based Fourier shape descriptor, referred to as GFD, and confirmed its advantage over other relevant shape descriptors. However, since there exists a high computational complexity with GFD, in this paper an improved GFD, referred to as PFD, is presented and used as shape descriptor.

Overall, in this paper, we show a new perspective, which carries out CBIR with using two salient objects contained in the target and candidate images repectively. The scheme is organized as follows:

- As Figure 3 shows, the MSRG is used to detect and extract the salient object and yield one binary image as Figure 3(c), and one salient object image as Figure 3(d).
- Against the binary image as Figure 3(c), we use PFD to extract the shape descriptor of the salient object; against the salient object image as Figure 3(d) we use the 10×3×3 quantization histogram of $L^*a^*b^*$ colour space as its colour descriptor.

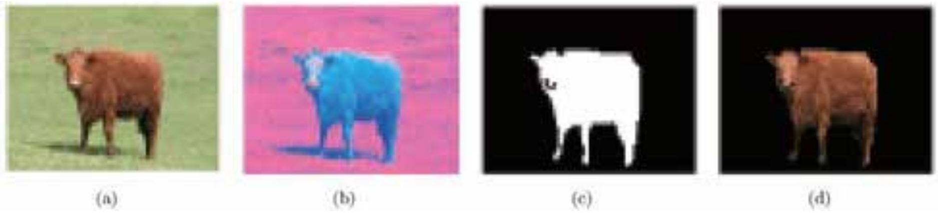

Figure 3: An example using MSRG to detect and extract. (a) The original image; (b) the segmented image; (c) the binary image; (d) the salient object image.

We define a similarity measure based on the linear combination of shape and colour features between the target and candidate salient objects.

- The candidate images are ranked in ascending order by their similarity measures.

The rest of the paper is organized as follows. In Section 2, image segmentation and salient object extraction are described. In Section 3, extraction and representation of the salient object are discussed. A similarity measure between two salient objects is defined in Section 4. In Section 5, evaluation and experiment comparisons are conducted.

2. Image segmentation and salient object extraction

The general process of the MS filtering and segmenting is shown in [11]. The MS filtering algorithm is used to quantize the colour space so that the smoothed image is formed by some homogeneous colour regions. Afterward, based on the smoothed image, we use the region growing algorithm to over-segment the smoothed image into such regions that $I_s = \bigcup_i R_i$, and, $R_i \cap R_j = \emptyset$ while i ≠ j.

2.1. Mean shift filtering

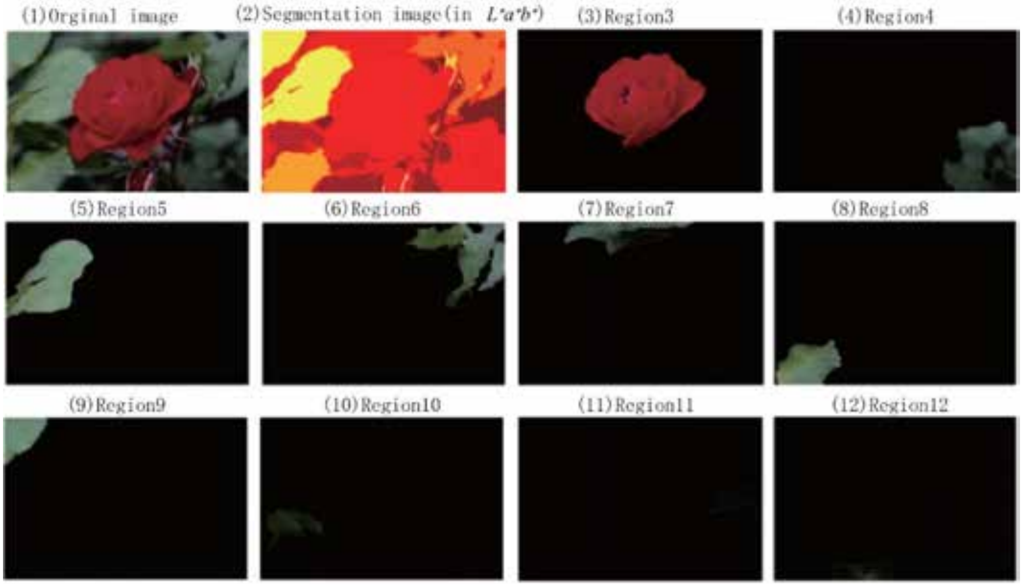

Figure 4: A rough segmentation generated by using the mean shift filtering with h = (6; 32) and region growing algorithms with Tr = 7. (1) The original image; (2) the segmented image in L*a*b* space; (3)-(12) the top 10 segmented regions ranked by regional sizes.

An image *I* is typically represented two dimensional lattice of p-dimensional vector, where *p* = 1 in the grey-level case, *p* = 3 for colour images [11]. The space of the lattice is known as the spatial domain, while the grey-level or colour feature is represented in the range domain. For both domains, Euclidean metric is used. When the location and range vectors are concatenated in the joint spatial-range domain of dimension *d* = 2 + 3 ,their different nature has to be compensated by proper normalization. Thus, the multivariate kernel is defined as the product of two radically systemic kernels and the Euclidean metric allows a single bandwidth parameter for each domain or kernel.

