Improved Generic Object Retrieval In Large Scale Databases By SURF Descriptor

Hasan Farsi*
Department of Electrical and Computer Engineering, Birjand University, Birjand, Iran
hfarsi@birjand.ac.ir
Reza Nasiripour
Department of Electrical, Faculty of Engineering, Birjand University, Birjand, Iran
reza.nasiripour@birjand.ac.ir
Sajjad Mohammadzadeh
Department of Electrical, Faculty of Engineering, Birjand University, Birjand, Iran
s.mohamadzadeh@birjand.ac.ir

Received: 26/Nov/2016 Revised: 21/Jun/2017 Accepted: 08/Jul/2017

Abstract

Normally, the-state-of-the-art methods in field of object retrieval for large databases are achieved by training process. We propose a novel large-scale generic object retrieval which only uses a single query image and training-free. Current object retrieval methods require a part of image database for training to construct the classifier. This training can be supervised or unsupervised and semi-supervised. In the proposed method, the query image can be a typical real image of the object. The object is constructed based on Speeded Up Robust Features (SURF) points acquired from the image. Information of relative positions, scale and orientation between SURF points are calculated and constructed into the object model. Dynamic programming is used to try all possible combinations of SURF points for query and datasets images. The ability to match partial affine transformed object images comes from the robustness of SURF points and the flexibility of the model. Occlusion is handled by specifying the probability of a missing SURF point in the model. Experimental results show that this matching technique is robust under partial occlusion and rotation. The properties and performance of the proposed method are demonstrated on the large databases. The average of retrieval rate by the proposed method applied on Oxford landmarks and Corel dataset are 69.68% and 65.79%, respectively. Also, the average of ANMRR measure by the proposed method applied on Oxford landmarks is 0.223 and this criterion for Corel dataset is 0.269. The obtained results illustrate that the proposed method improves the efficiency, speeds up recovery and reduces the storage space.

Keywords: Object retrieval; Speeded Up Robust Features (SURF); Large-scale; Supervised; Training-Free.

1. Introduction

Recently, the problem of specific object retrieval from an image database is a very challenging area that many researchers look for the best solution. In the other words, by selection of a particular object in a given query image, an object retrieval system should return a set of representative images that contain object. Object detection plays important roles in computer vision which includes image retrieval, intelligent transportation systems and surveillance. Object detection has used in recent researches on object tracking, object recognition, and other object-based approaches.

Object retrieval is faced many challenges:

1. Illumination condition/position: Illumination changes that occur during a day can be added a shadow to object interesting. Also, climate condition causes changes in illumination condition.
2. Geometric distortions: change in the position of object is called as geometric distortion. When two objects in images are matching, object is geometric distortion, to reduce rate of recognition.
3. Rotation: an object retrieval system must have ability adapt when position of object is rotation.
4. Scale: this challenge occurs when the size of object in an image is changed.
5. Occlusion: when object in an image is not completely visible that is called occlusion.

Object retrieval is mainly divided into two parts: category retrieval and detection. The aim of object category retrieval is to classify a given object into several predefined categories whereas the aim of object detection is to separate desired objects from the background in a target image.

Generally, the object detection seeks any object of a particular class in a test image to answer the question “how many objects are in the image, and where do they are located.

The detection area is divided to two parts, appearance-based and contour-based approaches [1-5]. In contour-based approaches, learning algorithms are more common to use. In the other words, these methods first extract some features such as color, texture, and illumination change. Then they use learning methods [6-8].

In appearance-based approaches, first, ‘interest points’ are selected at distinctive locations in the image, such as corner, blobs, and T-junctions. An interest point detector has most valuable property that is its repeatability,

* Corresponding Author

whether it reliably finds the same interest points under different viewing conditions. Then, the neighbors of every interest point are represented by a path descriptor. This descriptor has to be distinctive, and also robust against noise, detection errors, geometric, and photometric deformations [9-10].

Category of retrieval is caused limitation within computer vision. To achieve this purpose, Agarwal et al. and Barnard et al. have proposed the methods that use learning algorithm [11]. In other words, this method used many number of training data. Then, training process is done by the classifier known. Finally, the test image is added to classifier to matching process with training images.

In the recent years, machine vision methods extract features of image and object to achieve the purpose of object recognition. There are some approaches that use low-level features. For example, volumetric descriptor [14], surface distribution [15], geometry [16-17] have used to extract the feature for object retrieval. While these methods act efficiently well under an engineered environment where object pose and illumination are strictly controlled, it is no longer feasible under slight positions or illumination variations because of the limited computation power and nearly infinite possible combinations of pose and lighting.

To overcome this problem, it is proposed to use high-level feature based methods instead of searching all possible model positions through the image. High-level feature based methods extract object features which are mostly invariant among different positions, orientations and lightings. Scale Invariant Feature Transform (SIFT) [-18] is an efficient algorithm that is widely used in object recognition, image stitching, stereo vision and various computer vision researches.

One of the popular methods to detect objects by using a single query image without training is keypoint-based matching. Most methods use many keypoints which are relatively stable in the image and calculate the local invariant descriptors of the patches around the keypoints, such as SIFT and shape context [19]. Therefore, the object matching is translated into a set of local descriptors on the keyoints. Some approaches use densely computed descriptors [20]. The reported method in [21] used local regression kernels the descriptor with a matrix generalization of the cosine similarity measure as the comparison method.

There are other methods which are based on re-ranking. In [22], the authors have proposed a query expansion (QE) which causes improvements of retrieval performance. In [23], a discriminative query expansion (DQE) which uses matching learning method has been proposed. In [24], the authors proposed the method based on vector similarity.

