Pattern Recognition 90 (2019) 390–403

Contents lists available at ScienceDirect

Pattern Recognition

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

Improved adaptive image retrieval with the use of shadowed sets

Hongyun Zhang [a,c,1,*], Ting Zhang [a,c,1], Witold Pedrycz [a,b,d], Cairong Zhao [a,c,*], Duoqian Miao [a,c]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6G 2G7, Canada
[c] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, PR China
[d] System Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 8 May 2018
Revised 6 January 2019
Accepted 25 January 2019
Available online 28 January 2019

Keywords:
Shadowed sets
Saliency detection
Image retrieval
Image segmentation
Edge detection

ABSTRACT

Image retrieval algorithms based on the whole image exhibit high complexity due to background interference, low-level description abilities and large storage requirements, while image retrieval algorithms based on the saliency detection have been found to have low accuracy owing to the lack of important information in extracted salient regions caused by the uncertainty of the salient regions of the image. In this paper, we propose a shadowed-set-based image retrieval algorithm, and develop techniques of an automatic selection of two threshold parameters by combining saliency detection and edge detection, which automatically determine shadowed regions. The developed algorithm uses shadowed set theory to divide the image into salient regions, non-salient regions and shadowed regions, in order to extract the useful information of the image and ignore irrelevant one. As a consequence, this leads to the salient regions and the shadowed regions to be jointly involved in the retrieval process. The experimental results reported for several datasets show that the proposed algorithm can effectively improve the retrieval accuracy compared with the existing state-of-the-art algorithms.

1. Introduction

Currently, digital images have been widely encountered in everyday life, such as product images available on shopping platforms, life photos on social platforms, various digital pictures on major search platforms, and others. With the rapid growth of image databases, finding the target image from the massive image databases becomes an important research topic. Image retrieval is one of the active research pursuits. Its objective is to find the most similar images to the query image from the massive images.

The earliest image retrieval technique was Text Based Image Retrieval.(TBIR [1]), which allows users to type in some keywords and then keyword-related images are retrieved. Its biggest drawback is the need for a large number of manual annotations, because artificially describing an image with a few words only is inefficient, and is not sufficient to fully cover all the features of the image. Therefore, TBIR comes with evident limitations.

Next, Content Based Image Retrieval (CBIR [2]) algorithms were proposed. CBIR usually extracts low-level features such as color, texture and shape of the image to represent the image, and then selects the appropriate measure to calculate similarity between images. Common image features are color histogram [3], bag of word (BOW [4]), spatial envelop model (GIST [5]), scale invariant feature SIFT [6], to identify several alternatives. Content-based image retrieval technology can be divided into whole-image-based image retrieval algorithms [7–13] and salient-regions-based image retrieval algorithms [14–16]. The main problems of image retrieval based on the whole image are the noise of the extracted features, the limitation of the low level features, and the large storage space required. Actually, people tend to pay more attention to the objects with obvious distinguishing ability in the image. Therefore, in the field of image retrieval, in order to improve the efficiency and reduce the time complexity of retrieval, a series of saliency detection algorithms [17–19] were introduced. The salient regions of an image usually refer to the most appealing visual regions of the image. Due to the semantic information being taken into account, extracting and using the salient regions of an image works better than the whole-image-based image retrieval. However, image retrieval algorithms based on saliency detection have the problem that salient regions tend to easily lose important information due to the uncertainty of salient regions, so the binary mask based on saliency

* Corresponding authors at: Department of Computer Science and Technology, Tongji University, no 4800 Cao'an Road, Shanghai 201804, PR China.

E-mail addresses: zhanghongyun@tongji.edu.cn (H. Zhang), zhaocairong@tongji.edu.cn (C. Zhao), dqmiao@tongji.edu.cn (D. Miao).

[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.patcog.2019.01.029
0031-3203/

map segmentation can-not accurately reflect the salient information of the image. For example, the salient regions are classified as background regions, which will result in inaccurate retrieval results.

Methods based on deep learning networks are developed with the success of CNN framework. Xia et al. [20] first combined hashing with CNN. Afterwards, Lin et al. [21] developed CNN based hashing. Also, VGGNet [22] is applied to hash-based image representation. Saliency detection algorithms are also expanded to deep learning frameworks [23–25], which made saliency detection techniques more accurate. However, training a deep network is very time consuming and need big datasets. Without big data as an input, these methods can easily fall into over fitting and hardly achieve good results. In daily life image classification, there still exist some application scenarios whose dataset is not very big.

To address these problems, a new image retrieval algorithm based on shadowed sets is proposed in this study. The shadowed set theory is mainly used for data description and data selection. Essential (core) data and boundary data can be automatically obtained with the use of shadowed sets. The key problem in shadowed set theory is the determination of the two parameters of a shadow set (so-called upper approximation α and lower approximation β). Most recent research is concerned with some empirical determination of these values of the parameters. The algorithm proposed in this paper can automatically determine the values of these two parameters, and apply shadowed set theory to salient-region-based image retrieval. First, the upper approximation is calculated from the saliency map obtained by the saliency detection algorithm. We use the threshold of adaptive segmentation as upper approximation. Then the lower approximation is obtained by edge detection of the original image. It is calculated with the mean value of the gray value of edge regions. After determining the values of these two parameters, we can split the image into the three regions, namely salient regions, non-salient regions and shadowed regions. We consider the regions where the gray value of the pixel is greater than upper approximation as salient regions, and consider the regions where the gray value of the pixel is smaller than lower approximation as non-salient regions. The shadowed regions are located where the gray value of the pixels are between the upper approximation and lower approximation. Finally, salient regions and shadowed regions are jointly involved in the retrieval process. We adopt the iterative-quantization-based hash algorithm (ITQ [9]) as the retrieval algorithm.

Our method exhibits the following important characteristics:

1. The underlying originality of the proposed method relies on the application of the shadowed sets theory in image retrieval, which is a first approach involving shadowed sets in this category of problems.
2. Techniques of automatic parameter selection are developed for automatically obtaining shadowed regions by combining saliency detection and edge detection.
3. A three-way division approach of images is presented to automatically split an image into three regions: salient regions, non-salient regions, and shadowed regions.
4. The salient regions and the shadowed regions are jointly used as the correctly detected regions to improve performance of segmentation and accuracy of retrieval.

This paper is organized as follows: in Section 2, we briefly introduce the related works, including image retrieval algorithms based on the whole image and those based on the saliency detection. In Section 3, we elaborate on the image retrieval model combined with the shadowed set theory. The fourth section focuses on the experimental results and discussions. Finally in Section 5, conclusions are covered.

2. Related studies

In this section, we briefly introduce the relate work of image retrieval algorithms based on the whole image and those based on saliency detection,which are highly related to our work in this paper.

2.1. Image retrieval algorithms based on the whole image

Image retrieval algorithms based on the whole image have made significant progress in recent ten years. The bag of word model (BOW [4]) is a traditional model of text processing. It was applied to image retrieval in 1997 to “visualize” the features of the image and to contribute to large-scale image retrieval. Although the “bag of word” model has achieved good results in the applications, it still exhibits some problems. The biggest problem is the quantization loss of “characterization” of visual features. In order to address this problem, multiple allocation and soft allocation methods have been applied to the quantization of visual words. Some studies also combine local features with color features [11] or with spatial distribution as a weight [10], so as to reduce the error caused by local features. In addition, hash-based image retrieval [7–9,26] is a fast image retrieval algorithm proposed in recent years. It converts the image features into binary hash features, and only needs to calculate the Hamming distance between features during retrieval. The process is very fast. The main problem with such algorithms is the lack of attention paid to spatial information on features. In recent years, the visibility of deep learning has led to intensive research focused on deep learning, which is aimed to extract image features; notably very good results were reported. Xia et al. [20] first combined hashing with CNN. Afterwards, Lin et al. [21] developed CNN based hashing. Also, VGGNet [22] is applied to hash-based image representation.

Image retrieval algorithms based on the whole image take the feature of the whole image into account. These algorithms exhibit some limitations. First, they usually include some background noise and are affected by existing interferences. Second, the features of the whole image exhibit low-level description abilities. Third, these algorithms require a large storage space. Last but not least, the algorithms coming from this category are of relatively high complexity. In real world problems, people tend to pay more attention to the obvious objects in the scene. Based on this, some researchers proposed saliency-detection-based image retrieval algorithms, which can comprehensively capture and consider the semantic information of images and in this manner reduce the size of features and the complexity of the algorithm.

2.2. Image retrieval algorithms based on saliency detection

Compared with image retrieval algorithms based on the whole image, the salient-regions-based image retrieval algorithms not only can avoid the noise and interference caused by the background feature, but also can consider the semantic information of the image. The task of saliency detection [17–19] is to extract the set of pixels that have obvious differences from the surrounding pixels in the image, such as color, edge, texture, etc., as the salient regions of the image. Salient-regions-based image retrieval algorithms fall into two main categories: 1) Query pruning or early termination by using saliency map (saliency pruning, SP); 2) Using saliency map to weight the visual features or image patches (saliency weighting, SW).