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^3} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{1}$$

where x^s is the spatial part, $x^r$is the range part of a feature vector of *I* . *k*(*x*) the common profile used in both two domains, h_s and h_r the used kernel bandwidths, and *C* the corresponding normalization constant. In this paper, the gaussian kernel is used as *k*(*x*), the bandwidth parameter *h* = (h_s, h_r) is the only parameter which has to be set and able to determine the resolution of image segmentation, i.e., a small (h_s, h_r) leads a fine segmentation or over-segmentation, vice and versa.

Let x_i and z_i, *i* = 1, · · · , *n*, be the *n*-dimensional input *I* and filtered image I_f pixels in the joint spatialrange domain.

1. For each *i*, *i* = 1, · · · , *n*
2. Initialize *j* = 1 and $y_{i,j}$=x_i.
3. Compute $\boldsymbol{y}_{i;j+1}$ according to (3) until convergence, $\boldsymbol{y} = \boldsymbol{y}_{i;c}$

$$y_{i,j+1} = \frac{\sum x_i K_{h_s,h_r}\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum K_{h_s,h_r}\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} \tag{2}$$

4. Assign $z_i = (x_i^s, y^r)$
5. until *i* = *n*, otherwise go to step 1.

The assignment specifies that the filtered data at the spatial location x_i^s will have the range component of the point of convergence *y* and the filtered image I_f is gained.

2.2. Image segmentation

The region growing algorithm with threshold T^r is employed to partition I_f into *N* homogeneous regions, i.e., $I_f = \bigcup_{i=1}^N R_i$, and, $R_i \cap R_j = \emptyset$ while i ≠ j, and the segmented image is denoted as I_s. For each region R_k of I_s, *L* is used to represent the region-classified label matrix such that *L*(*i, j*) is assigned as *k* while the pixel *p*(*i, j*) of $I_f$belongs to R_k. Therefore, region R_k can be readily extracted from the segmented image *Is* by

$$L_k(i,j) = \begin{cases} 1\ if\ L(i,j) == k \\ 0\ otherwise \end{cases} \tag{3}$$

and

$$R_k = I_s \cdot * L_k$$

where L_k is the mask matrix to represent region R_k.

As Figure 4 shows, an input image is partitioned into many regions by the mean shift filtering with ***h*** = (6, 32) and the region growing algorithms with $T^r = 7$, and the top 10 regions ranked by the region size are extracted and displayed.

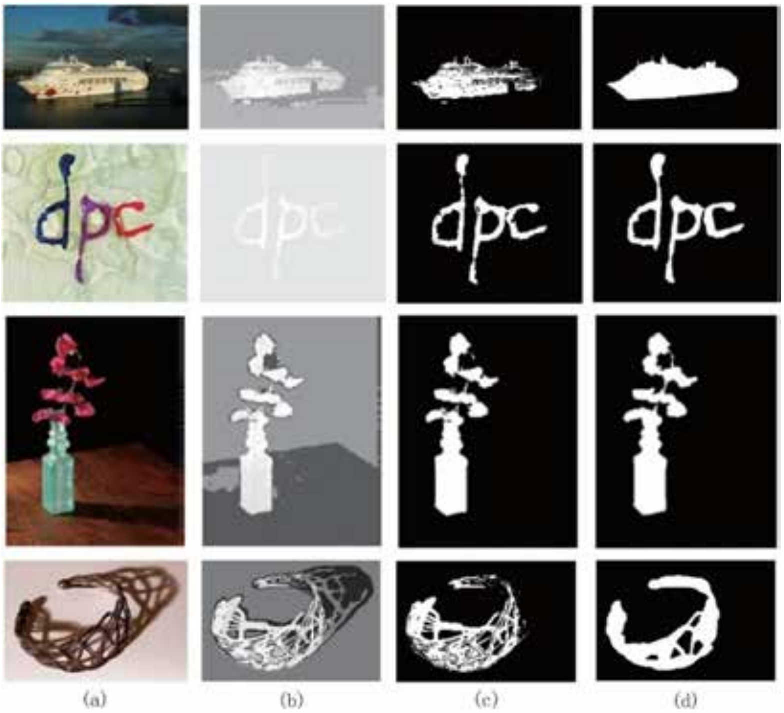

Figure 5: A demonstration extracting salient object. (a) The original image; (b) the saliency map; (c)the extracted and binarized salient object; (d) Ground True.

2.3. Salient object detection and extraction

Our saliency detection and extraction method is investigated based on the three basic facts of visual perception as follows:

1) for a region belonging a salient object, there exists a strong colour contrast to its surrounding;

2) for two regions belonging to the same object, there exist not only similar colours but also a nearer distance between them;

3) usually, a salient object is located near image center rather than image boundary.

According to the facts above and if an image *I* is partitioned into *N* regions, i.e., $\mathrm{I} = \{\bigcup_{k=1}^{N} R_k | R_i \cap R_j = \emptyset, i \neq j\}$, our saliency for region *Ri* can be defined as

$$S_i = C_i^g \omega_i^{cs} \sum_{j=1}^{N} \omega_{ij}^{cp} \tag{4}$$

where C_i^g is the color contrast between the average colours of region R_i and the image *I* and simply defined as

$$C_i^g = \|c_i^r - c^g\|_2 \tag{5}$$

where c_i^r denotes the average colour of region R_i, which is calculated as $c_i^r = \frac{\sum_{I_k^c \in R_i} I_k^c}{|R_i|}$. c^g denotes the average colour of the global image, which is calculated as $c^g = \frac{\sum_{k=1}^{|I|} I_k^c}{|I|}$. I_k^c is 3D vector in $L^*a^*b^*$ colour space, | • | denotes the number of pixels in a region or image.