In this paper, we propose a new method which uses only one query image to detect the object, without training, as shown in Figure 1. The proposed method can be decomposed into two parts: first, learning the probabilistic model for a specific object, and second, matching the model in a test image. The model is constructed with SURF points that are acquired from the image. Information of relative positions, scale and orientation between SURF points are calculated and constructed into a probabilistic model of the object. Then, dynamic programming is used to find the best combination among all SURF points. In the matching step, dynamic programming is also used to find the best model among all SURF points. Since a SURF point in the model can be lost during matching in the test image, this situation is handled by defining a missing probability for a SURF point. When a SURF point is lost, a virtual SURF point with the best match in the model is inserted; this is necessary to calculate the probability of the relation among SURF points in the model.

We evaluate the proposed model on two datasets. First dataset is the building dataset that comprise 5K images of Oxford landmarks where “landmarks” means a particular part of the building. We use a set of images comprising 11 different landmarks. The images for each landmark are retrieved from [12]. In addition, we use the Corel dataset to evaluate the proposed system [13]. This dataset includes 1000 different images. The images are divided into 11 classes, including early humans, elephants, flowers, buses, horses, etc. that we used to three classes, including planes, buses and dinosaurs. Some examples of the images from these dataset are shown in figure 2

Fig. 1 Object detection using only one query.

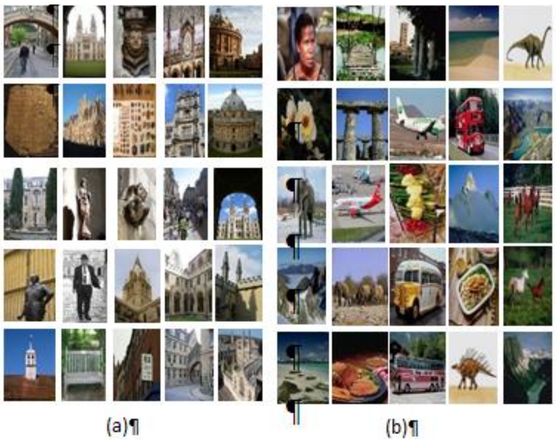

Fig. 2 Example images from dataset. (a) 25 randomly sampled images from Oxford dataset, (b) 25 randomly sampled images from Corel dataset.

This paper is organized as follows: In Section 2, the proposed object retrieval system is introduced. The proposed system can be divided into two parts, first part is object retrieval and second part is object detection. In section 3, experimental results have been explained. Finally, the conclusion is drawn in Section 4.

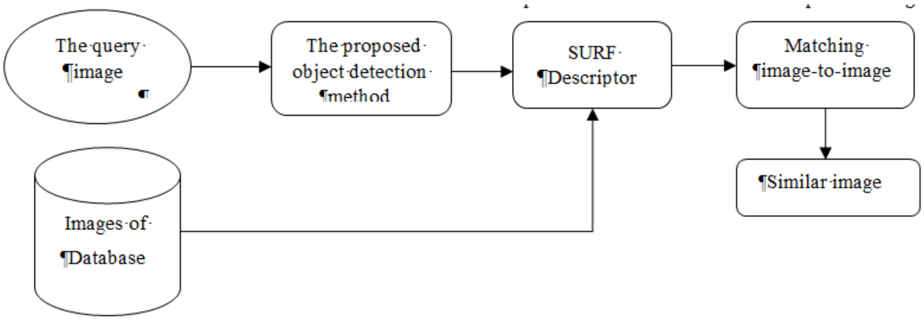

Fig. 3 Block diagram of the proposed Object Retrieval method.

2. Proposed Object Retrieval Method

The proposed method detects the objects with a single query image. The key problem is to represent the target class from the query image, which could be a typical real image. The detection process is very similar to “template matching”. The query image is used as a standard “template”, with the test images matched to this “template” to find the objects.

A block diagram of the proposed object retrieval method is illustrated inFigure3. The proposed method acts in different procedure using only one query image to detect the object without training. We extract features from query image and dataset images. This is achieved by SURF descriptor which is detailed in subsection 2.1. Since target image requires object detection, the proposed object detection for extraction of the target object is detailed in subsection 2.2. In subsection 2.3, the details of object matching using the proposed object retrieval method have been described.

2.1 SURF Algorithm

This section reviews the SURF algorithm which was proposed by Bay H, Tuytelaars, Gool L.V. in 2006. This algorithm is similar to SIFT algorithm. However, it is faster than SIFT in terms of calculation speed. In this section, SIFT descriptor is firstly described and then difference between SURF and SIFT descriptors is demonstrated. Both descriptors are applied in four steps:

2.1.1. Scale-space extreme detection
2.1.2. Key point localization
2.1.3. Orientation assignment
2.1.4. Key point descriptor

2-1-1- Scale-Space extreme detection

In SIFT, in order to detect the extreme or in other words position of interest point, image in a space called ‘scale-space’ is defined. The Interest points are stable features under different directions and scales. Scale space is performed by filtering image with sequence of Gaussian filter. Scale space of image is constructed as pyramid form. As observed in Figure 4, the scale space is composed of several octaves such that each octave is composed of five levels. In the first octave, the first level of image filtering is performed using Gaussian function with σ=0.5. Subsequently four levels are derived by convolution of image with Gaussian kernel $\sqrt{2}\sigma$, 2σ, $2\sqrt{2}\sigma$ and 4σ, respectively. In order to obtain the first level of the new octave, sub-sampling operation is performed by sampling the original image with 2:1 rate. The new levels in new octaves are constructed by using the same Gaussian kernels. The same process is performed for the construction of new octaves. The difference of levels in current scale is used to approximate the difference of Gaussian filters (DOG) or Laplace - Gaussian (LOG). Finally, the interest points are selected by local extreme points in the DOG scale space.