The retrieval algorithms using only the features of salient regions (saliency pruning) were almost developed regardless of background information. Reference [15] uses saliency map to filter the feature points, in order to improve the speed of large-scale image retrieval. Besides, it sorts the rest of the feature points according to

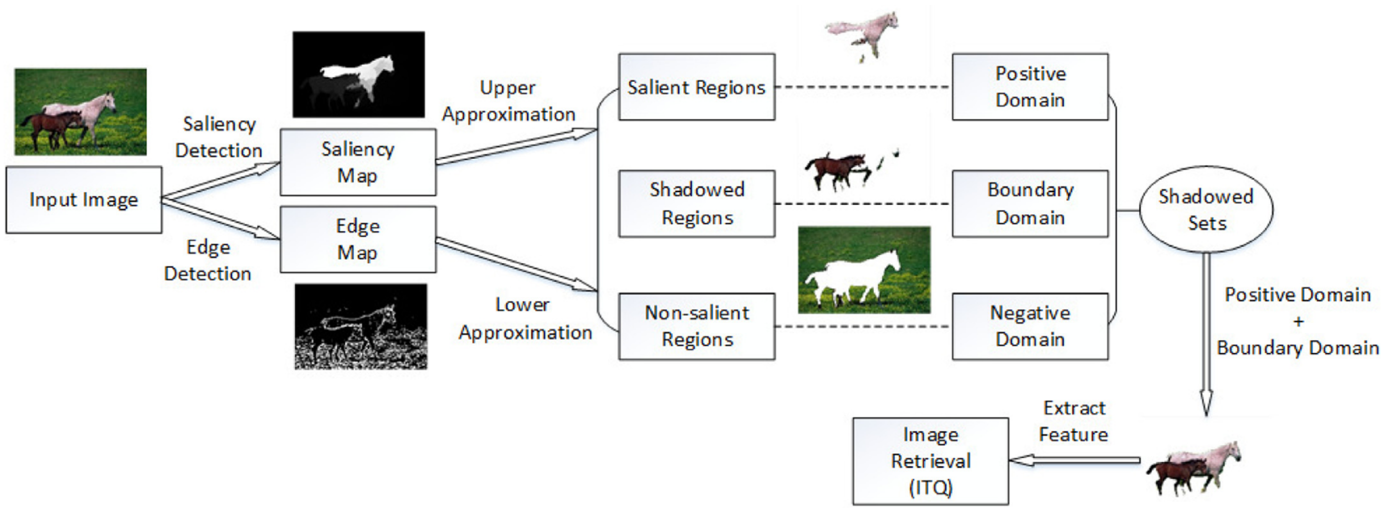


Fig. 1. Main idea of proposed algorithm: image retrieval based on shadowed sets.

the saliency map, which not only can improve the retrieval speed, but can also ensure the accuracy. Being different from saliency pruning, saliency weighting means that different regions can be assigned different weights by saliency map. In [14], the local features and global features are weighted by saliency.

Many improvements have been made based on the salient-regions-based image retrieval algorithm from different aspects, such as integrating spatial information, modifying segmentation methods or changing saliency detection algorithms. However, the importance of the easily lost fuzzy regions between the salient regions and non-salient regions is rarely considered in salient-regions-based image retrieval. In fact, due to the uncertainty of the salient regions of images, there exist fuzzy regions which cannot be simply defined as salient or non-salient regions. To solve this problem, in this study we propose an adaptive image retrieval algorithm based on shadowed sets theory.

3. Image retrieval based on shadowed sets

Traditional image detection algorithms based on saliency detection divide the image into salient regions and non-salient regions, which is a two-way (binary) division. The main problem of this type of algorithm is that it is unable to handle the uncertainty of the image. Hence, we introduce shadowed set theory to realize a new split of the image, which splits an image into three regions: salient regions, non-salient regions, and shadowed regions. First, graph-based manifold ranking [19] is used to generate the saliency map. Given this saliency map, adaptive threshold leads to the upper approximation to extract the salient regions. Then the edge of the image is detected by the Sobel edge operator. We obtain the lower approximation from the edge information, so that the shadowed regions are extracted automatically. Finally the salient regions and the shadowed regions are jointly involved in the retrieval. ITQ [9] is used as the retrieval algorithm. Fig. 1 shows the essence of the proposed approach.

3.1. Shadowed sets

The concept of shadowed sets was proposed by Pedrycz [27]. The shadowed set theory preserves the core fuzzy information of the object through a three-valued logic mapping, which essentially delivers a concise representation of the concept of fuzzy sets. Assuming that X is a fuzzy set, the shadowed set maps this fuzzy set to a three-valued space described in the following form {1, 0, [0, 1]}. 0 means the element does not belong to X (exclusion). 1 indicates that the sample belongs to X (inclusion). The interval [0, 1] indicates that the sample may or may not belong to X, which constitutes a shadow of the construct.

Definition 1. Assume that the membership function of fuzzy sets is $f(x)$. Let $f(x) = 1$, if $f(x) > \alpha$; let $f(x) = 0$, if $f(x) < \beta$. If $\beta < f(x) < \alpha$, we let $f(x) \in [0, 1]$. This is equivalent to mapping x to 0, 1 and the unit interval [0,1], ie f: $X \rightarrow \{0, 1, [0, 1]\}$. The mapping is referred to as a shadowed set.

Shadowed sets theory is usually used to deal with the problem of uncertainty. Since 1998, shadowed sets has emerged as a new way to model ways of representing and processing fuzzy sets, the theory has been used by many scholars in different fields. Cattaneo proposed an algebraic method to define the relationship between fuzzy sets and shadowed sets [28]. Pedrycz applied shadowed sets to fuzzy clustering in many papers in order to improve the clustering effect [29–32]. Zhou et al. applied the data selection method of shadowed sets into neural networks to improve its performance [33]. In 2017, there have been progress made in the shadowed set theory. Cai et al. [34] interpreted dynamic fuzzy sets by means of shadowed sets. Claudia [35] presented an approach to obtain the shadowed set for Triangular and Gaussian membership function. Yao proposed a framework for constructing shadowed sets and three-way approximations of fuzzy sets in [36].

3.2. Image segmentation combined with shadowed sets

Inspired by the theory of shadowed sets, if we can extract the shadowed regions of images, we may improve the retrieval accuracy. The crucial point is using the image information determine automatically these three regions. At present, the values of the two parameters α and β are empirically determined. Therefore, to be able to adapt to the image content, we develop techniques of an automatic selection of two parameters by combining saliency detection and edge detection, which automatically determine shadowed regions. In this way the image is split into salient regions, non-salient regions and shadowed regions (fuzzy regions).

3.2.1. Determination of parameter α based on the saliency detection and adaptive segmentation

Based on the saliency detection of early works, we use graph-based manifold ranking [19] algorithm to obtain the saliency map and adaptively obtain the parameter α on it. In order to avoid the accuracy being influenced by illumination variant and affine

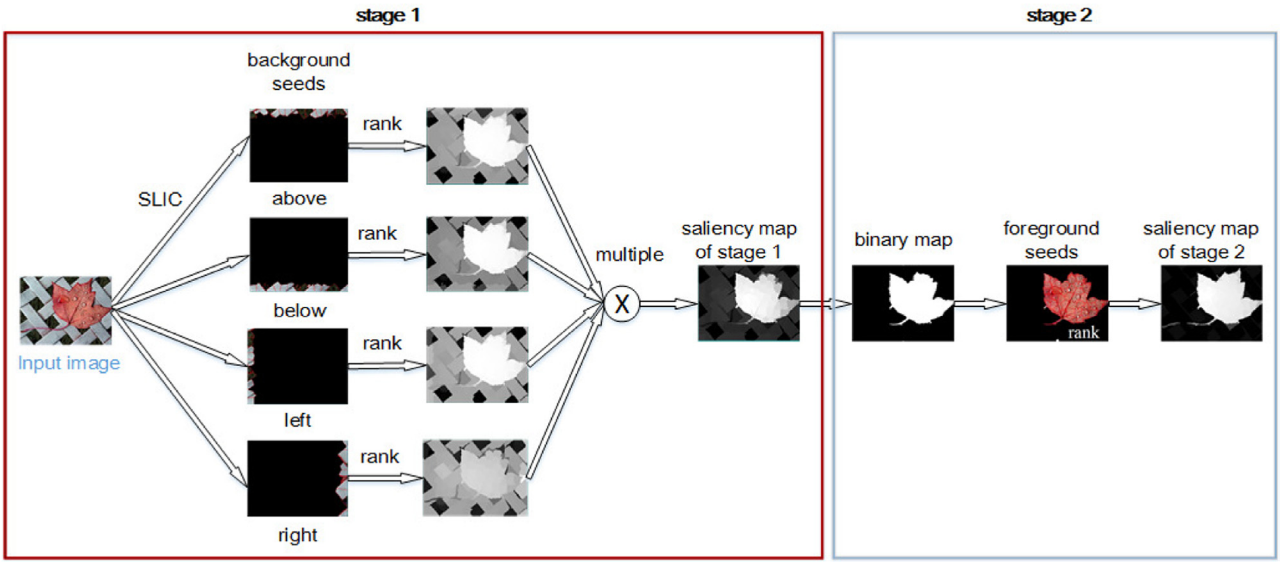


Fig. 2. Saliency detection via graph-based manifold ranking [19].

transformation, we extract SIFT (scale invariant feature) [6] feature before saliency detection. SIFT feature is represented by vectors in different directions, which is invariant to image scale and illumination. Specifically, graph-based manifold ranking is a bottom-up saliency detection algorithm. First, the graph is constructed by applying SLIC [37] algorithm to the SIFT feature extracted from original images. Second, the sparse color histogram features of each super pixel patches is extracted, and the super pixel patches at the four edges of the image are considered as the background seed nodes. Third, manifold ranking algorithm [38] is used to calculate the four initial saliency maps of the seed nodes. Finally, the four maps are multiplied and merged into a single saliency map and the overall flow of saliency detection is visualized in Fig. 2.

Saliency can be used as a weight or a selector. A method of assigning saliency is to use a threshold to binarize the saliency map, and generate a mask map to mask the original image. The threshold of the saliency map can be fixed or adaptive. The fixed threshold is determined in an empirical fashion, while the adaptive threshold is expressed as twice the average gray value of the saliency map. According to the obtained saliency map (the rightmost image in Fig. 4), we can find that there is an obvious difference in gray scale between salient regions and non-salient regions. So we adaptively set the upper approximation through the adaptive threshold.

$$\alpha = \frac{2\sum_{x,y} S(x,y)}{m \times n} \tag{1}$$

Where m and n denote the height and width of the image, respectively, $S(x, y)$ represents the saliency value of the saliency map at the position (x, y), x, y, m and n satisfy $1 \le x \le m$ and $1 \le y \le n$.