Centre weight ω_i^{cs} is used to highlight the contribution of the proximity of region R_i and the image centre to the saliency S_i, and the exponent function can be used to approximately reflect its change, i.e., $\omega_i^{cs} = e^{-\lambda D(p_i^r, p_c^I)}$, where p_i^r is the centre of region R_i and calculated by $p_i^r = \frac{\sum_{I_k^p \epsilon R_i} I_k^p}{|R_i|}$, and p_c^I is the centre of image *I* and calculated by $p_c^I = \frac{\sum_{I_k^p \epsilon I} I_k^p}{|I|}$.

We have carried out a large number of experiments and conclude that a suitable λ should be taken as 0.04.

ω_{ij}^{cp} is used to reflect fact 2), .i.e., for two regions belonging to the same object, there exists not only similar colours but also a nearer distance between them. Similar to center weight ω_i^{cs}, the relation of colours and positions between two regions can be also described by the exponent function as

$$\omega_{ij}^{cp} = e^{\left(-\left\|c_i^r - c_j^r\right\|_2 \left\|p_i^r - p_j^r\right\|_2\right)} \tag{6}$$

Figure 5 shows the saliency map, binary image and saliency image generated by the MSRG method, which exhibit good effectiveness.

3. Shape extraction and description of a salient object

In MPEG-7, shape is one of the key components for a describing digital image along with other features such as texture and colour. Zhang and Lu[21] proposed a generic Fourier descriptor (GFD), which is applied to the region and obtains a good effect. However, this method has a high computational complexity, which limits its popularity. To address this issue, two improvements for the GFD are carried out: the first is the GFD formulation which is remedied to promote the accuracy of the salient object shape description; the second is a regional lattice-based GFD which is proposed to largely reduce the high computational complexity of the GFD. The improved GFD is referred to as PFD.

3.1. Fourier transform over the polar coordinate system

Let us set $f(x, y)(0 \leq x \leq M,\ 0 \leq y \leq N)$ to be the input image, then its continuous and discrete Fourier transforms in the 2-D Cartesian space are given respectively by equations (7) and (8) .

$$\mathrm{F}(\mu, v) = \int_x \int_y f(x,y)\exp\big(-i2\pi(\mu x + vy)\big)dxdy, \qquad (7)$$

$$\mathrm{F}(\mu, v) = \textstyle\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)exp\big(-i2\pi(\mu x/M + vy/N)\big), \qquad (8)$$

Since the FT over the polar system produces rotation-invariant data which are particularly well suited for the accurate extraction of orientation feature. We put both the input image $I(x, y)$ and its spectra $F(u, v)$ over the polar system; let

$$x = rcos\theta,\ y = rsin\theta \qquad (9)$$

$$\mu = \rho cos\theta,\ v = \rho sin\theta \qquad (10)$$

where (r, θ) are the polar coordinates over the image plane and (ρ, ϕ) are the polar coordinates over the frequency plane. The transformation between $dxdy$ and $drd\theta$ is given by

$$dxdy = rdrd\theta \qquad (11)$$

Then the PFT is calculated by substituting x, y, μ, v of equation (7) with equations (9) and (10) respectively as follows:

$$\begin{aligned} PF(\rho,\phi) &= \int_r\int_\theta f(rcos\theta, rsin\theta)\exp(-i2\pi(\rho cos\phi rcos\theta\ +\ \rho sin\phi rsin\theta)rdrd\theta \\ &= \int_r\int_\theta rf(r,\theta)exp(-i2\pi\rho r(cos\phi cos\theta\ +\ sin\phi sin\theta))drd\theta \\ &= \int_r\int_\theta rf(r,\theta)exp\Big(-i2\pi\rho r\big(cos(\phi\ -\ \theta)\big)\Big)drd\theta \qquad (12) \end{aligned}$$

However, equation (11) in [21] is incorrectly derived as

$$= \int_r\int_\theta rf(r,\theta)exp(-i2\pi\rho r(sin(\phi\ +\ \theta)))drd\theta$$

The equation (11) is discretized as

$$\mathrm{PF}(\rho_s,\phi_t) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} r_m f(r_m,\theta_n)exp\big(-i2\pi r_m\rho_s cos(\theta_n - \phi_t)\big) \qquad (13)$$

$$\mathrm{s} = 0,1,\cdots,\mathrm{S}-1; \quad \mathrm{t} = 0,1,\cdots,\mathrm{T}$$

where M and N are the space resolutions along the radial frequency and angular frequency in polar coordinates respectively. S and T are the spectra resolutions along the radial frequency and angular frequency in polar coordinates respectively.