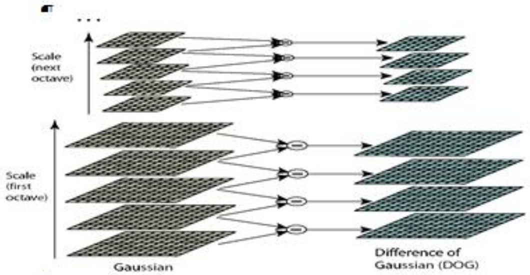

Fig. 4 Scheme of SIFT algorithm for extreme detection.

But in SURF descriptor uses a hessian matrix to find interest points. The determinant of a hessian matrix is an expression of the local change around the area. Given a point, $X = (x, y)$, in an image, I, the hessian matrix,H (x, σ), in Xat scale, σ, is defined as:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & l_{yy}(x,\sigma) \end{bmatrix} \tag{1}$$

Where $L_{xx}(x,\sigma)$, $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$ denote the convolution of the second order Gaussian derivative $\frac{\partial^2 g(\sigma)}{\partial x^2}$, $\frac{\partial^2 g(\sigma)}{\partial xy}$, $\frac{\partial^2 g(\sigma)}{\partial y^2}$ with the image at point $X = (x, y)$, respectively. $g(\sigma)$ is given by:

$$g(\sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2}$$

The convolution is very time-consuming. Hence, it is approximated and speeded-up by using integral images and approximated kernels-box filters. The integral image, I(X), at a location, X=(x,y), represents sum of all pixels in the input image, I, within a rectangular region formed between the origin and the position of X.

$$I(X) = \sum_{i=0}^{i\leq x}\sum_{j=0}^{j\leq y} I(x,y) \qquad (3)$$

Using the box filters, the hessian determinant can be approximated by (4):

$$\det(H_{appx}) = D_{xx}D_{yy} - \left(wD_{xy}\right)^2 \qquad (4)$$

$\mathrm{D_{xx}}$, $\mathrm{D_{xy}}$ and $\mathrm{D_{yy}}$ denote the convolution of the box filters with the image at point,$\mathrm{X} = (\mathrm{x},\mathrm{y})$, respectively.

Scale spaces are usually achieved by image pyramids. The image pyramids in SURF are constructed by changing the size of box filters rather than reducing the size of image. Initial scale layer is output of 9×9 filtering, and the corresponding scale, $\sigma = 1.2$. The following layers are obtained by filtering the image with gradually bigger masks, such as:9×9, 15×15 , 21×21, 27×27.

2.1.2. Key point localization

Both SURF and SIFT descriptors proceed at the way to locate the key points. In order to find the key points, each pixel is compared with 26 pixels. If the pixel value is lower or higher than all 26 pixels, this point is considered as a key point. Otherwise, this point is removed and the algorithm is applied on next pixel (or point). The computational complexity of the algorithm is low because most points in first stage of the algorithm are removed. Then the key points are interpolated in scale space image.

2-1-3- Orientation assignment

In SIFT method, in order to assign the orientation; a window is constructed around of each key point. Orientation histogram is constructed by using gradient directions of the points within the window. Each orientation histogram for each key point contains 36 parts which covers 360 degree of directions. In this histogram, the orientation having highest value is considered as the dominant orientation.

In SURF method, rotation invariance is achieved by detecting the dominant orientation of each feature point. The dominant orientation is estimated by calculating sum of the horizontal and vertical Haar wavelet responses within a sliding orientation window with angle of $\frac{\pi}{3}$. The two summed responses constitute a vector, and the longest vector lends its orientation to the feature point. The size of the Haar filter kernel is scaled to $4s \times 4s$where s is the scale of the feature point. The responses are weighted by a Gaussian function centred at the feature point.

Finally, these description vectors are normalized to unit vectors to provide robustness against contrast.

2.2 The proposed Object Detection Method

A block diagram of the proposed object detection method is illustrated in Figure 5. In this section, the gradients in different directions and scale have been calculated, and comparative operator is defined the inspired Retina eye model, and it is called Gabor functions. The simple method is proposed to combine multi-directional and multi-scale edge presentation. Eye map the input image to the various features that start in the retina. We have used a kind of Gabor filter by changing these parameters of filter. Therefore, many different scales of the edge of the image are extracted [25]. We have selected Gabor filter because frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination.

Equations (5) models human visual system by a Gabor function which includes direction, scale, frequency and spatial model of the desired shape [26].

$$R(x - x_c, y - y_c)_{\lambda,\sigma,\theta,\varphi,\gamma} = R_0 \exp\left(-\frac{u^2 +,\gamma^2 v^2}{2\sigma^2}\right)\cos\left(2\pi\frac{u}{\lambda} + \varphi\right)$$

