Pattern Recognition 46 (2013) 2890–2901

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Interest point detection using rank order LoG filter

Zhenwei Miao*, Xudong Jiang

School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Link, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 2 November 2011
Received in revised form
28 February 2013
Accepted 26 March 2013
Available online 12 April 2013

Keywords:
Interest point detection
Image matching
Weighted rank order filter
Repeatability

ABSTRACT

This paper proposes a novel nonlinear filter, named rank order Laplacian of Gaussian (ROLG) filter, based on which a new interest point detector is developed. The ROLG filter is a weighted rank order filter. It is used to detect the image local structures where a significant majority of pixels are brighter or darker than a significant majority of pixels in their corresponding surroundings. Compared to linear filter based detectors, e.g. SIFT detector, the proposed rank order filter based detector is more robust to abrupt variations of images caused by illumination and geometric changes. Experiments on the benchmark databases demonstrate that the proposed ROLG detector achieves superior performance comparing to four state-of-the-art detectors. Evaluation experiments are also conducted on face recognition problems. The results on five face databases further demonstrate that the ROLG detector significantly outperforms the other detectors.

1. Introduction

As a powerful tool for computer vision, interest point detection has drawn great attentions in the last two decades [1–3]. It has been used in a wide range of research, such as panoramic image stitching [4], image retrieval [5], image registration [6], texture classification [7], object categorization [8], object recognition [9,10], 3D object modeling [11], video shot retrieval [12], and face recognition [13]. Many interest point detectors have been proposed in the past few years to detect local structures of images [14–26]. They can be roughly classified into three categories: corner-based detectors, blob-based detectors and region detectors.

Corners correspond to points in the 2D images with high curvature [3]. Harris corner detector [14] uses the second moment matrix, also called the auto-correlation matrix, to analyze the principal intensity changes in two orthogonal directions in a neighborhood around a point. The Harris measure combines the trace and the determinant of the second moment matrix in a single measure. This measure is used to detect the image local structures where the principal intensity changes in two orthogonal directions are both large. However, this type of structures includes not only corners, but also textured patterns and noise [27]. Harris–Laplace/affine detectors [15] were proposed to be invariant with scale and affine changes. Corners are detected by the Harris corner detector in multi-scales, and their characteristic scales are determined by the Laplacian operator. As the shape of a corner does not match the shape of the Laplacian operator, scale estimations for corners are often unstable [28]. SUSAN detector [16] defines a corner as the smallest USAN (univalue segment assimilating nucleus) point, which is dissimilar from a majority of pixels within a neighborhood of it. This detector is sensitive to impulse noise and blur, and it fails to deal with scale changes.

Blobs refer to bright regions on dark backgrounds or vice versa [29]. Hessian detector [18] employs the Hessian matrix to analyze the second order Taylor expansion of the intensity surface. The Hessian matrix consists of the second order derivatives of image intensity. The trace and the determinant of this matrix are used to detect blobs in a single scale. Hessian–Laplace/affine detectors [15] were developed to detect blobs in multiple scales based on the Hessian detector and the Laplacian operator. These detectors are stable in estimating the characteristic scales of blobs, of which the shapes are similar to that of the Laplacian operator. SIFT detector [19] employs the difference of Gaussian (DoG) filter to approximate the normalized Laplacian of Gaussian (LoG) filter. The DoG filter significantly accelerates the computation process. SURF detector [20] employs the box filters and the integral images to further speed up the Hessian–Laplace detector. The box filters are approximations of the second order Gaussian derivative filters. The integral images allow for the fast convolutions of the box filters with the input image. Different approaches, which are not based on the second order derivative of image intensity, were proposed in [21–23] to detect blobs. Salient region detector [21] employs the image local complexity to detect blobs. The characteristic scales are determined by the entropy extrema of the local descriptors. A common computational concept is proposed in [22] to detect different types of local structures. The intensity variance in a local circular region is divided into three components, which are used to

* Corresponding author. Tel.: +65 9220 7858.
E-mail addresses: mi0001ei@e.ntu.edu.sg, mi0001ei@ntu.edu.sg (Z. Miao), exdjiang@ntu.edu.sg (X. Jiang).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.03.024

detect corners, blobs and high textured structures. A histogram-based similarity measure is introduced in [23] to bridge the gap between interest point detectors and descriptors.