3.2. Extraction the shape Polar Fourier descriptor

In order to preserve the rotation invariant, the FD2 method proposed in [21], firstly maps the input image in the Cartesian space to the polar space by changing equation (12) as follows:

$$\mathrm{PF}(\rho_s,\phi_t) = \textstyle\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} r_m f(r_m,\theta_n)exp\left(-i2\pi\left(\frac{r_m}{R} + \frac{2\pi\phi_t}{T}\right)\right) \qquad (14)$$

$$\mathrm{s} = 0,1,\cdots,\mathrm{S}-1; \quad \mathrm{t} = 0,1,\cdots,\mathrm{T}$$

However, the proposed PFD method is remedied in the equation above as

$$\mathrm{PF}(\rho_s,\phi_t) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} r_m f(r_m,\theta_n)exp\left(-i2\pi\left(\frac{r_m}{R} - \frac{2\pi\phi_t}{T}\right)\right) \qquad (15)$$

$$\mathrm{s} = 0,1,\cdots,\mathrm{S}-1; \quad \mathrm{t} = 0,1,\cdots,\mathrm{T}$$

Moreover, our polar coordinate system is built by using the shape regional centroid (x_c, y_c) as the original point. While the shape region in continuous $I(x, y)$ is homogeneous, the approximation to the centroid (x_c, y_c) is calculated by

$$\mathrm{x_c} = \frac{\int_x\int_y yI(x,y)}{\int_x\int_y I(x,y)},\quad y_c = \frac{\int_x\int_y xI(x,y)}{\int_x\int_y I(x,y)} \qquad (16)$$

For discrete $I(x, y)$, the approximation of the centroid (xc, yc) is calculated by

$$x_c = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} yI(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)},\quad y_c = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} xI(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)} \qquad (17)$$

The maximum distance from the centroid (x_c, y_c) of the shape region to the four vertexes of the input image is taken as the maximum radius R. The main difference from the method in [21] is that we use the latticed polar system over the shape region to approximately calculate the PFD of the shape region rather than over the whole input image, as in [21]. Later experiments show that the scheme exhibits a very high efficiency and largely reduces the time cost.

The final PFD is given by

$$\mathrm{PFD} = \left\{\frac{|PF(0,0)|}{area}, \frac{|PF(0,1)|}{|PF(0,0)|}, \cdots, \frac{|PF(0,T-1)|}{|PF(0,0)|}, \cdots, \frac{|PF(S-1,0)|}{|PF(0,0)|}, \cdots, \frac{|PF(S-1,0)|}{|PF(0,0)|}, \cdots, \frac{|PF(S-1,T-1)|}{|PF(0,0)|}\right\} \quad (18)$$

4. Similarity measure between two images

In order to promote precision and flexility of our image retrieval system, both colour and shape features are integrated into the proposed image retrieval method. The combined similarity measure is designed as follows:

We can use the MSRG to produce two resulting images, the binary image representing the corresponding salient objects and the image containing the corresponding salient object cut off from the original image. The colour histogram in $L^*a^*b^*$ colour space is used as the colour feature descriptor of a salient object, and the $L^*a^*b^*$ colour space is quantified into 10×3×3 bins along the L, a and b directions. Furthermore, the Bhattacharyya coefficient is used to measure the difference between the colour histograms. Let us set H_i and H_q to denote the colour histograms of the two objects contained in the input and query images respectively, then the Bhattacharyya coefficient $D_h(H_i,H_q)$ measuring the difference between the colour histograms is calculated by

$$D_h\left(H_i, H_q\right) = \sqrt{1 - \sum_{i=1}^{90} \frac{\sqrt{H_i(i)H_q(i)}}{\sum_{i=0}^{90} H_i(i)H_q(i)}} \quad (19)$$

where $D_h(\bullet)$ is a normalized function ranging from [0, 1]; while the colour features of two objects are completely uniform, $D_h(\bullet) = 0$ otherwise $D_h(\bullet) = 1$.

It is not viable to directly use the PFD given by equation (18) to measure the shape component in the combined similarity. We need to normalize the PFD prior to carrying out the image retrieval.

Let PFD_1 and PFD_2 be the PFDs of the input and query objects respectively in the forms as equation (19). For convenience, we write PFD_1 and PFD_2 as two vectors below.

$$PFD_1 = (PD_1(1), PD_1(2), \cdots, PD_1(N)) \quad (20)$$

$$PFD_2 = (PD_2(1), PD_2(2), \cdots, PD_2(N)) \quad (21)$$

Their magnitudes $|PFD_1|$ and $|PFD_2|$ are calculated by

$$|PFD_1| = \sqrt{(PD_1(1)^2 + PD_1(2)^2 + \cdots + PD_1(N)^2)} \quad (22)$$

$$|PFD_2| = \sqrt{(PD_2(1)^2 + PD_2(2)^2 + \cdots + PD_2(N)^2)} \quad (23)$$

Their distance between two shape descriptors is calculated by

$$D_s\left(I_i, I_q\right) = \frac{\sqrt{\left(PD_1(1)-PD_2(1)\right)^2 + \cdots + \left(PD_1(N)-PD_2(N)\right)^2}}{max\{|PFD_1|,|PFD_2|\}} \quad (24)$$

where N is the first N components of the PFD. It is easy to see that $D_s(I_i, I_q)$ is a normalized parameter ranging between [0, 1]. Where the shapes of two objects are completely similar $D_s(I_i, I_q) = 0$, otherwise $D_s(I_i, I_q) \approx 1$. Later experiments show that the shape measure is corrective and very effective.

We use the linear combination of the shape and colour features of a salient object to match the input and query images in our image retrieval system. The similarity measure $D_c(I_i, I_q)$ for the linear combination of shape and colour features is defined by

$$D_c(I_i, I_q) = \lambda D_h(H_i,H_q) + (1 - \lambda)D_s(I_i, I_q) \quad (25)$$

where λ is the weighting factor in the range of [0, 1]. The proposed method stresses the colour feature if λ is smaller, otherwise it highlights the shape feature. In our later experiments, we assess the effectiveness caused by changing λ.