$$u(x - x_c, y - y_c, \theta) = (x - x\_c)cos\theta - (y - y\_c\ )sin\theta$$

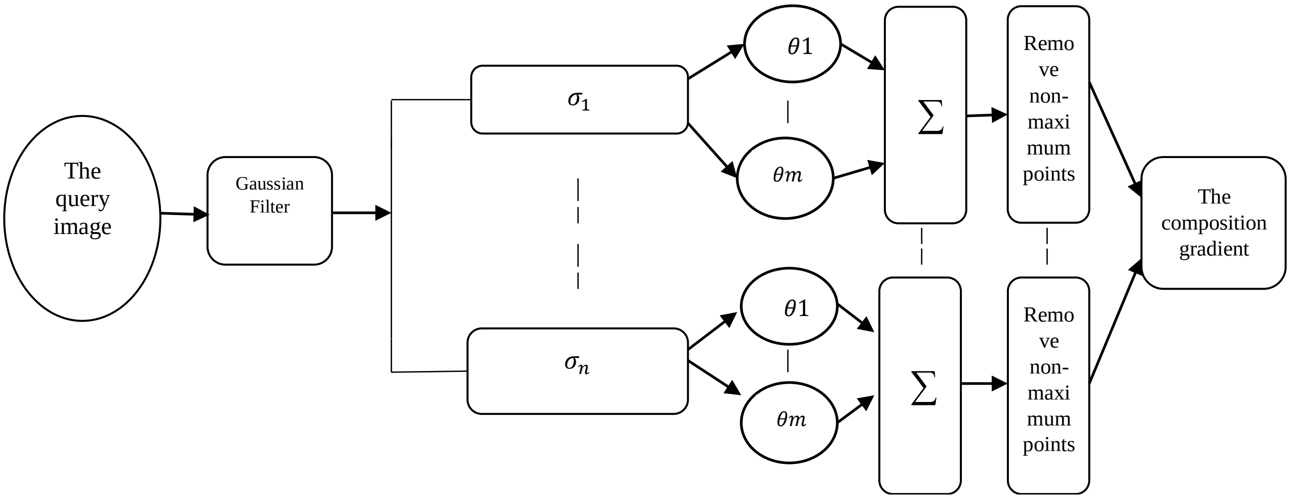

Fig. 5 Block diagram of the proposed object detection method.

$$v(x-x_c, y-y_c, \theta) = (x-x_c)sin\theta + (y - y_c)cos\theta \qquad (5)$$

In this equation, x_c and y_c are the rotation center of the filter to the preferred angle, θ, that are placed relative to the origin. σ is standard deviation, λ length wave, and φ filter phase difference. Equation (6) has obtained by using query image and Equation (5):

$$t(x_c, y_c, \theta) = \iint I(x,y)\, R(x-x_c, y-y_c) d_x d_y \cong \sum_{m=1}^{M}\sum_{n=1}^{N} I(m\Delta x, n\Delta y) R(m\Delta x - x_c, n\Delta y - y_c)\, \Delta x \Delta y \qquad (6)$$

In this section we have used from the following approximations: $\lambda = 2\pi\sigma^2, \emptyset = \frac{\pi}{2}$

First, we have used the Gaussian filter with fit σ to reduce noise input image. Second, for each scale, Gabor filter is convoluted with the original image in different scales (n) and different directions (m), -90 to +90 degrees. Therefore, the M × N gradient function are obtained from the original image. Then, the sum of weighted responses of Gabor filter is used to estimate the total gradient vector. In any scale, local gradient function with maximum values achieved by applying the remove the non-peak local maximum algorithm. So, in any scale, an approximate map has been achieved. Figure6 shows an illustration of target object (shape) extraction enhancement with the proposed object detection method.

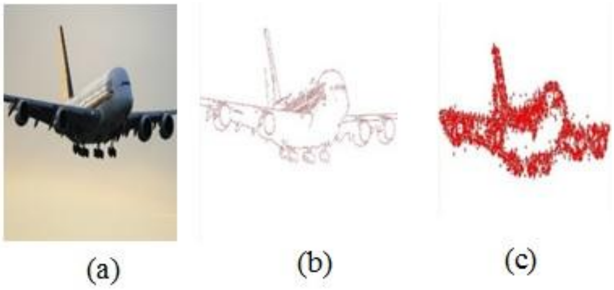

Fig. 6. Target object detection by the proposed Object Detection method and SURF descriptor. a) Original image, b) Detection method and c) extract SURF descriptor from object.

2.3 Object Matching

Figure7 shows object matching between two images for instance, where right image is query image and left image is matched image. We first acquire all the SURF points in the target image. In other words, each image is represented as isolated SURF points.

A model is defined by a set of N_m nodes. The set of nodes $M = m_a$is indexed by $\alpha = 1,2,\dots,N_m$ and each m_a will correspond to a SURF point in the image. Each node has attributes of $(z_\alpha, s_\alpha, \theta_a, A_\alpha)$ where z_α denotes the spatial location, $s_\alpha$indicates the feature size, θ_a denotes the orientation, and $A_\alpha$represents the appearance. There is also a binary value variable,u_α, that specifies whether a node in the model could be found in the image.

There are $N_m - 2$ triplet-cliques C in a model; each triplet-clique is also labeled by α and is a set of 3 nodes $C_\alpha = \{m_\alpha, m_{\alpha+1}, m_{\alpha+2}\}$.

We define the model parameters as $\omega = (\omega^A, \omega^S)$, where $\omega^A = \{\omega_\alpha^A\}$ are the appearance parameters and $\omega^S = \{\omega_\alpha^S\}$ are the shape parameters. The model parameters, ω , can be decomposed into N_m nodes $m_1, \dots, m_{N_M}$ and $N_M - 2$ cliques $C_1, \dots, C_{N_{M-2}}$; each node represents an appearance vector u_a^A , $\alpha = 1, \dots, m$; each clique represents a shape vector u_a^S. To match the model to the image, first we match the SURF points in the target image to the first clique in the model. To reduce the enormous possible combinations of any 3 SURF points in the image, for each SURF point, we first find the nearest neighbour of its appearance vector among the appearance vectors of the first 3 nodes (m_1, m_2, m_3). We restrict them to only match to the closest one. We then store each 3 SURF points that matches (m_1, m_2, m_3)in the model as the chain,H_k, and its chain probabilities given the model parameters as P_h^k. $P_h^k$is defined as the clique probability P^c times the observed probability P^0.

For the first iteration every SURF point in image I_1 are added to one of the available cliques that would give the highest product of maximum P_i^c among images $I_2 \dots I_{NI}$. The P_i^cin Image I_i is calculated by assuming the model parameter u^A and u^S is equal to the appearance vector and shape vector of the corresponding SURF points in C_{1j}^a; and the model parameter $\sum A$ and $\sum S$ is set to the identity matrix. The maximum clique probability in I_i is the maximum among all combinations of Surf points in I_i.