Region detectors extract regions with similar image structures and properties [3]. Edge-based region detector [24] uses the Harris points as the initial points. The two nearby edges and several intensity based functions are exploited to determine a parallelogram region. Intensity-based region detector [24] employs the local intensity extrema as the initial points. The affine invariant regions are determined by the significant changes of the intensity profiles, which are along rays going out of the extrema. MSER detector [25] starts with the local intensity extrema. The maximally stable extremal regions are extracted using a watershed like segmentation algorithm. The MSER detector works well for structured images, which have strong intensity changes on region boundaries. But it is sensitive to image blur which undermines the stability criterion [1].

In this paper, we propose a new interest point detector based on a nonlinear filter. The proposed detector is inspired by the SIFT detector [19], and alleviates the problems of this detector. As the most popular detector, the SIFT detector employs the DoG filter to generate the blob map. However, the response of the DoG filter is easily affected by the strong and abrupt structures near the structure to be detected. This makes the SIFT detector unstable to detect the low contrast image structures, where other strong and abrupt structures caused by illumination or geometric changes partially fall in the detection window. Moreover, due to the second order derivative nature of the SIFT detector, many unstable and spurious points are often detected around the structures. To solve these problems of the SIFT detector, we design our detector based on the rank order filter.

Most state-of-the-art rank order filters, such as the weighted rank order filter and the weighted median filter, are designed to remove noise or detect edges [30–32]. Very few studies have been carried out on applying rank order filters to interest point detection. Paler et al. [33] developed a corner detector based on the median filter. The difference between the median filtered image I_m and the original image I, $I_d = I_m - I$, is used as the corner map. As it functions as a high pass filter, this detector is sensitive to noise. Ren and Jiang [34] employed a rank order filter pair to detect human eyes. The difference of the outputs of the two rank order filters, one for the eyeball and the other for the surrounding pixels, is computed. The local maxima are used as the eye candidates. This eye detector is effective in dealing with iris reflections and other dark objects near eyeballs. However, it cannot be used as an interest point detector because it has the following three limitations. Firstly, it can only detect dark regions. Secondly, it has high response on edges and contours. Local extrema may be detected along edges and contours. These extrema are unstable because they are sensitive to small intensity changes in their neighborhoods. Thirdly, it is not invariant to scale changes.

In this work, a novel rank order filter with weights proportional to the coefficients of the LoG filter is proposed and, hence, is named rank order Laplacian of Gaussian (ROLG) filter. It is used to detect the image local structures where a majority of pixels are brighter or darker than a majority of pixels in their corresponding surroundings. The new interest point detector is built on the proposed ROLG filter to detect image local structures in multiple scales.

2. The proposed ROLG filter

As a necessary preliminary of the study, the properties of the LoG filter and the weighted rank order filter are discussed in Sections 2.1 and 2.2.

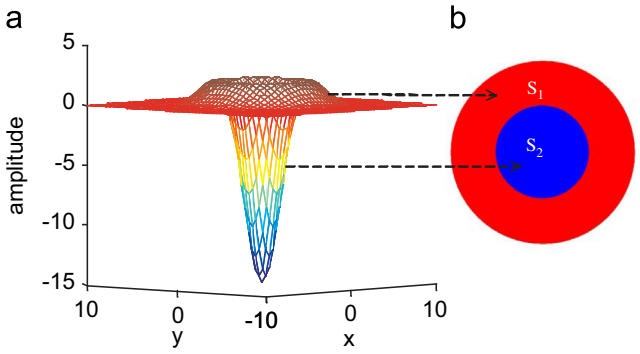

Fig. 1. LoG filter. (a) Shape of the LoG filter. (b) Two parts of the LoG filter. S_1 corresponds to the surrounding ring containing positive weights. S_2 corresponds to the inner circular disk containing all negative weights.

2.1. LoG filter

The LoG filter with the shape shown in Fig. 1(a) is defined by

$$w(x,y,\sigma) = -\frac{1}{\pi\sigma^4}\left[1-\frac{x^2+y^2}{2\sigma^2}\right]e^{-(x^2+y^2)/2\sigma^2}, \tag{1}$$

where σ is the standard deviation of the Gaussian function, and also named scale factor. For input image $I(x,y)$, the output of the LoG filter at (x,y) can be expressed as the difference of two weighted averages

$$\begin{aligned} r(x,y,\sigma) &= \sum_{(m,n)\in S} w(m,n,\sigma)I(x-m,y-n) \\ &= \sum_{(m,n)\in S_1} w_+(m,n,\sigma)I(x-m,y-n) \\ &\quad - \sum_{(m,n)\in S_2} w_-(m,n,\sigma)I(x-m,y-n), \end{aligned} \tag{2}$$

where S is the region of the filter, S_1 and S_2 (as shown in Fig. 1(b)) are the two parts of S with $S_1 \cup S_2 = S$ and $S_1 \cap S_2 = \phi$ (ϕ is the null set). S_1 corresponds to the surrounding ring containing the positive weights of the filter, and S_2 corresponds to its inner disk containing all the negative weights. w_+ and w_- are the absolute values of the LoG coefficients in S_1 and S_2, respectively.