5. Experiments

Experiments are divided into three parts. Firstly, we survey the efficiency of the MSRG. Secondly, a comparison between the PFD and GFD is carried out. Finally, the colour-shape-based image retrieval method, referred to as CSIR, is evaluated against the widely used common precision and recall and compared with the latest two relevant methods [6, 7].

The widely used common benchmark databases, such as the MSRA10K Database, referred to as MSRADB, of [10], CorelDB of [22], MPEG-7 Shape Set B, referred to as Set B, and Leaf database [20], are used as evaluation datasets. MSRADB contains 10,000 images with consistent bounding box labeling in it, therefore it has a ground true database. CorelDB manually divided 10,800 images from the Corel Photo Gallery into 80 concept groups; each group includes more than 100 images and the images in each group are category-homogeneous; Set B consists of 1400 shapes of 70 groups, and there are 20 similar shapes for each group. The Leaf database contains 25 groups and about 3600 images of leaves. All methods are conducted on the HP Compaq 8000 with Dual 3.00 Intel Core2 Duo E8400 and 8GB RAM.

Let us set the precision and the recall to be P and R respectively, which are calculated according to the traditional formulae. We represent the numbers of the images in group i as g_i, the numbers of the images returned by query as r_i, and the numbers of the images returned by query and falling into g_i as s_i. Then, the precision and the recall are calculated by

$$\mathrm{P} = \frac{s_i}{r_i} \quad R = \frac{s_i}{g_i} \quad (26)$$

5.1. Evaluation of the MSRG

Two main drawbacks of the RC are its highly computational complexity and being nonadaptive for the threshold. Clearly, it is inappropriate to use it to extract the salient object during CBIR. To address the issue, we substitute the graph-cut with mean shift and 4 neighbours growing algorithms with $h_s = 3$, $h_r = 3$, *threshold* = *20* to smooth and over-segment the input image in order to reduce the time cost in the RC; and then we use the adaptive threshold equation (27) to yield the binary images from the saliency map as column (c) in Figures 8 shows; the adaptive threshold is given by

$$\text{threshold} = \frac{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1} S(x,y)}{W \times H} \tag{27}$$

where H and W are the pixel numbers of row and column of the saliency map S respectively. Furthermore, we use the binary image to cut off the corresponding salient object from the input image as column (d) in Figures 5 show.

We used MSRG and RC methods to generate the binary and salient object images for all the images in MSRADB, CorelDB and Leaf database. Table 1 shows the average time taken by each method for each database, and that the MSRG has a substantial advantage over the RC. Figure 6 depicts the average precision-recall graph of all the databases and it confirms that the MSRG slightly overtakes the RC. Overall, experiments show that the MSRG outperforms the RC in all apspects and lays the foundation for further conducting future image retrieval.

Table 1: The computational complexity between the RC and the MSRG.

Database Method	MSRADB (400×300)	CorelDB (640×480)	Leaf Database (320×250)
RC	0.213	0.580	0.150
MSRG	0.049	0.123	0.033

5.2. Evaluation of the PFD

In the subsection, we are to test the retrieval effectiveness of the proposed PFD and carry out a comparison with the GFD based on the three indicators precision, recall and computational complexity.

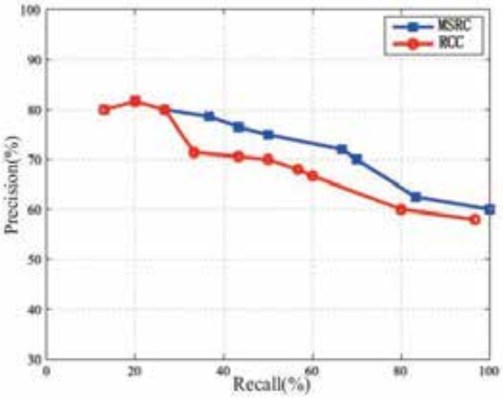

Figure 6: The average precision-recall graph using MSRG and RC methods on the databases.

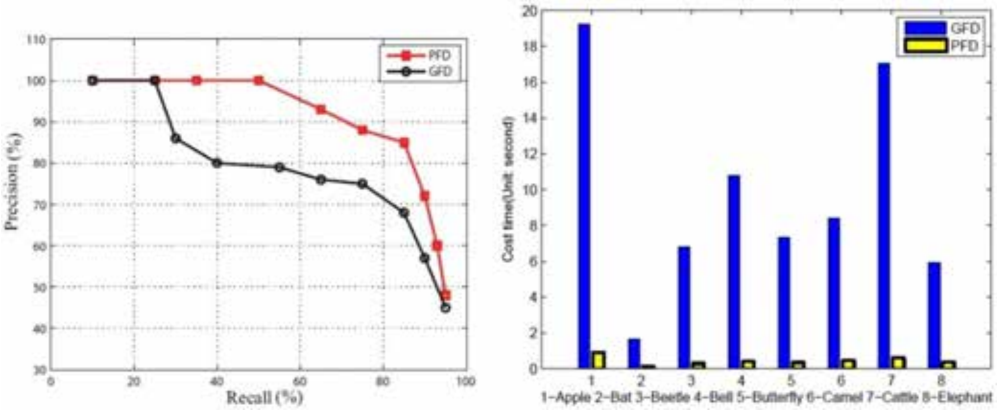

Figure 7. Left: The average precision and recall of the retrieval using GFD and PFD. Right: The average time costs taken by using both GFD and PFD for some groups in MPEG-7 Shape Set B.