Adding a new node to the clique would create a new clique which is composed of one new node and two previous nodes. In further iterations, new SURF points in I_1 will be added it this set of new clique instead of the previous ones; each SURF points would be added to the clique that would give the highest product of maximum clique probabilities among images$I_2 \dots I_{NI}$ times the chain probabilityP_{kt}^h.

The iteration is stopped when none of the chain probability is higher than a threshold.

In addition to 3 SURF points stored as the chain, we also store a case in which 1 SURF point is lost. We generate a virtual SURF point which has the appearance vector and the location giving the maximum$\mathrm{P^c}$, and store 2 actual SURF points and 1 virtual SURF point into a chain $\mathrm{H_k}$ with chain probability $\mathrm{P_h^k = P^cP^0}$. The only restriction is that one node in a clique could be a virtual SURF point. For each iteration, we match one of the SURF points in the image to one of the chains from the last iteration which corresponds to the maximum$\mathrm{P_h^k}$. The iteration ends when all nodes in the model are matched. The chain H_k containing the highest P_h^k is the most possible location of the object in the image.

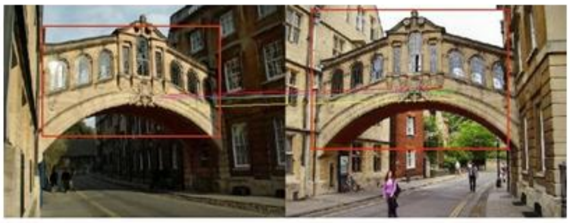

Fig. 7 Example of object matching.

3. Experimental Results

3.1 Evaluation Measures

There are many measures for evaluation of retrieval methods. In this paper, we use three measures ANMRR, precision and recall. Precision and recall are used for evaluation of most object retrieval methods.

The Average Normalized Modified Retrieval Rank (ANMRR) is an objective measure which summarizes the performance of system into a scalar value. It is defined from MPEG-7 research group [27].

First, we denote NG(q), K(q), R(k) as follows:

NG(q) : The number of the ground truth images for a query image, q.

K(q) =min[4. |NG) q)|,2. max{|NG(q)|, ∀q}]

R(k)= rank of an image, k, in retrieval results.

Rank(k) is obtained by:

ank(k) (7)

$$\begin{cases} R(k) & \text{if } R(k) \leq K(q) \\ 1.25K & \text{otherwise} \end{cases}$$

Using Equation 9, Average Rank, AVR(q), for query q, is given by:

VR(q) = <Rank(k)>. (8)

However, for ground truth sets with different sizes, the AVR(q) value depends on NG(q). To minimize the influence of variations in NG(q), Modified Retrieval Rank, MRR(q), is obtained by:

[RR(q) = AVR(q)-0.5[1+|NG(q)|] (9)

The upper bound of MRR(q) depends on NG(q). To normalize this value, Normalized Modified Retrieval Rank, NMRR(q), is obtained by:

$$\text{MRR(q)} = \frac{AVR(q) - 0.5[1+|\text{NG(q)}|]}{1.25K(q) - 0.5[1+|\text{NG(q)}|]} \quad (10)$$

This measure is zero for perfect performance and approaches to one as performance worsens. The ANMRR of the dataset is finally given by averaging the NMRR(q) over all the q's

NMRR = <NMRR(q)>. (11)

Precision is the fraction of returned images that are relevant to the query image. Recall is the total number of relevant images with respect to the total number of relevant images in the dataset according to a priori knowledge. If we denote T as the set of returned images and R as the set of all images relevant to the query image, then the precision and recall criteria are given by Equations (12) and (13), respectively [28]

$$Precision = \frac{|T \cap R|}{|T|} \quad (12)$$

$$Recall = \frac{|T \cap R|}{|R|} \quad (13)$$

The number of relevant images is computed and the precision and recall in reach of retrieved images for all query images are obtained. We next consider the average of these precisions and recalls for each number of retrieved images as the precision and recall of each method for each number of retrieved images.

We use various methods to evaluate the proposed system. Montagna and Finlayson [29] proposed a method using the combination of precision and recall criteria as the performance measures for object retrieval method. According to Montagna and Finlayson [29], the following measures have been adopted:

P (0.5), precision at 50% recall (i.e. precision after retrieving 1/2 of the relevant documents).

P (1), precision at 100% recall (i.e. precision after retrieving all of the relevant documents, P(1) is the percent of crossover point of precision and recall). We use these values because precision and recall are considered in relation to each other and they are not meaningful if taken separately. To evaluate performance we use Mean Average Precision (MAP) for the landmark.

3.2 Indexing Results

In this section, we evaluate the performance of the proposed object retrieval on the building dataset of Oxford landmarks and Corel datasets. We use a set of images in Oxford dataset that comprising 11 different landmarks. The images for each landmark are retrieved from [12]. Also, Corel dataset include 1000 different images. These images are divided into 11 classes, including early humans, elephants, flowers, buses, horses, etc. that we use them into three classes, including planes, buses and dinosaurs.

3.2.1. The Corel Dataset

We evaluate the proposed method with color layout descriptor [30], dominant color descriptor [31], Patch based HOG-LBP [36] and Scalable color descriptor [30] in Corel dataset. We apply the proposed method to retrieve relative images, therefore we use query image. The P(0.5), P(1) and ANMRR of the proposed method, color layout descriptor, dominant color descriptor and Scalable color descriptor on Corel dataset are represented in Figure 8.