It is not difficult to see that the LoG filter produces extrema at corners and blobs. Several detectors are built on the LoG filter [15,18–20]. The SIFT detector [19], which is the most famous one, employs the DoG filter to approximate the normalized LoG filter that significantly accelerates the computation process.

The LoG filter is ineffective in dealing with the sparse but strong noise, such as the salt and pepper noise. Even a small portion of pixels can greatly affect the output adversely if their grey values largely deviate from those of the image structure to be detected. However, a small portion of pixels have almost no influence on the output of the rank order filter even if their grey values are extremely high or low. This motivates us to design a weighted rank order filter with weights proportional to those of the LoG filter for interest point detection.

2.2. Weighted rank order filter

The output of the weighted rank order filter [30,31] is defined as follows. Assume the weights for the input series $\mathbf{x} = \{x_1, x_2, \ldots, x_q\}$ are $\mathbf{w} = \{w_1, w_2, \ldots, w_q\}$. For the ascending sorted $\tilde{\mathbf{x}} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_q\}$, their corresponding weights are rearranged as $\tilde{\mathbf{w}} = \{\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_q\}$. The output of the weighted rank order filter with rank r_w, $r_w \in \{1, 2, \ldots, \sum_{j=1}^{j=q} w_j\}$, is given by

$$\begin{aligned} y_{r_w} &= rank_{r_w}\{w_1 \diamond x_1, w_2 \diamond x_2, \ldots, w_q \diamond x_q\} \\ &= rank_{r_w}\{\tilde{w}_1 \diamond \tilde{x}_1, \tilde{w}_2 \diamond \tilde{x}_2, \ldots, \tilde{w}_q \diamond \tilde{x}_q\}, \end{aligned} \tag{3}$$

where $\diamond$ is the replication operator defined by

$$w_i \diamond x = \underbrace{x, x, \ldots, x}_{w_i \text{ times}}. \tag{4}$$

Take the input series $\mathbf{x}=\{10,8,9\}$, the weights $\mathbf{w}=\{2,3,1\}$, and the rank $r_w=4$ as an example. The output of the weighted rank order filter is

$$\begin{aligned} y_4 &= rank_4\{2\diamond 10, 3\diamond 8, 1\diamond 9\} \\ &= rank_4\{3\diamond 8, 1\diamond 9, 2\diamond 10\} \\ &= rank_4\{8,8,8,9,10,10\} = 9, \end{aligned}$$

which is the 4th element of the expanded data.

In order to avoid replicating the data, which is time consuming and needs more storage spaces, a cumulative sum of the sorted weights is defined by

$$c_i = \frac{1}{w_s}\sum_{j=1}^{i} \tilde{w}_j, \tag{5}$$

where $w_s = \sum_{j=1}^{q} \tilde{w}_j$ is the total sum of the weights, $i \in \{1,2,\ldots,q\}$ and $c_0 = 0$. Then, the output of the weighted rank order filter with a normalized rank $r_{nw} \in [0 \quad 1]$ is given by

$$y_{r_{nw}} = \tilde{x}_{i_o}, \{i_o : c_{i_o-1} < r_{nw} \le c_{i_o}\}. \tag{6}$$

2.3. The proposed ROLG filter

One direct way to apply the weighted rank order filter to interest point detection is to replace the weighted average in (2) by the weighted median, as

$$\begin{aligned} r_{wm}(x,y,\sigma) = &\underset{(m,n)\in S_1}{median}(\hat{w}_+(m,n,\sigma)\diamond I(x-m,y-n)) \\ &-\underset{(m,n)\in S_2}{median}(\hat{w}_-(m,n,\sigma)\diamond I(x-m,y-n)), \end{aligned} \tag{7}$$

where $\hat{w}_+(m,n,\sigma) = w_+(m,n,\sigma)/\sum w_+$ and $\hat{w}_-(m,n,\sigma) = w_-(m,n,\sigma)/\sum w_-$. With these weighting coefficients, pixels near the boundary between S_1 and S_2, which are uncertain to be grouped to S_1 or S_2, are assigned with small weights to weaken their influence on the filter output. The difference of the two weighted median filters (7) has similar role to the LoG filter (2) and, hence, can be used to detect interest points.