Mathematically, the computational complexity of the PFD is $O(C \times S \times T)$, where C is $R \times 90$, R is the maximum radius, the angular resolutions in the space domain are set to 90 in our system, M and N are the width and height of the input image, S and T are radial and angular resolutions in spectral space respectively. But the computational complexity of the GFD is $O(M \times N \times S \times T)$. Since $C \ll M \times N$ so that $O(C \times S \times T) \ll O(M \times N \times S \times T)$; therefore the computational complexity of the PFD is far lower than that of the GFD.

The next retrieval experiments are conducted on all the shapes of Set B. The running parameters for the proposed PFD method are set as follows: The shape descriptor is represented by the top six components of the PFD; The resolutions in the space domain over the polar coordinate system: the radial resolution is set to 10 and the angular resolution is set to 15; The resolutions in the spectral domain over the polar coordinate system: the radial resolution is calculated dynamically as mentioned above and the angular resolution is set to 90 in the next experiments. All the shapes in Set B are used as queries. The common retrieval measurement precision-recall is used for evaluation of the retrieval effectiveness. The average precision and recall of the retrieval using the GFD and PFD on Set B are shown in Figure 7. In order to verify the time costs, a comparison between the PFD and GFD is carried out, and the average time costs of some groups in Set B are drawn in Figure 7.

As can be seen from Figures 7, the precision for GFD (black curve) is descending more steeply than that for PFD (red curve) when the recall percentage increases. Moreover, Figure 7 also shows that compared with the GFD, the PFD has much more lower computational complexity than that of the GFD. All experiments confirm that the PFD has an overall advantage over the GFD.

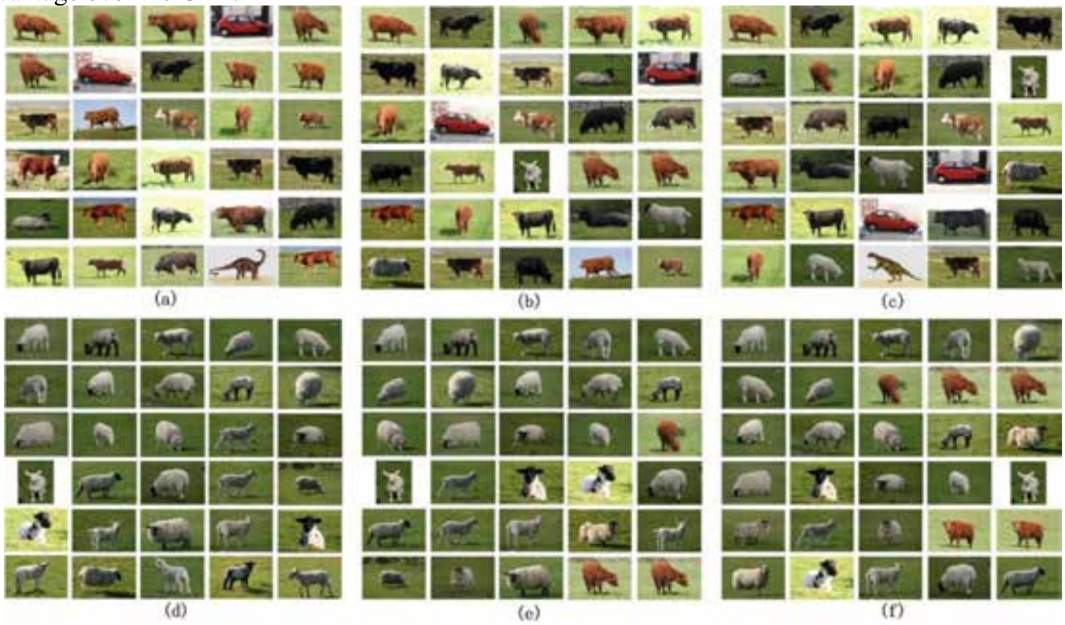

Figure 8: Some screen shots of the results returned by using different values of the weighting factor λ. (a) and (d) are the result set queried by setting λ = 0:3; (b) and (e) are the result set queried by setting λ = 0:5; (c) and (f) are the result set queried by setting λ = 0:7.

5.3 Evalution of the CSIR

We test the retrieval effectiveness of the proposed CSIR method. The experiments are organized in two parts: the first part is to assess effect of the weighting factor λ on the retrieval result; the second part is to conduct a comparison between the CSIR and the latest two relevant methods, which are the method proposed by [6], referred to as CTS, and the method proposed by [7], referred to as CDS.

5.3.1. Effect of the weighting factor λ on the retrieval result

In order to test the effectiveness of changing weighting factor λ, all experiments are focused on both the CorelDB and Leaf databases. The integrated distance measure between images is calculated by equation(25). The images queried by the CSIR are ranked in ascending order according to their distances from the input image; the higher the image the smaller the distance. For the λ in equation (25), mathematically, the larger value is to highlight the shape similarity between images.

In order to simplify the experiments, we only take into consideration over $\lambda = \{0.3, 0.5, 0.7\}$, which should be a reasonable choice since it represents the λ three changing trends against the colour, trade-off and shape features. We carried out our experiments by using the different λ values on all the 105 groups in the CorelDB and Leaf databases. We have produced very interesting and excellent results. Figure 8 shows some screen shots by using the three different λ values on two groups: cow and sheep.