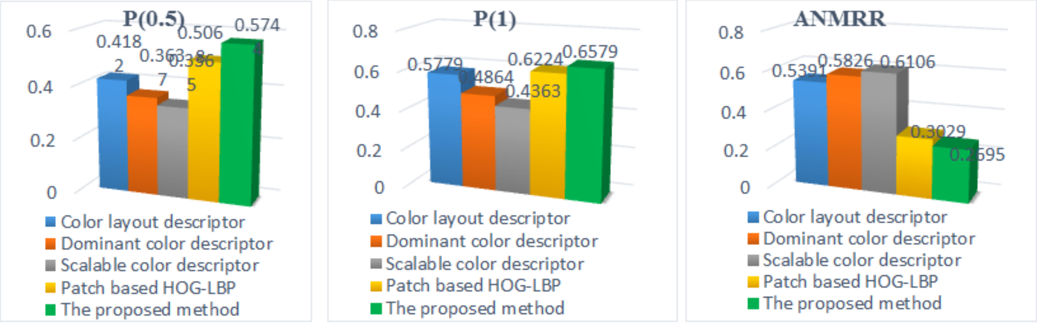

Fig. 8 P(0.5), P(1) and ANMRR for different landmarks on Corel dataset.

3.2.2. The Oxford Landmarks Datasets

In order to demonstrate the effectiveness of the proposed method, we apply the proposed method on Oxford landmarks datasets. We compare the proposed method with cosine model [32], general (content-unaware) language modeling approach (LM) [33] and two variants of the matting-based COR model [34]. The performance of the proposed method, Cosine, CORm, CORa and LM methods of the 11 landmarks on the Oxford 5K dataset is shown in Figure9. As observed, the performance of the proposed method in "All Souls"is better than other methods. After the proposed method, CORm, CORa, Cosine and LM have better performance, respectively. As observed, in each landmarks, the performance of the proposed method is better than other methods except in "Balliol" and "Keble" landmarks.

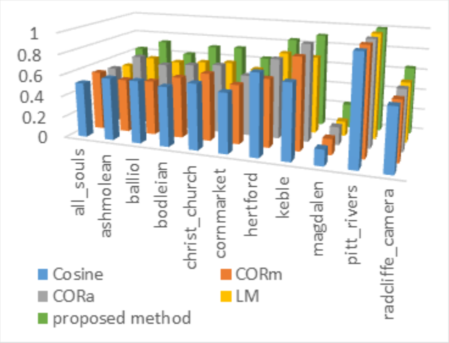

Fig. 9 P(1) for different landmarks on Oxford dataset.

We compare the proposed method with Ng et al [37], SPoC [38], R-MAC [39], CroW [40] and uCroW [40]. The MAP performance of the proposed method, Ng et al, SPoC, R-MAC, CroW and uCroW methods of the Oxford dataset is shown in Figure 10. Table 1 summarizes the test results on the Oxford 5K dataset. The P(0.5), P(1), ANMRR and average of these parameters of the proposed method, CORm, CORa and LM methods are represented in this table. Once again, in Table 2 the best score for each metric is in bold face.

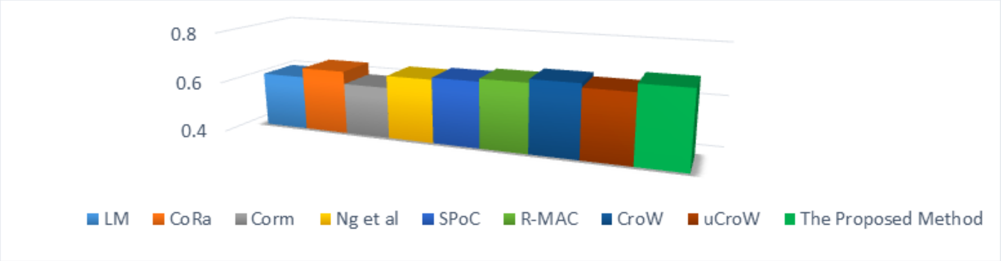

Fig. 10 MAP for different approaches on Oxford dataset.

Figure 11 shows the object matching of the proposed method on 11 landmarks of Oxford dataset. In this figure, each sub-figure has two images where right image is query image and left image is matched image. We present matched points by using color lines. These points are obtained by using SURF descriptor explained in section 3. According to these figures, the proposed matching method is robust under illumination, rotation, scaling and partial occlusion.

4. Conclusions

We proposed a new method for object retrieval object using a single query image without training or free-training. We used SURF algorithm for object matching. The detection process is very similar to "template matching". The query image is used as a standard "template", with the test images matched to this "template" to find the objects. The obtained results

Table 1. P(1), P(0.5) and ANMRR of different method in Oxford 5K dataset.