However, when noise exists, filter (7) produces very strong response on an edge if one median filter captures one side of the edge while the other median filter happens to capture the other side of the edge. Such strong response results in many local extrema being detected along edges, which are undesirable for interest point detection. Therefore, additional rules are imposed to enhance the robustness of the detector. Median filter is a special case of the rank order filter, as median is equal to rank 0.5. Replacing the weighted median filter by the weighted rank order filter, (7) is reformulated as

$$\begin{aligned} r_{wr}(x,y,\sigma,\lambda_1,\lambda_2) = &\underset{(m,n)\in S_1}{rank_{\lambda_1}}(\hat{w}_+(m,n,\sigma)\diamond I(x-m,y-n)) \\ &-\underset{(m,n)\in S_2}{rank_{\lambda_2}}(\hat{w}_-(m,n,\sigma)\diamond I(x-m,y-n)), \end{aligned} \tag{8}$$

where λ_1 and λ_2 are the rank factors for the two weighted rank order filters.

In order to suppress the edge response, we require a significant majority of pixels ($>50\%$, e.g. 60%) in the surrounding ring brighter than a significant majority of pixels ($>50\%$, e.g. 60%) in the inner disk, or a significant majority of pixels in the surrounding ring darker than a significant majority of pixels in the inner disk. Otherwise, the outputs are set to zero to suppress noise and edges. This idea can be realized by introducing a positive nonzero offset parameter δ and two functions, $\mathcal{P}$ and $\mathcal{N}$, as

$$\mathcal{P}(x,y,\sigma,\delta) = r_{wr}(x,y,\sigma,0.5-\delta,0.5+\delta), \tag{9}$$

and

$$\mathcal{N}(x,y,\sigma,\delta) = r_{wr}(x,y,\sigma,0.5+\delta,0.5-\delta). \tag{10}$$

$\mathcal{P}(x,y,\sigma,\delta) > 0$ ($\mathcal{N}(x,y,\sigma,\delta) < 0$) is used to check that a significant majority of pixels in the surrounding ring is brighter (darker) than a significant majority of pixels in the inner circle at point (x,y). With this idea, the proposed ROLG filter is defined by

$$r_{ROLG}(x,y,\sigma,\delta) = \begin{cases} \mathcal{P}(x,y,\sigma,\delta), & \text{if } \mathcal{P}(x,y,\sigma,\delta) > 0 \\ \mathcal{N}(x,y,\sigma,\delta), & \text{if } \mathcal{N}(x,y,\sigma,\delta) < 0 \\ 0. & \text{otherwise} \end{cases} \tag{11}$$

3. Analyses of the ROLG filter

In this section, illustrative analyses of the ROLG response on blobs, corners and edges are presented.

3.1. Responses of the ROLG filter on blobs

One drawback of the LoG filter in detecting blobs is that spurious local extrema are produced around the blobs. Examples are shown in Figs. 2(b) and 3(b). For a 1D blob, one negative peak at the center and two positive peaks on both sides of the center are generated by the LoG filter. For a 2D blob, besides a peak at the center of the blob, a ring (indicated by the red circle shown in Fig. 3(b)) is produced around it. Many unstable extrema may be detected on this ring in the presence of even very small noise.

Figs. 2(c) and 3(c) show the responses of the proposed ROLG filter on a 1D blob and a 2D blob, respectively. It is clear to see that the ROLG filter produces a peak at the center of the blob, and does not generate any peaks around the blob. This is explained by the filtering process on a 1D blob illustrated in Fig. 4. When the 1D-ROLG-filter mask is on one side of the blob, input values within the mask are monotonically increasing or decreasing. Take mask 1 in Fig. 4 as an example. Within this mask, input values are monotonically increasing. Compared to its inner part, half of its surrounding parts (left part) is darker, and the other half (right part) is brighter. The ROLG response is set to 0, because it does not satisfy that a significant majority of pixels in the surrounding parts

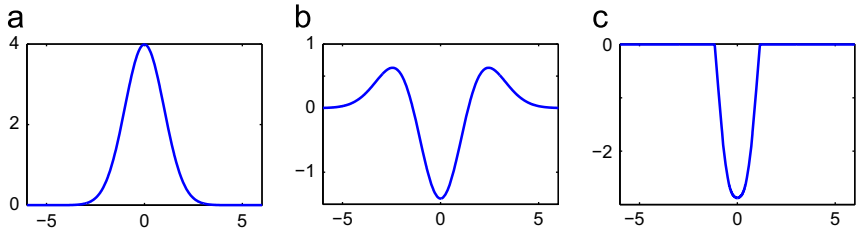

Fig. 2. Response on a 1D blob. (a) A 1D blob. (b) Response of the LoG filter. (c) Response of the ROLG filter.