Therefore, the salient region $S(x, y)$ should satisfy the following inequality:

$$\frac{2\sum_{x,y} S(x,y)}{m \times n} \le S(x,y) \le 255 \tag{2}$$

Once the threshold value has been set, the regions of the saliency map that are lower than the threshold is set as 0, and the regions that larger than the threshold is set to 1. This binarized map is a mask. We use it to mask the original image several times, in order to obtain the salient regions of the image; refer to Fig. 3.

3.2.2. Development of the parameter β based on edge detection

From Fig. 3, we note that the salient regions of the images can express the most obvious parts of the images, but still lose some useful regions of the images that are similar to the background or are different from the saliency regions. Those are the dark pony in the first sample image, some white portion of the scarf in the second sample image, the bottom of the color bucket in the third sample image, the medal of the circle in the fourth sample image, and the golden pattern of the carpet in the fifth sample image. As these regions tend to include obvious boundary between salient regions or background regions, large gradient changes in pixel values can occur on the digital representation of the images. Therefore, in order to “fill in” these missing “fuzzy” regions, we compute the lower approximation of the saliency map in conjunction with the edge detection algorithm to obtain the boundary regions of the images.

We propose to employ Sobel-operator-based edge detection because of its stability and robustness. The Sobel operator is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. As a small, separable, and integer-valued filter, it is used to convolve the image in the horizontal and vertical directions. If we define I as the source image, G_X and G_Y are two images which at each point contain the horizontal and vertical derivative approximations, respectively. The final gradient G can be calculated as the sum of the squares of G_X and G_Y. The computations are completed as follows:

$$G_X = Sobel_X * I,\ G_Y = Sobel_Y * I,\ G = \sqrt{G_X^2 + G_Y^2} \tag{3}$$

Where $Sobel_X = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$, $Sobel_Y = \begin{bmatrix} -1 & -2 & +1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$.

Fig. 4 shows the Sobel edge detection result obtained for a sample image, in which the edge regions are outlined with white outline. Recall that the gray value at the edge regions is denoted as $edge(x, y)$.

Since the missing useful regions tend to include obvious boundary between salient regions and background regions, we take the adaptive mean value of the edge regions as the lower approximation β.

$$\beta = \frac{\lambda \sum_{x,y} edge(x,y)}{N} \tag{4}$$

Fig. 3. The first line shows sample pictures and the second line shows the salient regions we extract.

Fig. 4. sample image and its edge detection.

Where N represents the number of pixels in the edge regions, λ denotes the adaptive coefficient, and the shadowed regions are extracted by the following inequality:

$$\frac{\lambda \sum_{x,y} edge(x,y)}{N} \leq S(x,y) < \frac{2\sum_{x,y} S(x,y)}{m \times n} \tag{5}$$

In addition, the extraction of the non-salient regions should satisfy the relationship:

$$0 \leq S(x,y) < \frac{\lambda \sum_{x,y} edge(x,y)}{N} \tag{6}$$

In this way, we can automatically obtain the upper and lower approximations based on the shadowed sets. Hence, the image is divided into the three regions: salient regions, non-salient regions, and shadowed regions. As shown in Fig. 5, the first line displays the original sample images, the second line shows the salient regions, while the third line shows the shadowed regions, and the fourth line shows the non-salient regions. In the figure, the black regions represent the filtered regions and the colored regions are the reserved areas.

In the proposed algorithm, the parameter λ is crucial for extracting shadowed regions and non-salient regions. Different values of the parameter λ determine whether certain shadowed regions are reasonable or not. However, it is difficult to select proper value for this parameter because it would be different for data sets with different images. Since β is smaller than α and λ is between 0 and 1, we inspect several possible values to find optimal result. Empirically, it was found that when we set $\lambda = 2/7$, the parameters form a sound option. In the reported experiment in Section 4.5, the value of the parameter λ is further verified.

The retrieval algorithm extracts visual feature on salient regions and shadowed regions, as shown in Fig. 6, which preserves more valid information than saliency-detection-based image retrieval algorithms that extract features only on salient regions.

3.3. Shadowed-sets-based image retrieval

After extracting the features of the salient regions and shadowed regions of the image, iterative quantization (ITQ [9]) based algorithm is used for image retrieval. ITQ is an unsupervised alternating minimization scheme inspired by multi-class spectral clustering and the orthogonal Procrustes problem. It is simple and efficient. The main idea of ITQ is to find a rotation of zero-centered data so as to minimize the quantization error of mapping this data to the vertices of a zero-centered binary hypercube. Hence, the similarity between different images can be expressed by XOR operation, which is very suitable for large data sets. So we employ ITQ in this version of work. We first extract the GIST [5] features based on the Gabor filtering as the input of ITQ. It is extracted by Gabor filters at 4 scales and 8 directions to convolute with the image, so as to generate 32 features maps of the input image. Each feature map is divided into 16 regions. Next, the mean value of each region is computed. Finally, the 16 mean values of 32 feature maps are aggregated to 16 $\times$ 32 GIST feature. Suppose the feature of each image extracted by a dataset X with n images is $\{x_1, x_2, \ldots, x_n\}$, $x_i \in R^d$, that form the rows of the data matrix $X = [x_1, x_2, \ldots, x_n]^T \in R^{n\times d}$, $d = 512$. We assume that the points are zero-centered, i.e., $\sum_{i=1}^{n} x_i = 0$. Our goal is to learn a binary code matrix $B \in \{-1, 1\}^{n\times c}$, where c presents the code length. The detailed algorithm can be found in [9].

Overall, the shadowed-sets-based image retrieval algorithm is divided into two steps. The first step is an offline indexing of all the images present in the dataset. Then we form the binary representation of the images. The second step is the online indexing of query image, and then calculate the distance between all the images in the dataset and the query image, so as to identify the most similar images.

In the offline indexing step, we firstly calculate the saliency map $S \in R^{a\times b}$ according to the saliency detection algorithm with the aid of the graph-based manifold ranking [19]. Then we form the two parameter α and β according to Sections 3.2.1 and Section 3.2.1, respectively. With the use of the two parameters, we extract the salient and shadowed regions $(I \times M)$, where $\times$ means a multiplication of the value of the matrix I with the corresponding position in the matrix M while M denotes the mask map which satisfies the following relationship: $M(x, y) = 1$ if $\beta \leq S(x, y) \leq 255$, and $M(x, y) = 0$ if $0 \leq S(x, y) < \beta$.

Finally, we binarize the GIST feature as $B \in R^{1\times c}$ according to ITQ [9], where c represents the number of bits.

In the online retrieval step, the first step is indexing the query image according to Algorithm 1, and getting $x_{feature} \in R^{1c}$ as the index of the query image. Then, the similarity between the query image and images in dataset is calculated by the xor operation.

Algorithm 1 offline indexing.

Input: Original image $I \in R^{a\times b}$
Output: The binary representation $B \in R^{1\times c}$
1: Calculate the saliency map $S \in R^{a\times b}$
2: Form the upper approximation α according to 3.2.1
3: Form the lower approximation β according to 3.2.2
4: Extract the salient and shadowed regions $(I \times M)$ implied by α and β
5: Extract the GIST feature on salient and shadowed regions $(I \times M)$
6: Binarize the GIST feature as $B \in R^{1\times c}$

Fig. 5. A set of sample images and their division results.

Fig. 6. Salient Regions + Shadowed regions.

Finally, the retrieval result is obtained by sorting the entries of the similarity matrix.

Algorithm 2 online retrieval.

Input: The query image $x \in R^{a\times b}$ and the output of Algorithm 1 $B^n \in R^{n\times c}$
Output: similarity matrix $Similarity \in R^{n\times 1}$
1: Binarize the query image into $x_{feature} \in R^{1c}$ by Algorithm 1
2: **for all** $B^i \in B^n$ **do**
3: $\quad Similarity(i) = x_{feature} \quad xor \quad B^i$
4: **end for**
5: Sort Similarity

4. Experimental studies

4.1. Datasets and code

MSRA10K Dataset: MSRA10K dataset which contains 10,000 images with the ground truth of salient region marked by bounding boxes.

Corel-1000 Dataset: This dataset contains 10,000 images belonging to 100 categories, which include different themes such as portraits, landscapes, buildings, and butterflies. In the experiment, we use the first 9000 images as a training set and the remaining 1000 images as a test set.

SIVAL Dataset: The dataset is designed for partial image retrieval. It contains only the salient features of the image. It consists of 1500 images in 25 categories, with 60 images in each category.

CIFAR-10 Dataset: This dataset contains 60,000 images. Each image is represented by 512-dimensional GIST feature vector extracted from the original color image of size 32 × 32. Each image is manually labeled to belonging to one of the ten classes. Two images are considered to be semantically similar if they share the same class label. Otherwise, they are treated as semantically dissimilar.

The core code of the proposed method can be obtained on the github.[2]

4.2. The comparison of performance of the segmentation

We compared the proposed algorithm with several state-of-the-art saliency detection algorithms on the MSRA10K database. In our method, the correctly detected salient regions include salient regions and shadowed regions. Performance was evaluated by assessing precision, recall, and *F*-measure. Precision is the ratio of correctly detected salient pixels to the total detected salient pixels, while recall is the ratio of correctly detected salient pixels to the ground truth salient pixels. In addition, we introduced the *F*-measure to evaluate the overall performance. The *F*-measure value is defined as:

$$F_\beta = \frac{(1+\beta^2)Precision * Recall}{\beta^2 Precision + Recall} \tag{7}$$

Where $\beta^2 = 0.3$, as presented in [19].

According to the number of references, the proposed time and the diversity criteria, we select four algorithms for the comparative analysis, which are CB [39], RC [40], SEG [41] and FT [42]. Since the saliency map obtained by the above algorithms gives the salient value of each pixel but not binary, we adaptively divide the saliency map into the corresponding binary mask map in the experiment. That is, the value in the set {0, 1}, where the value of 1 means salient, the value of 0 means non-salient. The algorithm proposed in this paper takes the salient regions and the shadowed regions as retrieval regions, so the salient regions and the shadowed regions are set to 1, and non-salient regions are set to 0.