As Figure 8 shows, the first image in each result set is the input image. Firstly, looking at the result sets (a) and (d), which are the result sets queried by setting $\lambda = 0.3$, it is easy to see that the images among the two sets are of high colour similarity. Focusing on result set (a), the number of brown cows among (a) is obviously greater than that of the black cows in (b) and (c). Similarly, all sheep among the result set (d) are almost white. These results indicate that the smaller λ value given highlights to a greater extent of the component of the colour feature. Conversely, looking at the result sets (c) and (f), with the larger λ value given, the image similarities are leading to the component of the shape and pose features while the component of the colour features are weak. Therefore these two result sets (c) and (f), among which there are the brown and black cows but their shapes are very similar, highlight the shape similarity. Finally, for (b) and (e) with λ = 0.5 given, the components of colour and shape features are traded off and imposed on the CSIR so that the result sets embody the similarities of these images in either colour or shape.

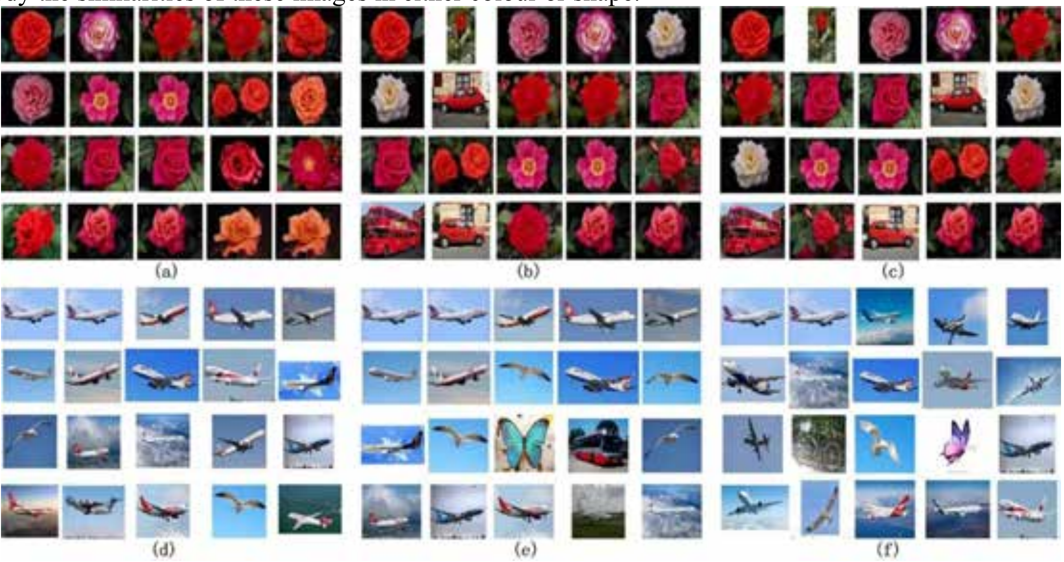

Figure 9: The three image retrieval methods are used to retrieve an image by our system, (a) and (d) the result set queried by using the CSIR; (b) and (e) the result sets queried by using the CTS; (c) and (f) the results sets queried by using the CHD.

5.3.2. Comparison to other methods

In the last experiment, we compare the CSIR with the CTS and CDH methods. The CTS is an image retrieval scheme combining all the colour, texture and shape information. The CDH defines a colour difference histogram, which counts perceptually the uniform colour difference between two points under different backgrounds with regard to colours and edge orientations in the $L*a*b*$ colour space.

We evaluate the two indicators, i.e. recall and precision, for all the three methods on the single object subset of the CorelDB and all the Leaf database. In particular, we assess the ability that each method has to discriminate the shape. We randomly pick up the input image from each group, and set the number of query images to be exhibited to 5, 10, 15, 20, 25, 30, 35,40,45,50 to observe the changes in the precision and recall. Figures 9 show some screen shots of these experiments with the numbers 20 and 50 of the query images and λ = 0.3 respectively. Looking at Figure 9, for (a) and (d), the precision of the CSIR with the weighting factor λ = 0.3 are 1.0 and 18/20 respectively; for (b) and (e), the precision of the CTS with the quantization number 10 × 3 × 3 in the $L*a*b*$ colour space, at which the CTS shows its best performance, are 17/20 and 14/20 respectively; for (c) and (f), the precision of the CHD are 17/20 and 16/20 respectively.

Table 2 shows the average recalls and precisions with respect to the 10 groups by using all three algorithms. As Table 2 shows, for the 10 groups, the overall average recall and precision of the CDH are 0.53 and 0.21 respectively, those of the CTS are 0.56 and 0.22 respectively; however those of the CSIR are 0.73 and 0.31 respectively. Moreover, Table 3 shows us the average recalls and precisions for the different numbers of the images returned by using the CSIR, CTS and CDH for the 10 groups; the average recalls and precisions of the CSIR overall outperform those of the CTS and CDH. As Table 3 also shows, in the total average precision-recall chart the blue curve of the CSIR exhibits a very obvious advantage over the red and blue curves of the CTS and CDH and the changing trend of the blue curve is more robust than the other two curves.

With the comparison to the other two methods above, the experimental results confirm that the overall effectiveness of the CSIR remarkably outperforms those of the CTS and CDH.

6. Conclusion

In this paper, a novel image retrieval method, which combines both the colour and shape features of a salient object, is proposed. Firstly, we proposed the MSRG method to detect and extract the salient object, which is represented by the binary and the salient object images. Secondly, we proposed the PFD method to describe and represent the shape features of the salient object. We also build a distance measure, based on the colour and shape descriptors, to match the input and query images. A large number of experiments confirm that our scheme is remarkable in effectiveness, robustness and computational complexity, and a further comparison with the CTS and CDH also indicates that the CSIR overall outperforms the other two methods.