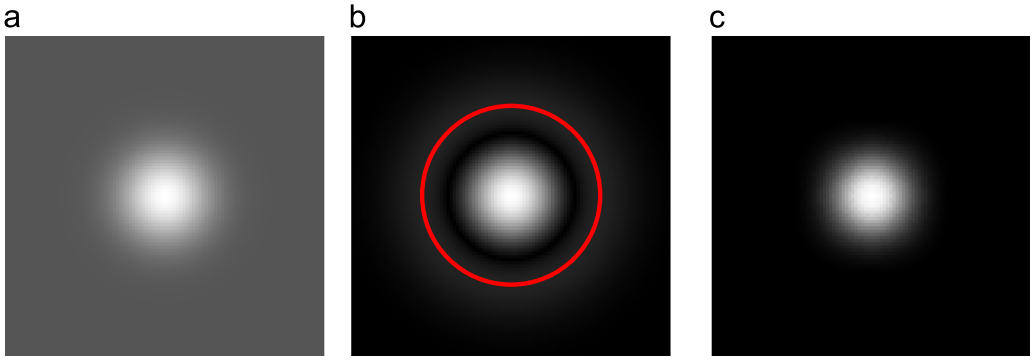

Fig. 3. Response on a 2D blob. (a) A 2D blob. (b) Absolute value of the LoG response. (c) Absolute value of the ROLG response. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

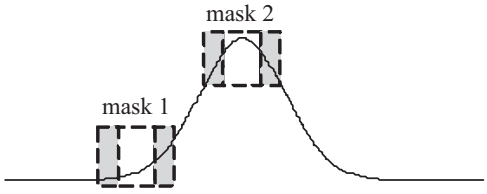

Fig. 4. A 1D blob and two 1D-ROLG-filter masks on the blob. In each mask, the white region corresponds to the inner part of the ROLG filer mask, and the gray region corresponds to its surrounding parts.

are brighter or darker than a significant majority of pixels in the inner part. When the ROLG mask is at the center of the blob, e.g. mask 2 in Fig. 4, a major part of the inner region is brighter than a major part of its surrounding regions. Therefore, a peak is generated by the ROLG filter at the center of the blob. Similar to the 1D case, the ROLG filter produces a peak at the center of the 2D blob, and does not generate a ring around the blob. This avoids detecting spurious points around the 2D blob.

3.2. The ROLG responses on edges and corners

Another problem of the LoG filter in detecting interest points is that local extrema are often detected along edges. For a 1D edge, two peaks near the edge are produced by the LoG filter as shown in Fig. 5(b). For a 2D corner, besides a peak on the corner, strong responses are generated along the edges as shown in Fig. 6(b). Local extrema may be detected along edges.

The ROLG responses on a 1D edge and a 2D corner are shown in Figs. 5(c) and 6(c), respectively. These two figures demonstrate that the ROLG filter suppresses the responses of edges. As shown in Fig. 7, when the 1D-ROLG-filter mask is on the edge, values within the mask are monotonically increasing. Half of the surrounding parts is brighter than the inner part, and the other half is darker. No majority of pixels in the inner part are brighter or darker than a majority of pixels in the surrounding parts. Thus, the responses of the ROLG filter on edges are set to 0.

4. Interest point detection by the ROLG filter

In this section, we propose the ROLG interest point detector based on the ROLG filter. The advantages of the ROLG filter in detecting interest points in a single scale are discussed in Section 4.1. The algorithm to eliminate ridge responses is presented in Section 4.2. The proposed ROLG detector to detect interest points in multiple scales is given in Section 4.3.

4.1. Interest point detection in a single scale

Previous sections have shown that the ROLG filter has the following advantages in detecting interest points:

1. The sparse but strong structures has small or no influence on the output of the ROLG filter. Thus, the ROLG filter is robust to the strong and abrupt variations of images.
2. Structures, which partially fall in a detection window, have limited influence on the response of the ROLG filter. Hence, the mutual influence of structures on the response of the ROLG filter is limited.
3. Only one peak is produced at the center of a blob, and no ring is generated around the blob. This property of the ROLG filter avoids detecting spurious points around a blob.
4. The ROLG filter suppresses the response of edges. Therefore, no point is detected along edges.