Fig. 7 shows the results of comparative study. Some examples of visualizations are shown in Fig. 8. Fig. 8(a) is the input image, Fig. 8(b) is the saliency labels at the pixel level given by the MSRA10K dataset, and Fig. 8(c)–(f) are the division results of CB, RC, SEG and FT. From Fig. 7, we can see that our algorithm is 5% ∼ 15% more accurate than CB, RC, SEG and FT. The Precision, Recall, and *F*-measure of the algorithm CB are 0.87, 0.81 and 0.85, respectively, which are only 1% lower than our algorithms on Recall. However, the Precision and the *F*-measure are 7% and 6% lower than ours respectively. The Precision, Recall and *F*-measure of the algorithm RC are 0.92, 0.79 and 0.89, respectively, which are almost the same as our algorithm in Precision and *F*-measure. However, the algorithm SEG's Precision, Recall, *F*-measure are 0.85, 0.59 and 0.77 respectively, which are 9%, 23% and 14% respectively

[2] https://github.com/zhangtingwaa/mycode.

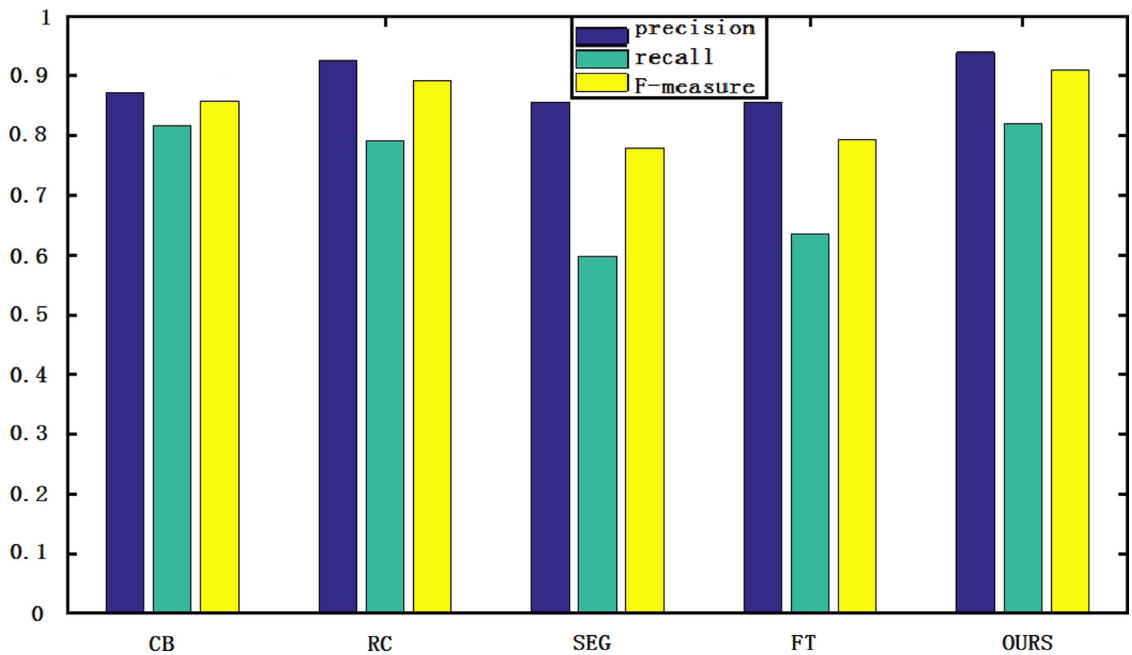


Fig. 7. The histogram of the comparative analysis.

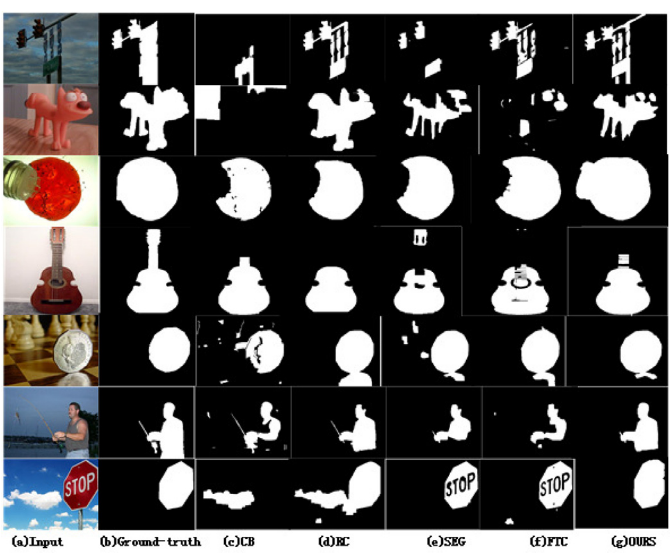


Fig. 8. Some examples of the obtained visualizations.

lower than for our algorithm. Algorithm FT's Precision, Recall, *F*-measure are 0.85, 0.63 and 0.79, respectively, which are 9%, 19% and 12% lower than for our algorithm. Through observations coming from experiments, it is not difficult to find that due to the idea of shadowed sets, our algorithm can more accurately capture salient regions of the image, which not only can be seen visually in Fig. 8, but also is reflected in the high Precision and Recall shown in Fig. 7. In addition, from Fig. 8, we can observe that the proposed algorithm can produce clearer and smoother boundaries of the salient regions.

4.3. Accuracy evaluation of the retrieval

On the Corel-1000 dataset, SIVAL dataset and CIFAR-10 dataset, we use the *Precision-recall* (PR) curve to measure the quality of the retrieval, where the *Precision* = the number of correct images retrieved / the number of all retrieved images, while the *recall* = the number of correct images retrieved / the number of images in the dataset. Besides, we use MAP to measure the accuracy of the retrieval, which is equal to the area under the *Precision-recall* curve.

Our algorithm is based on the ITQ algorithm, so we take the ITQ algorithm as a baseline for comparison. In addition, we select five other state-of-the-art hash-based image retrieval algorithms as comparison algorithms:

RR: This algorithm is similar to ITQ, except that the orthogonal matrix R in the second step is replaced by a random orthogonal matrix.

LSH [7]: Locality-Sensitive Hashing (LSH) was originally proposed in 1999. The basic idea is that if two data points are adjacent in the original data space, the probability of their adjacency in the new data space is very large after a mapping or a projection transformation. On the contrary, if two data points are not adjacent in the original data space, the probability of their adjacency in the new data space is very small after a mapping or a projection transformation.

SKLSH [8]: In 2009, Raginsky et al. proposed a Shift-Invariant Kernelized Locality-Sensitive Hashing (SKLSH) model. SKLSH takes the random mapping as the core idea and proposes a mapping method which is not related to data distribution. The advantage is that the Hamming distance between hash codes corresponds to the translational invariant kernel value of the corresponding eigenvector.

ITQ [9] (baseline): At the CVPR meeting, in 2013, Gong et al. proposed an iterative quantization-based image retrieval algorithm (ITQ). Firstly, it uses principal component analysis to reduce the dimension of the data, and then orthogonally transforms the projection data to minimize the quantization error.

COSDISH [12]: In 2016, considering the errors caused by the relaxation in graph based hashing, column sampling based discrete supervised hashing was proposed. It directly learns the discrete hashing code from semantic information. It is an iterative method and shows a constant-approximation bound in each step of the alternating optimization procedure.

LGHSR [13]: In 2017, Li et al. introduced spectral rotation technique into graph based hashing algorithm. It was proposed to address the problem of the spectral solution in real value in such methods deviating from the discrete solution. The binary codes are

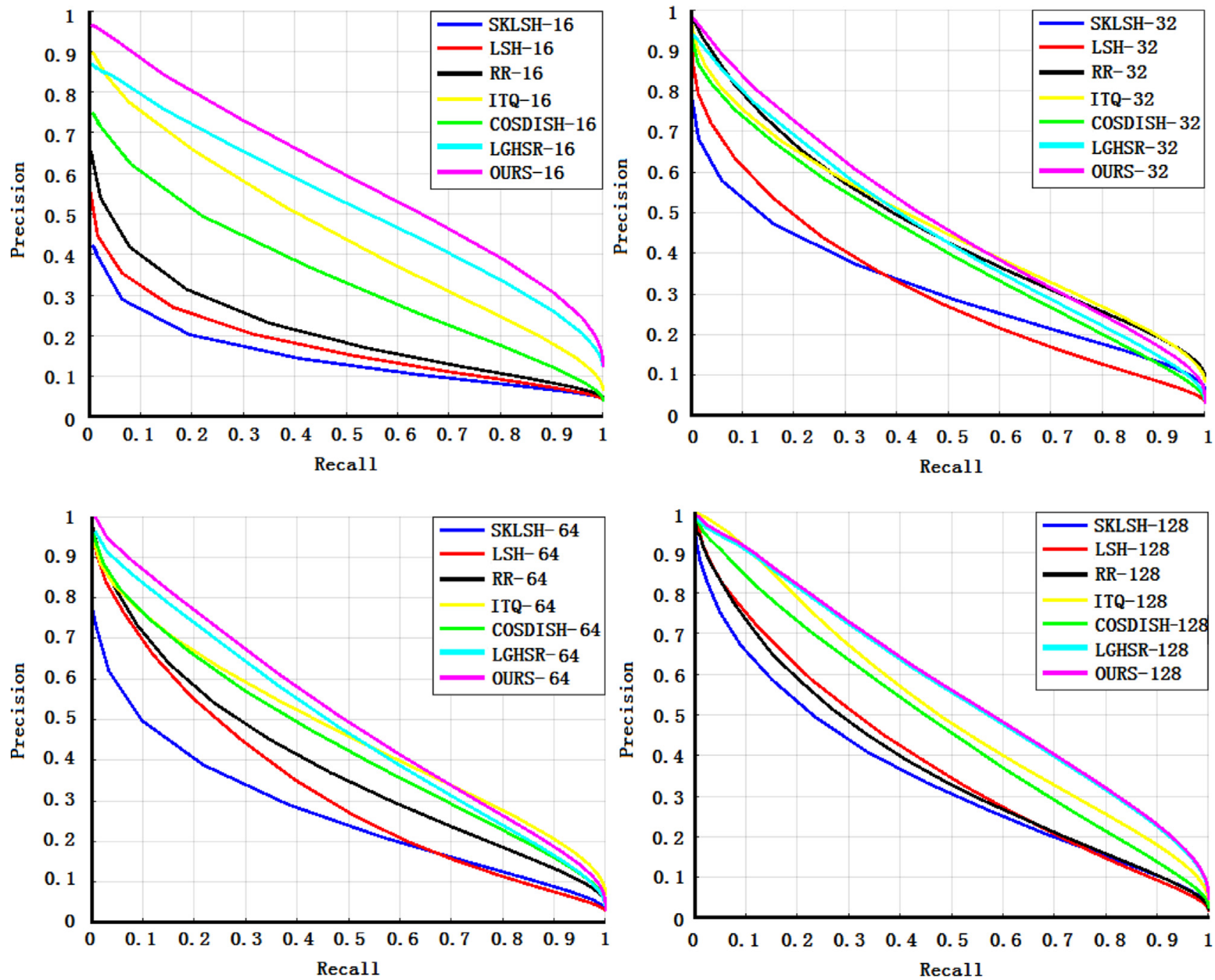


Fig. 9. The comparison results (PR curve) on Corel-1000: (a) 16 bits; (b) 32 bits; (c) 64 bits; (d)128 bits.

obtained from the modified solution via minimizing the Euclidean distance.