In the future, we shall continue to survey all types of detection and extraction methods based on salient object and further explore the new methods and their application in practice.

Table 2: The average recall and precision using the CSIR, CTS and CDH for 10 groups in the CorelDB (single object) and LeafDatabase

Group	CSIR		CTS		CDH	
	P	R	P	R	P	R
cow	0.74	0.32	0.54	0.21	0.50	0.20
butterfly	0.76	0.32	0.52	0.21	0.48	0.19
rose	0.81	0.34	0.56	0.22	0.52	0.20
lotus	0.77	0.33	0.55	0.21	0.51	0.20
plane	0.64	0.26	0.54	0.22	0.48	0.20
sheep	0.78	0.32	0.64	0.27	0.56	0.23
ship	0.76	0.33	0.59	0.23	0.68	0.28
chicken	0.59	0.26	0.59	0.23	0.68	0.28
star	0.85	0.36	0.57	0.23	0.52	0.20
bus	0.74	0.32	0.54	0.21	0.50	0.20
average	0.73	0.31	0.56	0.22	0.53	0.21

Table 3: The average recall and precision at each result set returned by using the CSIR, CTS and CDH for 10 groups.

Method	The number of the images of each result set	5	10	15	20	25	30	35	40
CSIR	Precision(%)	85	84	79	75	71	66	63	60
	Recall(%)	8	17	24	30	35	40	44	48
CTS	Precision(%)	72	70	68	59	50	46	42	40
	Recall(%)	7	14	21	24	25	27	30	31
CDH	Precision(%)	67	68	54	48	42	42	39	37
	Recall(%)	7	14	19	21	24	25	28	30

Acknowledgements

This work was supported in part by the National Meteorology Industry Special Project of China under Grant GYHY201306043 and GYHY201406029 and the Beijige Open Fund of Jiangsu Meteorology Science Research Institute under Grant BJG201504.

7. References

[1] M. Swain, D. Ballard, Colour indexing, International journal of computer vision 7(1) (1991) 11-32.

[2] C. L. Novak, S. A. Shafer, Anatomy of a colour histogram, in: IEEE Conference on Computer Vision and Pattern Recognition, 1992, pp. 599 - 605.

[3] E. A. Bashkov, N. S. Kostyukova, To the Estimation of Image Retrieval Effectiveness Using 2D-colour Histograms, Journal of Automation and Information Sciences 38(11) (2006) 84-89.

[4] A. Vedaldi, B. Fulkerson, An open and portable library of computer vision algorithms, in: Proceedings of the ACM international conference on Multimedia, 2010, pp. 1469-1472.

[5] X. Y. Wang, J. F. Wu, H. Y. Yang, Robust image retrieval based on colour histogram of local feature regions, Multimedia Tools and Applications 49(2) (2010) 323-345.

[6] J. Yue, Z. Li, L. Liu, Content-based image retrieval using colour and texture fused features, Mathematical and Computer Modelling 54(3) (2011) 1121-1127.

[7] G. H. Liu, J. Y. Yang, Content-based image retrieval using colour difference histogram, Pattern Recognition 46(1) (2013) 188-198.

[8] R. Achanta, F. Estrada, P. Wils, Computer Vision Systems: Salient region detection and segmentation, Springer, 2008, pp. 66-75.

[9] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597-1604.

[10] M. M. Cheng, G. X. Zhang, N. J. Mitra, Global contrast based salient region detection, in: IEEE
Conference on Computer Vision and Pattern Recognition, 37(3), 2011, pp. 569 - 582.

[11] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 17(8) (1995) 790-799.

[12] D. Comanicui, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5) (2002) 603 - 619.

[13] R. T. Collins, Mean-shift blob tracking through scale space, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 234-240.

[14] B. Georgescu, I. Shimshoni, P. Meer, Mean shift based clustering in high dimensions: A texture classification example, in: IEEE Conference on Computer Vision, 2003, pp. 456-463.

[15] A. Miguel, C. Perpinan, Gaussian Mean-Shift Is an EM Algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 29(5) (2007) 767-776. 23

[16] A. H. Greenspan, An adaptive mean-shift framework for MRI brain segmentation, IEEE Transactions on Medical Imaging 28(8) (2009) 1238-1250.

[17] L. Zheng, J. Zhang, Q. Wang, Mean-shift-based colour segmentation of images containing green vegetation, Computers and Electronics in Agriculture 65(1) (2009) 93-98.

[18] M. Gong, Y. Liang, J. Shi, Fuzzy c-means clustering with local information and kernel metric for image segmentation, IEEE Transactions on Image Processing 22(2) (2013) 573-584.

[19] F. Mokhtarian, A. K. Mackworth, A theory of multiscale, curvature-based shape representation for planar curves, IEEE Transactions on Pattern Analysis and Machine Intelligence 14(8) (1992) 789-805.

[20] A. Kadir, L. E. Nugroho, A. Susanto, Leaf classification using shape, colour, and texture features, arXiv preprint arXiv (2013) 1401-4447.

[21] D. Zhang, G. Lu, Shape-based image retrieval using generic Fourier descriptor, Signal Processing: Image Communication 17(10) (2002) 825-848.

[22] D. Zhang, G. Lu, A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval, Journal of Visual Communication and Image Representation 14(1) (2003) 39-57.