Fig. 8 shows some image structures and the detected points based on the LoG filter and the proposed ROLG filter. The input image of Fig. 8(a) contains a blob and a small black stripe. The abrupt structure drastically changes the LoG response to the blob. The peak of the LoG response deviates from the true position of the blob. Many false peaks are detected around the blob. The response of the ROLG filter shows that the abrupt structure has small impact on the output of the ROLG filter. The blob is correctly detected by the ROLG filter. Two close blobs are contained in the input image of Fig. 8(b). The mutual influence of these two blobs results in many spurious points being detected on the LoG response. However, their mutual influence on the ROLG response does not generate spurious peaks. Thus, these two blobs are detected correctly. Fig. 8(c) compares the response of the ROLG filter to that of the LoG filter on a blob. It is clear to see that many false peaks are detected on the LoG response around the blob, while no false peak is detected on the response of the ROLG filter. Fig. 8(d) shows that many local intensity extrema are detected along the edges on the LoG response. In contrast, the response of the ROLG filter on the edges is 0 and, hence, no false peak is detected on the edges.

4.2. Eliminating ridge responses

The response of the ROLG filter on a ridge is strong if its scale is close to the width of the ridge. Points detected on the ridge are unstable to small amounts of noise. We employ the algorithm given in [19] to remove such kind of unstable points.

Points on the ridge have a small principal curvature along the ridge but a large one in the perpendicular direction. The two principal curvatures at the location and scale of the interest point

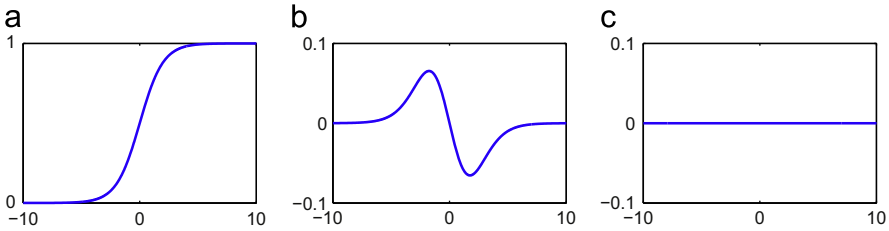

Fig. 5. Response on a 1D edge. (a) A 1D edge. (b) Response of the LoG filter. (c) Response of the ROLG filter.

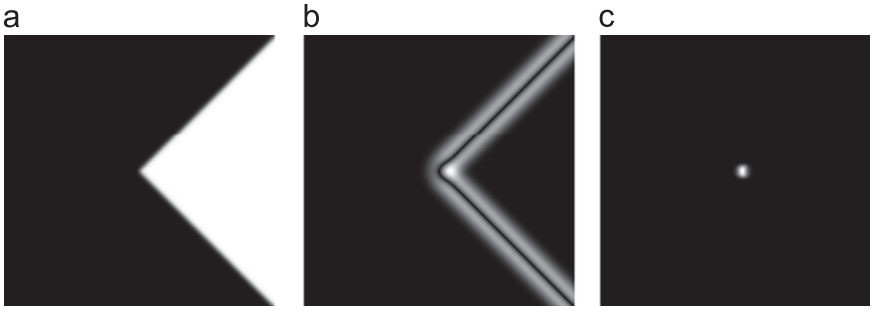

Fig. 6. Response on a corner. (a) A 2D corner. (b) Absolute value of the LoG response. (c) Absolute value of the ROLG response.

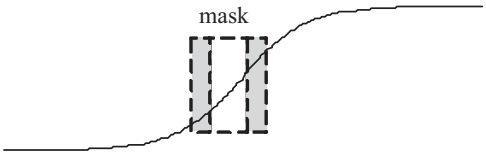

Fig. 7. A 1D edge and a 1D-ROLG-filter mask on the edge. In this mask, the white region corresponds to the inner part of the ROLG filter mask, and the gray region corresponds to its surrounding parts.