Figs. 9–11 show the PR curve of the seven retrieval algorithms for different hash bits on the Corel-1000, SIVAL and CIFAR-10 datasets, respectively.

Fig. 9 clearly shows that our method achieve far better performance than ITQ. The main difference between our method and ITQ is that ITQ is based on the whole image while our method combines shadowed set theory to extract the salient regions and shadowed regions. Hence, the intuitive viewpoint that using the precisely useful regions is reasonable has been successfully verified by our experiments.

The results on the SIVAL dataset (Fig. 10) shows that the accuracies of all the algorithms are slightly improved over the accuracies obtained for the Corel-1000, because the same class of images on this dataset has only the background regions changed while the salient regions are exactly the same. However, the advantages of the algorithm are not as obvious as that for the Corel-1000. We can note that the proposed algorithm has a significant effect on the complex and fuzzy image, but few effects on the images with clear foreground. An advantage in short code length is lower than that in long code length.

Fig. 11 shows the comparison result for the CIFAR-10 dataset expressed in terms of precision and recall. We can see that the proposed method outperforms the other methods with almost all

Table 1
MAP of the 7 image retrieval algorithms on Corel-1000.

Bits/algorithm	SKLSH	LSH	RR	COSDISH	ITQ	LGHSR	OURS
16 bits	0.12	0.25	0.24	0.24	0.26	0.31	0.33
32 bits	0.20	0.25	0.39	0.41	0.42	0.54	0.54
64 bits	0.25	0.35	0.45	0.47	0.47	0.57	0.62
128 bits	0.48	0.46	0.54	0.51	0.51	0.58	0.69

the code lengths. Furthermore, the performance of our method is also comparable, if not superior, to the state-of-the-art methods, such as LGHSR and COSDISH.

Tables 1–3 show the values of the MAP (mean average precision) for the seven retrieval algorithms on the three datasets. They show that as the code length increases, the MAP of all algorithms is improved. As the number of bits increases from 64 bits to 128 bits, the MAP of SKLSH, LSH and RR are slightly improved, while the MAP of ITQ and the our algorithm show a little improvement. This indicates that the influence of the code length gradually decreases as the code length increases.

In addition to comparing the proposed algorithm with the image retrieval algorithms based on the whole image, we also compare it with the saliency-detection-based image retrieval algorithm. In order to show the validity of the shadowed-sets-based division proposed in this paper, GB [19] and ITQ [9], respectively are used as the saliency detection algorithm and the retrieval

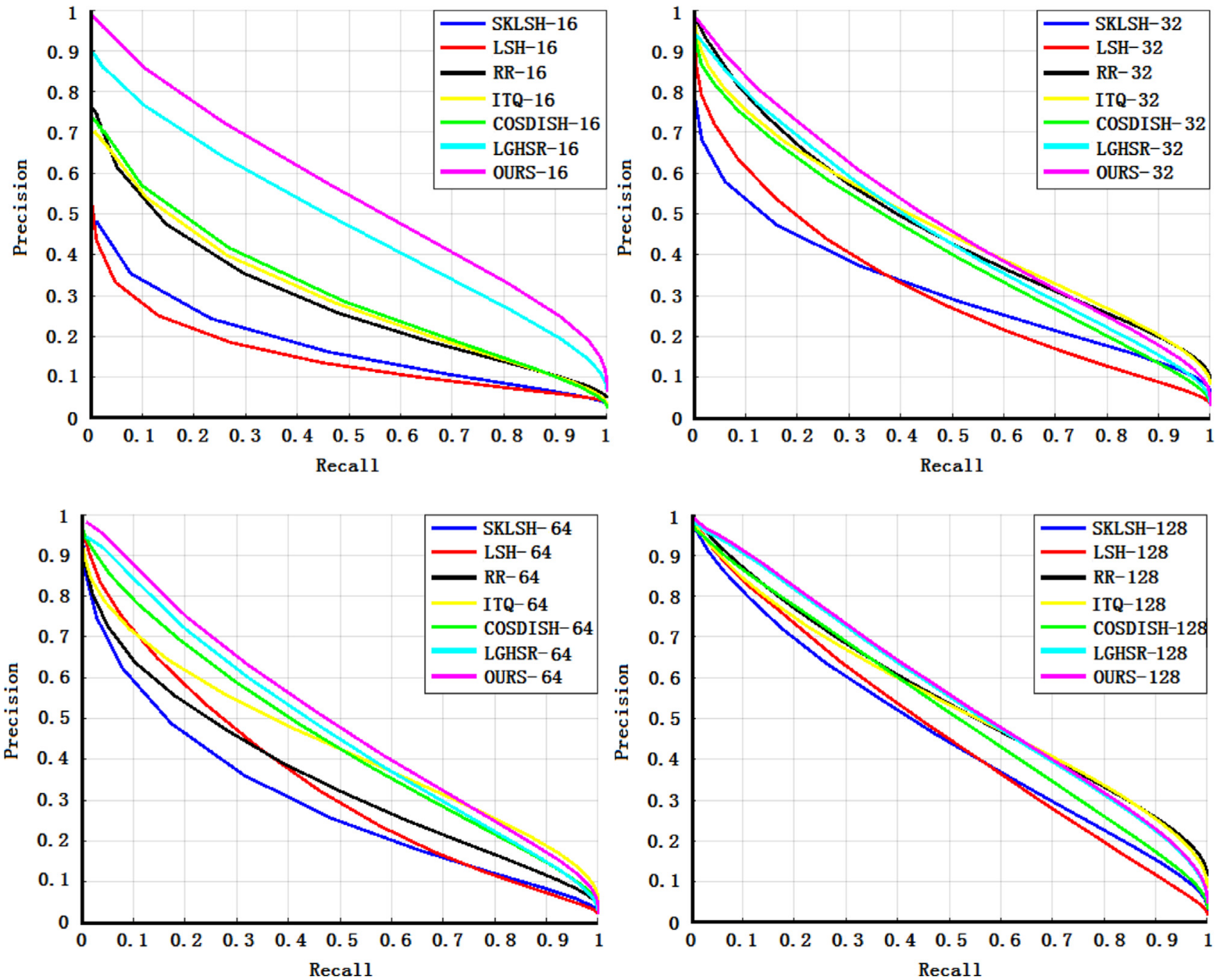


Fig. 10. The comparison results (PR curve) on SIVAL: (a) 16 bits; (b) 32 bits; (c) 64 bits; (d) 128 bits.

Table 2
MAP of the 7 image retrieval algorithms on SIVAL.

Bits/algorithm	SKLSH	LSH	RR	COSDISH	ITQ	LGHSR	OURS
16 bits	0.18	0.29	0.31	0.34	0.35	0.37	0.37
32 bits	0.24	0.31	0.43	0.47	0.49	0.55	0.56
64 bits	0.26	0.41	0.59	0.61	0.60	0.62	0.64
128 bits	0.48	0.49	0.64	0.66	0.66	0.64	0.65

Table 3
MAP of the 7 image retrieval algorithms on CIFAR-10.

Bits/algorithm	SKLSH	LSH	RR	COSDISH	ITQ	LGHSR	OURS
16 bits	0.14	0.27	0.34	0.36	0.35	0.38	0.38
32 bits	0.22	0.29	0.42	0.43	0.47	0.55	0.56
64 bits	0.29	0.31	0.50	0.65	0.58	0.64	0.67
128 bits	0.38	0.48	0.61	0.66	0.64	0.67	0.68

algorithm in the comparative experiment for the sake of fairness, of which the code length in ITQ is 128 bits. In the feature extraction section, we use three features: color histogram, BOW and GIST. Figs. 12, 13 and 14 show the results on the Corel-1000, SIVAL and CIFAR-10 datasets, respectively. We observe from Figs. 12 and 14 that since the difference between the background and the foreground is rather vague, the segmentation combined with the shadowed sets can be more accurate than the one without shadowed sets, and the PR curve is much higher. However, due to the clear difference between background and foreground, there is no obvious advantage in combining shadowed sets theory on the SIVAL dataset. Furthermore, the use of shadowed set theory exhibits a certain degree of significance on the retrieval accuracy.

Tables 4–6 show the concrete value of MAP on the Corel-1000, SIVAL and CIFAR-10 datasets, respectively. It is clear that on the Corel-1000 dataset, the final retrieval accuracy is improved by 6%, 8% and 7%, respectively, whereas on the SIVAL dataset the improvements are only 1%, 2% and 6%. This indicates that the proposed algorithm is more effective for fuzzy images in the salient regions, and the GIST feature exhibits stability in the three visual features.