Method	Datasets											
		All Souls	Ashmolean	Balliol	Bodleian	Christ Church	Cornmarket	Hertford	Keble	Magdalen	Pitt rivers	Radcliffe Camera	Average
Cosine	P(0.5)%	61.72	59.09	71.21	66.34	62.87	59.71	82.21	70.01	17.13	50	68.89	60.83
	P(1)%	53.21	62.17	61.18	57.09	62.50	57.21	78.24	71.15	13.76	100	59.31	61.43
	ANMRR	0.321	0.182	0.361	0.241	0.290	0.354	0.072	0.061	0.721	0	0.352	0.269
CORm	P(0.5)%	65.70	51.08	63.12	68.34	65.37	58.90	70.02	86.32	15.13	50	66.79	60.07
	P(1)%	56.09	49.47	53.33	58.12	64.01	57.05	65.29	85.81	8.76	100	64.10	60.19
	ANMRR	0.289	0.191	0.426	0.236	0.275	0.349	0.090	0.056	0.721	0	0.371	0.273
CORa	P(0.5)%	63.74	68.02	**73.25**	74.34	67.32	61.71	82.00	**92.00**	17.21	50	68.91	64.41
	P(1)%	55.11	67.47	**63.31**	65.18	67.14	58.14	78.07	**91.11**	11.71	100	59.09	65.12
	ANMRR	0.311	0.156	**0.335**	0.185	0.262	0.338	0.065	**0.034**	0.710	0	0.351	0.250
LM	P(0.5)%	60.68	61.10	70.00	71.30	63.87	62.02	82.17	74.22	14.45	50	67.65	61.59
	P(1)%	53.17	59.47	59.24	61.20	64.41	59.78	77.90	73.39	8.61	100	59.02	61.47
	ANMRR	0.329	0.168	0.361	0.205	0.289	0.332	0.081	0.074	0.741	0	0.364	0.268
The proposed method	P(0.5)%	**73.10**	**74.28**	72.09	**81.41**	**73.37**	**65.72**	**88.17**	91.11	**24.45**	50	**73.91**	**69.78**
	P(1)%	**64.17**	**73.02**	61.72	**71.30**	**72.05**	**63.21**	**84.05**	90.06	**22.81**	100	**64.11**	**69.68**
	ANMRR	**0.261**	**0.137**	0.352	**0.106**	**0.231**	**0.298**	**0.059**	0.039	**0.670**	0	**0.305**	**0.223**

showed that using the proposed system object retrieval could improve the performance in the two datasets; Oxford landmarks and Corel datasets. In addition, the proposed method results in improving efficiency, speeds up recovery and reduces the required space for storage. The experimental results also show that the proposed matching technique is robust under partial occlusion, rotation and scaling. This method is very useful for generic or immediate object detection tasks, because of using single query image.

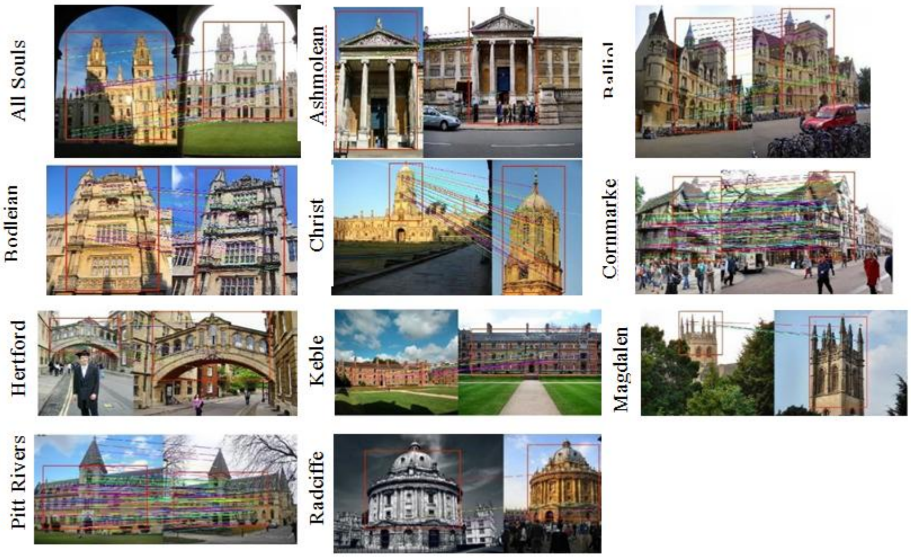

Fig. 11 Example of object matching on the Oxford landmarks database. Right image is query image and left image is image matched.

References

[1] P. Kontschieder, H. Riemenschneider, M. Donser, H. Bischof, “Discriminative learning of contour fragments for object detection”, In Proc. Brit. Mach. Vis. Conference. pp. 1-12, 2011.#

[2] X. Meng, Z. Wang, L. Wu, “Building global image feature for scene recognition”, Pattern Recognition. pp. 373-380, 2012.#

[3] B. Leibe, K. Schindler, N. Cornelis, L. Van Gool, “Coupled object detection and tracking from static cameras and moving vehicles", IEEE Transaction. Pattern. Anal Mach Intelligence, pp. 1683-1698, 2008.#

[4] H. Riemenschneider, M. Donoser, H. Bischof, “Using partial edge contour matches for efficient object category localization”, in Proc. Europa Conference Computer, pp. 29-42, 2010.#

[5] X. Yang, H. Liu, Latecki, “Contour-based object detection as dominant set computation”, Pattern Recognition, pp. 1927-1936, 2012.#

[6] V. Ferrari, T. Tuytelaars, and L. Van Gool, “Object detection by contour segment networks”, In (ECCV), 2006.#

[7] V. Ferrari, F. Jurie, and C. Schmid, “Accurate object detections with deformable shape models learnt from images”, In (CVPR), 2007.#

[8] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, ‶Groups of adjacent contour segments for object detection″, (PAMI), 2008.#

[9] X. Meng, Z.Wang, and L.Wu, ‶Building global image features for scene recognition″, Pattern Recognition, Vol.45, No.1, pp. 373–380, 2012.#

[10] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool, ‶Coupled object detection and tracking from static cameras and moving vehicles″, IEEE Trans. Pattern Anal. Mach. Intell, Vol.30, No.10, pp. 1683–169, 2008.#

[11] S. Agarwal, A. Awan, and D. Roth, ‶Learning to detect objects in images via a sparse, part-based representation″, IEEE PAMI, 2004.#

[12] Pages. Available: http://www.flickr.com.#

[13] Pages. Available http://wang.ist.psu.edu/docs/related.#

[14] J. Tangelder, R. Veltkamp, ‶Polyhedral model retrieval using weighted point sets″, Int. J. Image Graph, 2003.#