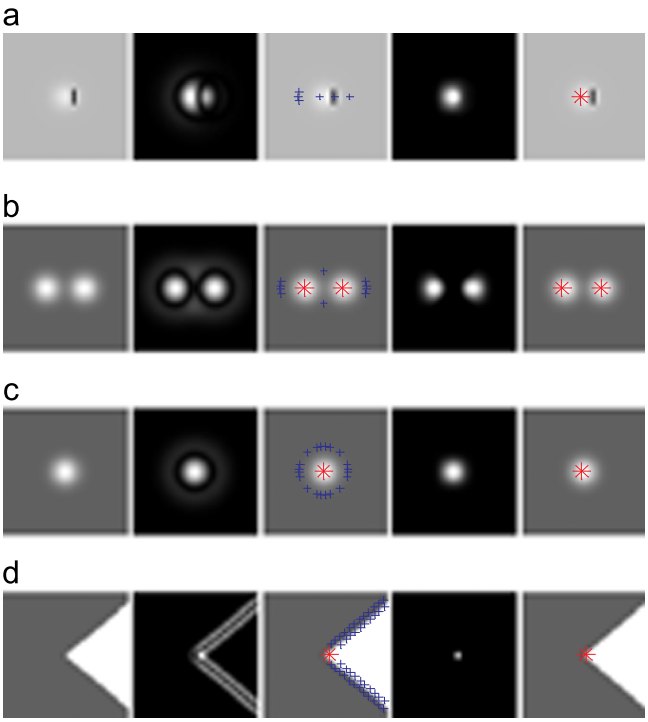

Fig. 8. Points detected on a single scale. '*' denotes desired point and '+' denotes undesired point. From left to right of (a), (b), (c) and (d) are input images, absolute value of the LoG responses, local extrema of the LoG responses, absolute value of the ROLG responses, and local extrema of the ROLG responses, respectively.

can be computed from the 2×2 Hessian matrix $\mathbf{H}$

$$\mathbf{H} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}, \tag{12}$$

where I_{xx}, I_{xy}, and I_{yy} are the second order derivatives. The derivatives are estimated by the differences between neighboring sample points.

The principal curvatures of I are proportional to the eigenvalues of $\mathbf{H}$. Therefore, the ratio between the larger eigenvalue and the smaller eigenvalue of $\mathbf{H}$ can be used to remove the points on the ridge. If the ratio is larger than some threshold r, it means that the principal curvatures in one direction is larger than r times of that in the perpendicular direction. In order to avoid explicitly computing the eigenvalues, the function given in [19]

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} < \frac{(r+1)^2}{r} \tag{13}$$

is used to check that the ratio of the eigenvalues of $\mathbf{H}$ is below r. The experiments in this paper use $r=10$, as suggested in [19].

4.3. Algorithm for ROLG detector

Interest point detection in multiple scales is an important issue in vision applications. Similar to the Hessian–Laplace detector [35], we can weight the responses in different scales, and choose the local maximum points both in spatial and scale dimensions as the interest points. In the implementation, we employ a straightforward method by detecting the interest points in each scale as done in [23].

The proposed algorithm for the ROLG detector is summarized below:

1. Initialize the ROLG filter by setting the offset parameter δ and the scale parameter σ.
2. Generate the corner/blob map by filtering the input image with the ROLG filter (11).
3. Detect peaks on the corner/blob map, and remove peaks which are on ridges. Remaining peaks are the interest points in this scale.

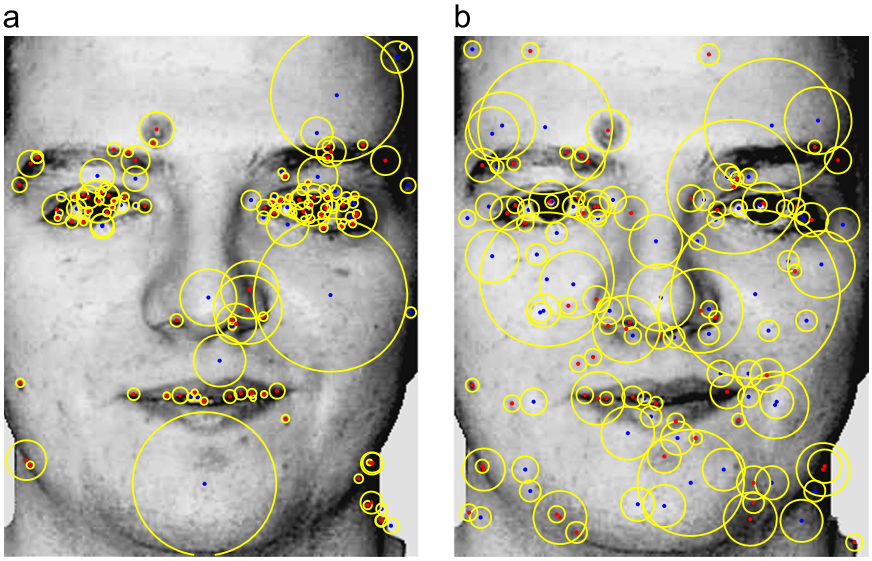

Fig. 9. Interest points detected by two detectors. (a) SIFT detector, (b) ROLG detector. Local maxima are indicated in blue, and local minima are indicated in red. The radius of the yellow circle is two times the scale σ of the interest point at its center. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Update the ROLG filter by a larger scale σ, and go back to step 2 to detect interest points in a new scale until the maximal scale is reached.