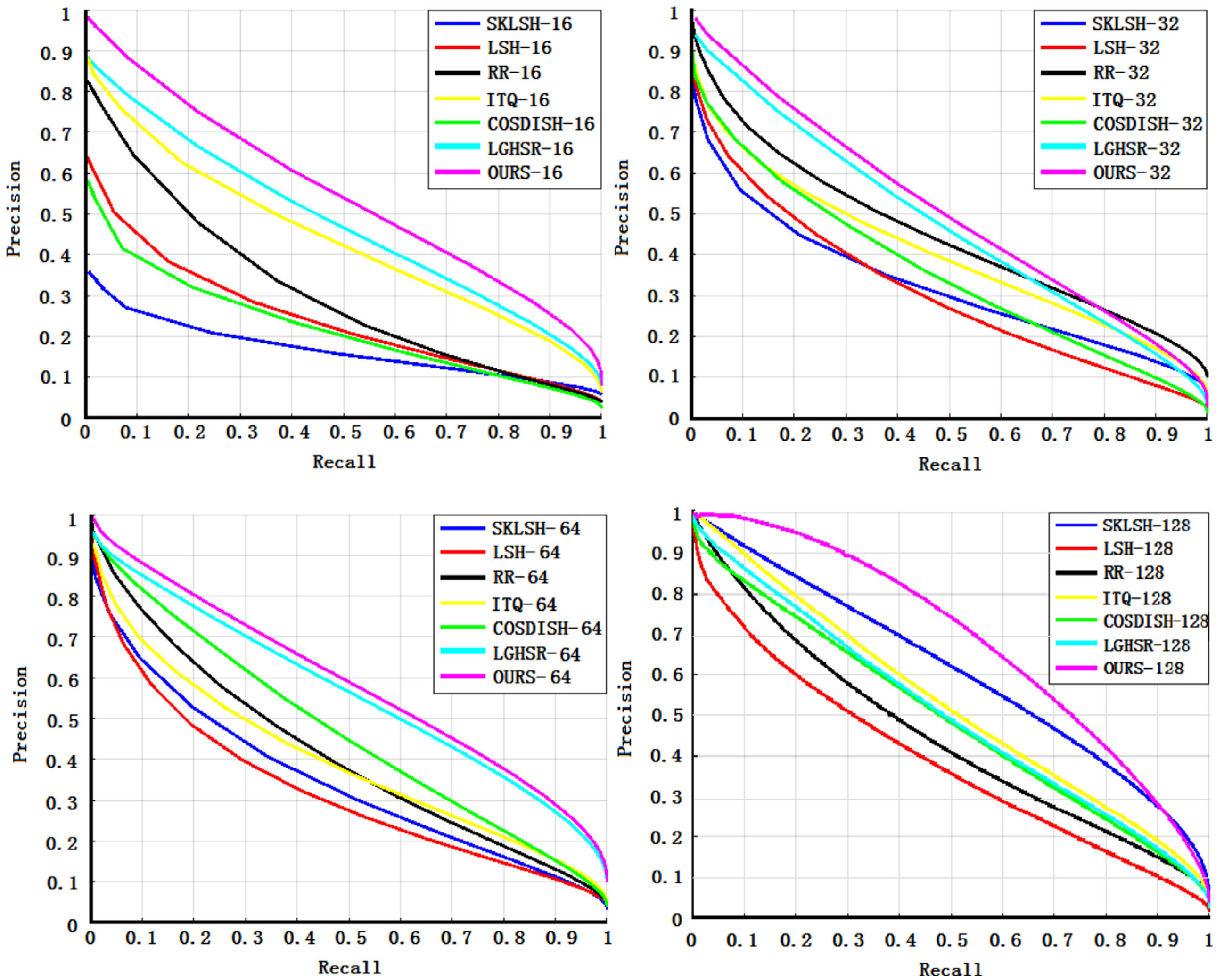


Fig. 11. The comparison results (PR curve) on CIFAR-10: (a) 16 bits; (b) 32 bits; (c) 64 bits; (d) 128 bits.

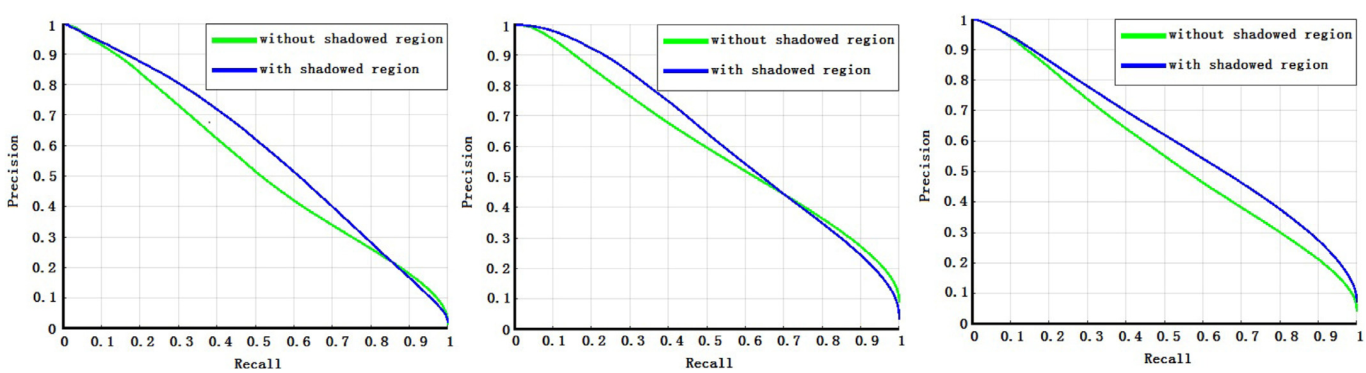


Fig. 12. Verification of the validity (PR curve) of the shadowed sets theory on Corel-1000: (a) Color Histogram; (b) BOW; (c) GIST.

Table 4
MAP on Corel-1000.

Method	Color histogram	BOW	GIST
Without shadowed region	0.53	0.52	0.55
With shadowed region	0.59	0.60	0.68

Table 5
MAP on SIVAL.

Method	Color histogram	BOW	GIST
Without shadowed region	0.56	0.56	0.57
With shadowed region	0.57	0.58	0.69

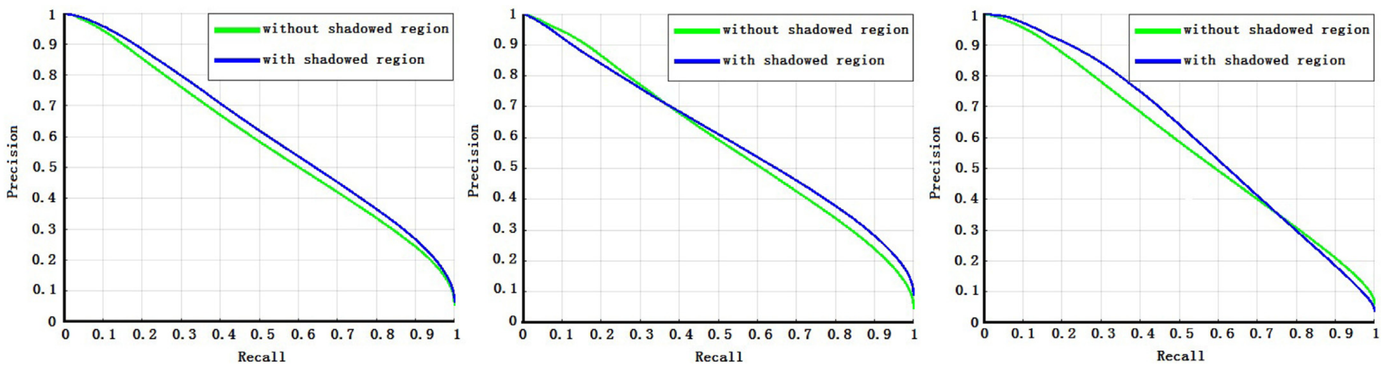


Fig. 13. Verification the validity (PR curve) of the shadowed sets theory on SIVAL: (a) Color Histogram; (b) BOW; (c) GIST.

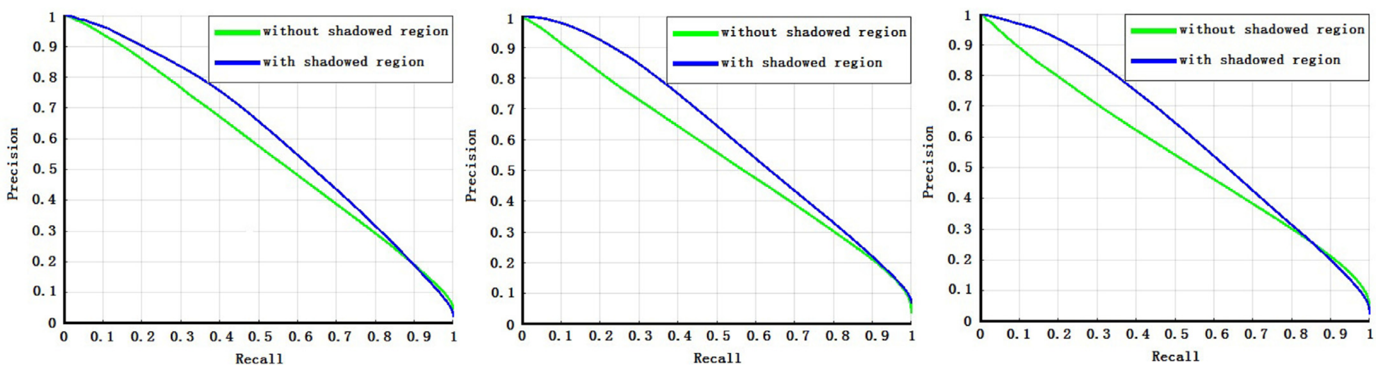


Fig. 14. Verification the validity (PR curve) of the shadowed sets theory on CIFAR-10: (a) Color Histogram; (b) BOW; (c) GIST.

Table 6
MAP on CIFAR-10.