[15] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, ‶Shape distributions″, ACM Transaction, 2002.#

[16] A.Makadia, K.Daniilidis, ‶Spherical correlation of visual representations for 3D model retrieval″, Int.J.Computer, Vol.89, No.2, 2010.#

[17] L. Zhu, Y. Chen, A. Yuille, ‶Unsupervised learning of a probabilistic grammar-Markov models for object categories″, Pattern Analysis and Machine Intelligence, IEEE Transactions, pp. 114-128, 2009.#

[18] D. G. Lowe, ‶Object recognition from local scale-invariant feature″, In Proceedings of the International Conference on Computer Vision, pp. 1150-, 1999.#

[19] S. Malik, J. Puzicha, ‶Shape matching and object recognition using shape context″, IEEE Transaction. pp. 509-522, 2002.#

[20] E. Shechtman, M. Irani, ‶Matching local self-similarities across images and videos″, IEEE Conference on Computer Vision and Pattern Recognition, 2007.#

[21] H. Seo, P. Milanfar, ‶Training-free, generic object detection using locally adaptive regression kernels″, IEEE Transaction, pp. 1688-1704, 2010.#

[22] O. Chum, J. Philbin, J. Sivic, M. Isard, A.A. Zisserman, ‶Total recall: automatic query expansion with a generative feature model for object retrieval″, IEEE 11th International Conference, 2007.#

[23] R.Arandjelović, A.Zisserman, ‶Three things everyone should know to improve object retrieval″, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.#

[24] D.Qin, S.Gammeter, L.Bossard, T.Quack, L.V.Gool, ‶accurate object retrieval with k-reciprocal nearest neighbors″, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.#

[25] S. E. Grigoresco, N. Petkov, P. Kruizinga, ‶Comparision of texture features based on Gabor filters″, IEEE Transaction, pp. 1160-1167, 2002.#

[26] C. Grigorescu, N. Petcov, M. A. Westenberg, ‶Contour detection based on nonclassical receptive field inhibition″, IEEE Transaction, pp. 729-739, 2003.#

[27] Chun, Kim, Jang, ‶Content-based image retrieval using multi-resolution color and texture features″, IEEE Transaction, pp. 1073-1084, 2008.#

[28] Veganzones M. A, Grana, M, ‶A spectral/spatial CBIR system for hyper spectral images″, IEEE J. Sel. Top. Earth Obs. Remote Sens, pp. 488-500, 2012.#

[29] Montagna R, Finlayson, G.G, ‶Padua point interpolation and L^P-norm minimization in color-based image indexing and retrieval″, IET Image Process. pp. 139-147, 2012.#

[30] R. Troncy, B. Huet, S. Schenk, ‶Multimedia semantics, desktop edition (XML): metadata, analysis and interaction″, John Wiley & Sons Inc., New York, (1st edition), pp. 36-54, 2011.#

[31] A. Ibrahim, A. Zou'bi, R. Sahawneh, M. Makhadmeh, ‶Fixed representative colors feature extraction algorithm for moving picture experts group-7 dominant color descriptor″, Journal of Computer Science, pp. 773-777, 2009.#

[32] J. Philibin, O. Chum, M. Isard, J. Sivic, A. Zisserman, ‶Object retrieval with large vocabularies and fast spatial matching″, Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2007.#

[33] B. Geng, L. Yang, C. Xu, ‶A study of language model for image retrieval″, IEEE Int. Conf. Data Mining Workshops, IEEE Computer Society, pp. 158-163, 2009.#

[34] Linjun Yang, Bo Geng, Yang Cai, Alan Hanjalic, Xian-Sheng Hua, ‶Object retrieval using visual query context″, IEEE Transactions on multimedia, 2011.#

[35] B. Herbert, E. Andreas, T. Tinne, V.G. Luc, ‶Speeded-up robust features (SURF)″, Computer vision and image understanding (CVIU), pp. 346-359, 2008.#

[36] J. YU, Z. C, T. W AND X. ZHANG, "Feature integration analysis of bag-of-features model for image retrieval", Nero computing, [On-line], 2013.#

[37] J. Ng, F. Yang, and L. Davis. "Exploiting local features from deep networks for image retrieval", In Computer Vision and Pattern Recognition Workshops (CVPRW), 2015.#

[38] A. Babenko and V. Lempitsky, "Aggregating local deep features for image retrieval", In International Conference on Computer Vision (ICCV), 2015.#

[39] G. Tolias, R. Sicre, and H. J_egou, "Particular object retrieval with integral max-pooling of CNN activations", arXiv preprint arXiv:1511.05879, 2015.#

[40] Y. Kalantidis, C. Mellina, and S. Osindero, "Cross-dimensional weighting for aggregated deep convolutional features", arXiv:1512.04065, 2015.#

Hasan Farsi received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as professor in communication engineering in department of Electrical and Computer Eng., University of Birjand, Birjand, IRAN.

Reza Nasiripour was born in Mashad in 1990. He received the B.Sc. and M.Sc. degrees in electrical communication engineering from University of Birjand, Birjand, Iran in 2012 and 2014, respectively. He is currently Ph.D. student in Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. His research interests include Image and Video Processing, Pattern Recognition and Machine Learning

Sajjad Mohammadzadeh received the B.Sc. degree in communication engineering from Sistan & Baloochestan, University of Zahedan, Iran, in 2010. He received the M.Sc. and Ph.D. degree in communication engineering from South of Khorasan, University of Birjand, Birjand, Iran, in 2012 and 2016, respectively. Now, he works as assistant professor in Faculty of Technical and Engineering of Ferdows, University of Birjand, Birjand, Iran. His area research interests include Image and Video Processing, Retrieval, Pattern recognition, Digital Signal Processing, Sparse Representation, and Deep Learning.