Method	Color histogram	BOW	GIST
Without shadowed region	0.57	0.59	0.62
With shadowed region	0.61	0.64	0.69

4.4. Comparison with deep learning algorithms

This section first compares the accuracy and the run time of the shadowed-set-based segmentation algorithm with those of deep segmentation algorithms. Then, we compare the accuracy and the run time of the proposed image retrieval algorithm with those of deep hashing algorithms. Meanwhile, we analyze the advantages and disadvantages of our algorithm.

4.4.1. The comparison of the accuracy and run time to deep segmentation algorithms

Dataset: PASCAL VOC 2012 [43]. PASCAL VOC 2012 is a well-known segmentation evaluation dataset which consist of 20 object categories and one background category. This original dataset is split into a training set, a validation set and a test set, which contain 1464, 1449 and 1456 images, respectively. This dataset is augmented by the extra annotations provided by Hariharan et al. [44], resulting in 10,582 training images. The performance is measured in terms of pixel intersection-over-union (IOU) averaged across the 21 classes.

The quantitative results of the proposed algorithm and the competitors are presented in Table 7. We compared our method with three deep learning based methods: DeepLab [25], SegNet [23] and FCN [24]. The quantitative performance of DeepLab and SegNet can be obtained from the study which proposed the two algorithms [23,25], while the quantitative performance of FCN is obtained by using the public source code available on github.[3] The performance of our algorithm is competitive to the state-of-the-art deep learning methods. We outperform competing methods in most categories, especially in categories which report relatively low accuracy under other methods (bike, boat, chair, table and sofa).This confirms that, with the consideration of fuzzy regions between foreground and background, our method can deal with more complex image data. Besides, in some categories already reported high accuracy (areo, cow, mbk), our method also achieve a little bit improvement. However, in the categories of bgk, bird, bus, cat, and person, our method can not achieve the best performance among the four methods. It is because the five categories have one thing in common, that is, the foreground and background are relatively clear there. In other words, there are nearly no fuzzy regions existing between the foreground and the background. Hence, the advantage of our method can not be manifested.

Table 8 demonstrates the run time comparison of the methods. The configurations of GPU is MSI GeForce GTX 1080 Ti, and we have two gpus on the computer. Since our algorithm is not based on deep networks, our training time does not contain two parts (forward time and backward time). FCN, SegNet which have fully connected layers (turned into convolutional layers) train much more slowly. Here we note also that over-fitting was a big issue in training these network models on small datasets. Instead, our method is simple and efficient compared to these deep learning based methods. We achieve the best performance of run time on this dataset.

4.4.2. The comparison of the accuracy and run time to deep hashing algorithms

To verify the effectiveness of the 2-threshold parameter-based image retrieval algorithm, the state-of-the-art deep hashing based

[3] https://github.com/shekkizh/FCN.tensorflow.

Table 7
Segmentation accuracy on PASCAL VOC 2012 test set.

Method	bkg	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow
DeepLab	**92.1**	78.4	33.1	**78.2**	55.6	65.3	81.3	75.5	78.6	25.3	69.2
SegNet	89.6	76.5	32.2	73.5	50.1	71.2	**83.4**	76.1	**82.1**	22.7	67.7
FCN	91.2	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5
OURS	90.2	**86.1**	**45.7**	77.1	**67.8**	**75.9**	81.4	**76.3**	80.3	**42.5**	**80.5**
Method	table	dog	horse	mbk	person	plant	sheep	sofa	train	tv	mean
DeepLab	52.7	75.2	69.0	79.1	**77.6**	54.7	78.3	45.1	73.3	56.2	66.4
SegNet	56.4	76.5	69.6	77.7	58.2	52.4	72.9	45.2	69.8	54.3	64.7
FCN	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	62.2
OURS	**64.7**	**80.4**	**75.0**	**80.8**	76.0	**61.5**	**79.2**	**53.8**	**73.9**	**66.0**	**72.1**

Table 8
Run time of segmentation on PASCAL VOC 2012 dataset.

Method	Forward pass(ms)	Backward pass(ms)	Training time(ms)	Test time(ms)
DeepLab	297.06	360.73	657.79	25.52
SegNet	632.50	708.71	1341.21	63.36
FCN	516.09	694.11	1210.20	59.24
OURS	–	–	197.34	5.93

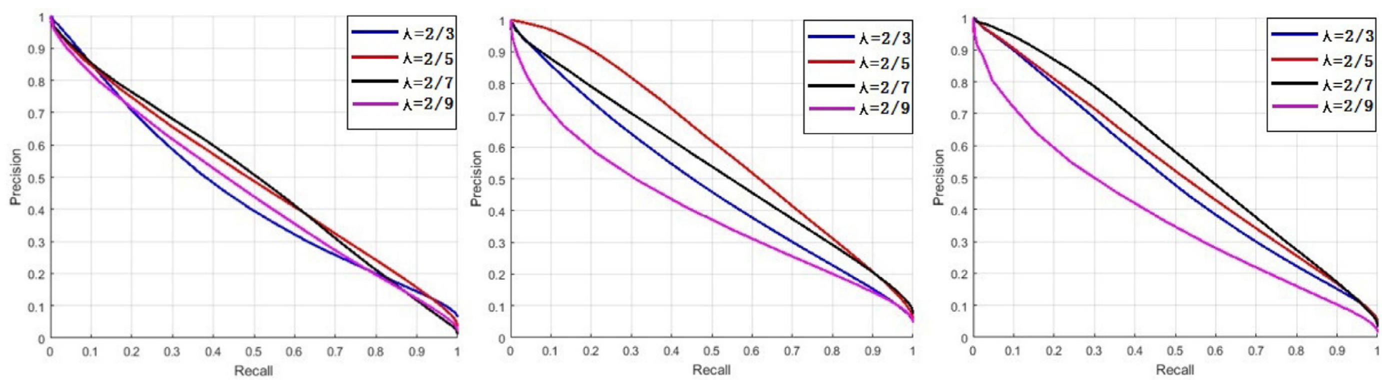


Fig. 15. The influence of the adaptive parameter λ on **Corel-1000, SIVAL** and **CIFAR-10** respectively.

Table 9
Performance comparison (mAP, %) of different bits on CIFAR-10 dataset.

bits/method	CNNH	DLBH	DeepBit	OURS
16 bit	31.9	77.5	17.8	38.4
32 bit	54.7	80.8	23.9	56.2
64 bit	58.4	84.4	24.7	67.9
128 bit	63.2	89.6	33.8	68.4

image retrieval algorithms are also taken into consideration in the comparison experiments. We compared our method with CNNH [20], DLBH [21] (Deep Learning of Binary Hash Codes for Fast Image Retrieval) and DeepBit [22] on the CIFAR-10 dataset.

Table 9 shows the CIFAR-10 retrieval results based on the mean Average Precision (mAP) of the top 1000 returned images with respect to different bit lengths. Our method improves the performance of DeepBit by 10.6%, 32.3%, 43.2% and 34.6% mAP with respect to 16, 32, 64 and 128 hash bits, respectively and achieve competitive results compared to CNNH. However, DLGH achieve the best performance in the four methods. This may account for the effective feature extraction technique of DLGH. Generally speaking, our method reports competitive performance results compared to deep-hashing-based image retrieval algorithms. This indicates the proposed method is effective to learn binary descriptors. The experiments also reveal that 2-threshold parameter-based segmentation technique can improve the hashing performance.

Since run time is another aspect used to measure the quality of algorithms, we also test the training time of the four algorithms. Being different from deep learning methods, the training time does not contain two parts (forward time and backward time), while it is composed of segmentation time and retrieval time, which are recorded before and after a plus symbol, respectively. The hash bit is set to 128 because the longer the hash bits, the higher mAP these algorithm achieves. We show the time pass under relatively high retrieval accuracy. According to Table 10, CNNH and DLBH need relatively high train time. The commonality of the two method is that they are both based on CNN framework. In the CNN framework, the fully connected layers will cost too much time. DeepBit makes much effect on optimizing the training time of the deep network. However, it still costs more time than our method. Our method can outperform DeepBit even if plus the segmentation time (183.71ms). About test time, our method also outperforms other methods. Hence, the proposed method is able to achieve the competitive performance with the shortest training time.

Although our method outperforms most of the other algorithms in the experiments, larger datasets such as ImageNet are not concluded because of the limits to the hardware. As is well known, DL methods achieve excellent performance on big datasets. They extract efficient features from numerous images. Therefore, our method provides a simple and quick

Table 10
Run time comparison (128 bit) on CIFAR-10 dataset.

Method	Forward pass(ms)	Backward pass(ms)	Training time(ms)	Test time(ms)
CNNH	885.49	1063.32	1948.81	93.20
DLBH	894.55	1108.53	2003.08	132.23
DeepBit	202.06	310.46	512.52	22.83
OURS	–	–	183.71+223.85	6.43+7.95

solution to image retrieval task and is suitable for relatively small datasets.

4.5. Evaluation of the influence of the adaptive parameter λ on the retrieval results

Fig. 15 shows the influence of the adaptive parameter λ on the retrieval results. On Corel-1000 and CIFAR-10, the best performance is achieved when λ is set to 2/7. While on the SIVAL dataset, 2/5 is the best choice of λ. Considering that SIVAL is clear in the task of distinguishing the salient regions and non-salient regions. We advise 2/7 to be the best value of the adaptive parameter λ.

Hence, we can set its value as 2/7, as this value produces the best results. If it is smaller than 2/7, the shadowed regions are too large and some background interference information cannot be ruled out. When it is bigger than 2/7, the shadowed regions are too small, and the result is close to the image retrieval algorithm based on saliency detection, which does not combine shadowed sets. This choice of parameter is provided with a certain degree of robustness since it is validated on two public datasets.

5. Conclusions

The study developed a shadowed set-based image retrieval algorithm. Based on the saliency-detection-based image retrieval algorithm, our approach exploits the shadowed set theory, splits the image into salient regions, non-salient regions and shadowed regions, and uses the shadowed regions and salient regions as the useful information for the retrieval. For the images having clear salient regions, the algorithm does not increase the redundant regions. In other words, it exhibits a significant level of robustness. A number of comparative experimental studies completed for the two publicly available large datasets show that the approach developed in the study can effectively improve the accuracy of image retrieval. However, there is still some room for further improvement in terms of feature selection. The division of salient regions is sensitive to color, and the salient regions can-not be easily identified if they have the color similar to the color of the background regions. Future studies may focus on ways on how to establish better feature extraction methods so that more semantic content of images could be captured and efficiently used in the classification and retrieval processes. Also, more state-of-the-art methodology will be employed to make our framework more creative.

Acknowledgements

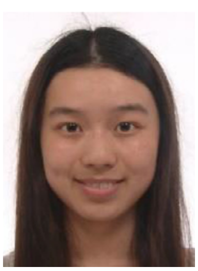
Authors would like to thank the anonymous reviewers for their constructive comments and valuable suggestions. This work is supported by the National Science Foundation of China (Grant No. 61573255, 61673301, 61673299, 61573259), National Key R&D Program of China (Grant No. 213), and Major Project of Ministry of Public Security (Grant No. 20170004).

References

[1] J. Kalpathy-Cramer, A.G. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, H. MLler, Evaluating performance of biomedical image retrieval systems–an overview of the medical image retrieval task at ImageCLEF 2004–2013, Comput. Med. Imaging Graph. 39 (2015) 55–61.
[2] J.M. Patel, N.C. Gamit, A review on feature extraction techniques in content based image retrieval, in: International Conference on Wireless Communications, Signal Processing and NETWORKING, 2016, pp. 2259–2263.
[3] Swain, J. Michael, Ballard, H. Dana, Color indexing, Int. J. Comput. Vis. 7 (1) (1991) 11–32.
[4] A. Gupta, Visual information retrieval, Commun. ACM 40 (5) (1997) 70–79.
[5] A. Oliva, A. Torralba, Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope, Kluwer Academic Publishers, 2001.
[6] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[7] P. Indyk, R. Motwani, Approximate nearest neighbor: towards removing the curse of dimensionality, Theory Comput. 604–613 (11) (2012) 604–613.
[8] M. Raginsky, S. Lazebnik, Locality-sensitive binary codes from shift-invariant kernels, Adv. Neural Inf. Process. Syst. 22 (2009) 1509–1517.
[9] Y. Gong, S. Lazebnik, Iterative quantization: a procrustean approach to learning binary codes, in: Computer Vision and Pattern Recognition, 2011, pp. 817–824.
[10] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, T.X. Han, Contextual weighting for vocabulary tree based image retrieval, in: International Conference on Computer Vision, 2011, pp. 209–216.
[11] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: coupled multi-index for accurate image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1947–1954.
[12] W. Kang, W. Li, Z. Zhou, Column sampling based discrete supervised hashing, in: AAAI Conference on Artificial Intelligence, 2016, pp. 1230–1236.
[13] X. Li, D. Hu, N. Feiping, Large graph hashing with spectral rotation, in: AAAI Conference on Artificial Intelligence, 2017, pp. 2203–2209.
[14] X. Qian, X. Qian, X. Tan, J. Han, Y. Tang, Sketch-based image retrieval by salient contour reinforcement, IEEE Trans. Multimedia 18 (8) (2016) 1604–1615.
[15] A. Papushoy, A.G. Bors, Image Retrieval based on Query by Saliency Content, Academic Press, Inc., 2015.
[16] L. Zheng, S. Wang, Z. Liu, Q. Tian, Fast image retrieval: query pruning and early termination, IEEE Trans. Multimedia 17 (5) (2015) 648–659.
[17] G. Li, Y. Yu, Deep contrast learning for salient object detection, CoRR abs/1603.01976 (2016) 478–487.
[18] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, S.M. Hu, Global contrast based salient region detection, in: Computer Vision and Pattern Recognition, 2011, pp. 409–416.
[19] C. Yang, L. Zhang, H. Lu, X. Ruan, M.H. Yang, Saliency detection via graph-based manifold ranking, in: Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
[20] R. Xia, Y. Pan, H. Lai, C. Liu, S. Yan, Supervised hashing for image retrieval via image representation learning (2014).
[21] K. Lin, H.F. Yang, J.H. Hsiao, C.S. Chen, Deep learning of binary hash codes for fast image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 27–35.
[22] K. Lin, J. Lu, C.S. Chen, J. Zhou, Learning compact binary descriptors with unsupervised deep neural networks, in: Computer Vision and Pattern Recognition, 2016, pp. 1183–1192.
[23] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for scene segmentation, IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2017) 1–1.
[24] E. Shelhamer, J. Long, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, IEEE Computer Society, 2017.
[25] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848.
[26] J. Tang, Z. Li, X. Zhu, Supervised deep hashing for scalable face image retrieval, Pattern Recognit. 75 (2017) 25–32.
[27] W. Pedrycz, Shadowed sets: representing and processing fuzzy sets., IEEE Trans. Syst. Man Cybern. B Cybern. 28 (1) (1998) 103–109.
[28] G. Cattaneo, D. Ciucci, An algebraic approach to shadowed sets, Electron. Notes Theor. Comput. Sci. 82 (4) (2003) 64–75.
[29] W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, Pattern Recognit. Lett. 26 (15) (2005) 2439–2449.
[30] S. Mitra, W. Pedrycz, B. Barman, Shadowed -means: integrating fuzzy and rough clustering, Pattern Recognit. 43 (4) (2010) 1282–1291.
[31] J. Zhou, W. Pedrycz, D. Miao, Shadowed sets in the characterization of rough–fuzzy clustering, Pattern Recognit. 44 (8) (2011) 1738–1749.
[32] W. Pedrycz, G. Vukovich, Granular computing with shadowed sets, Int. J. Intell. Syst. 17 (2) (2010) 173–197.

[33] Y. Zhou, Z. Anfu, et al, Sample data selection method for neural network classifiers, J. Huazhong Univ. Sci. Technol. 40 (6) (2012) 39–43.
[34] M. Cai, Q. Li, G. Lang, Shadowed sets of dynamic fuzzy sets, Granular Comput. 2 (2) (2016) 1–10.
[35] C.I. Gonzalez, P. Melin, O. Castillo, D. Juarez, J.R. Castro, Toward general type-2 fuzzy logic systems based on shadowed sets, Adv. Intell. Syst. Comput. 642 (2017) 131–142.
[36] Y. Yao, S. Wang, X. Deng, Constructing shadowed sets and three-way approximations of fuzzy sets, Inf. Sci. 412–413 (2017) 132–153.
[37] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. SSstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[38] D. Zhou, Ranking on data manifolds, Adv. Neural Inf. Process. Syst. 16 (2003) 169–176.
[39] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior, in: British Machine Vision Conference, 2011, pp. 1–12.
[40] M.M. Cheng, N.J. Mitra, X. Huang, P.H.S. Torr, S.M. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582.
[41] E. Rahtu, J. Kannala, M. Salo, J. Heikkil, Segmenting Salient Objects from Images and Videos, Springer Berlin Heidelberg, 2010.
[42] N.D.B. Bruce, Saliency, attention and visual search: an information theoretic approach, J. Vis. 9 (3) (2009) 5.1.
[43] M. Everingham, S.M.A. Eslami, L.V. Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: a retrospective, Int. J. Comput. Vis. 111 (1) (2015) 98–136.
[44] B. Hariharan, P. Arbelaez, L. Bourdev, S. Maji, J. Malik, Semantic contours from inverse detectors, in: International Conference on Computer Vision, 2011, pp. 991–998.

Hongyun Zhang received the Ph.D. degree in pattern recognition and intelligence system from Tongji University, Shanghai, China, in 2005. She is currently an Associate Professor at Tongji University. She is the author or coauthor of nearly 60 journal papers and conference proceedings in principal curves, pattern recognition, machine learning, granular computing, and rough set. Her current research interests include principal curves, pattern recognition, data mining, image retrieval, and granular computing. E-mail: zhanghongyun@tongji.edu.cn.



Ting Zhang is a master at Tongji University, Shanghai, China. She has participated in the national innovation project for colledge students on image processing since 2015. She was recommended as a master student for admission to Tongji University, in 2016, and received the bachelor degree in Computer Science and Technology from Tongji University in 2017. Her main research interests include image retrieval, shadowed set theory, and granular computing. E-mail: ZT_waa@tongji.edu.cn.

Witold Pedrycz is a professor and Canada Research Chair in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences. He is actively pursuing research in Computational Intelligence, fuzzy modeling, pattern recognition, knowledge discovery, neural networks, granular computing and software engineering. He has published vigorously in these areas. He is an author of eleven research monographs and numerous journal papers in highly reputable journals. Dr. Pedrycz has been a member of numerous program committees of international conferences in the area of Computational Intelligence, Granular Computing, fuzzy sets and neurocomputing. He currently serves as an Associate Editor of IEEE Transactions on Fuzzy Systems, IEEE Transactions on Neural Networks. He is also on editorial boards of over 10 international journals. Dr. Pedrycz is also an Editor-in-Chief of Information Sciences and IEEE Transactions on Systems, Man, and Cybernetics part A. He is the past president of IFSA and NAFIPS. He is a Fellow of the IEEE. E-mail: wpedrycz@ualberta.ca.

Cairong Zhao is currently an Associate Professor at Tongji University. He received the Ph.D. degree from Nanjing University of Science and Technology, M.S. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, and B.S. degree from Jilin University, in 2011, 2006 and 2003, respectively. He is the author of more than 30 scientific papers in pattern recognition, computer vision and related areas. His research interests include computer vision, pattern recognition and visual surveillance. E-mail: zhaocairong@tongji.edu.cn.

Duoqian Miao is a professor of Department of Computer Science and Technology at Tongji University, Shanghai, China. He has published more than 100 papers in international proceedings and journals. His research interests include soft computing, rough sets, pattern recognition, data mining, machine learning and granular computing. E-mail: dqmiao@tongji.edu.cn.