Fig. 9 gives a visual comparison between the SIFT detector and the ROLG detector on a face image. The parameters are chosen so that the numbers of points detected by both the detectors are the same. It is observed that SIFT detects many spurious points around the eyeballs while it misses a lot of points in other areas. The ROLG detector alleviates these problems. In the next section, statistical experiments are carried out for further comparison between the ROLG detector and other state-of-the-art ones.

5. Experiments

Experiments in Section 5.1 test the repeatability and the discrimination of the interest points. Experiments in Section 5.2 test the performance of the interest points in the application of face recognition. In both experiments, the SIFT descriptor [19] is used to represent the detected regions.

The proposed ROLG filter has an offset parameter δ, $0 \le \delta \le 0.5$. The Oxford database [1] is used to evaluate the sensitivity of the offset parameter δ with respect to the detection result. Both the absolute and relative measures given in Section 5.1 are employed. The experimental results are shown in Fig. 10. It is not a surprise that both the number of the repeated points and the number of the matched points decrease with the increase of the offset parameter δ, as shown in Fig. 10(a). However, as shown in Fig. 10 (b), the repeatability and the matching score increase first and then decrease. Both of them reach and maintain the maximum within the range of $0.05 \le \delta \le 0.15$. This clearly shows the noise- and edge suppression function of a nonzero offset parameter δ in the proposed ROLG filter. Therefore, we choose $\delta = 0.1$ for the proposed ROLG filter in all experiments of this paper.

5.1. Repeatability and discrimination tests

The goal here is to evaluate the ROLG detector under different image variations. The evaluation is based on the protocols suggested in [1]. Detectors are evaluated by both the absolute and relative measures. The absolute measures include the number of repeated interest points and the number of matched points. Each interest point corresponds to a detected region. Two regions are repeated if their overlap is the maximal and above some threshold (in our experiments, the threshold is 60%). Two regions are matched if they satisfy two conditions: (1) the two regions are repeated and (2) their descriptors are the nearest-neighbor in the descriptor space.

The relative measures include the repeatability and matching score. Note that the definitions of the repeatability and matching score in this paper are slightly different from that in [1]. In [1], the repeatability for a given pair of images is computed as the ratio between the number of the repeated points and the smaller number of the detected points in the pair of images. This may cause some inaccuracy problems in some cases. For example, assume 100 points are detected in image 1. Due to some variations, such as illumination change, assume 50 points are detected in image 2. If the 50 points are repeated, by the definition of repeatability in [1], the repeatability is 100%. This result is undesirable because 50 points are not detected in image 2. Therefore, in this paper we define the repeatability as the ratio between the number of the repeated points and the larger number of the detected points in a given image pair. Similarly, we also define the matching score as the ratio between the number of correct matches and the larger number of the detected points in a given image pair. All these points should be within the common area and the common scales of the image pair.

The publicly available Oxford database provided by [1] is used to evaluate the detectors. This database contains eight data sets, from which images are shown in Fig. 11. These data sets include five different changes in imaging conditions for structured and textured scenes: scale change, viewpoint change, JPEG compression, image blur, and lighting change. Each data set consists of six images with five homographies between the first image and the other five images. In all experiments reported here, interest points are detected on the downsampled images.

The parameter setting for the ROLG detector is as follow. Interest points are detected in 12 scales: $\{\sigma_n\}_{n=1,2,\ldots,12} = \{1.6 \times 2^{1/3}, 1.6 \times 2^{2/3}, 3.2, \ldots, 1.6 \times 2^4\}$. Instead of continuously increasing the ROLG mask size, the 12 scales are divided into four octaves by downsampling the previous octave. Each octave contains three scales $\{\sigma_{no}\}_{no=1,2,3} = \{1.6 \times 2^{1/3}, 1.6 \times 2^{2/3}, 3.2\}$.

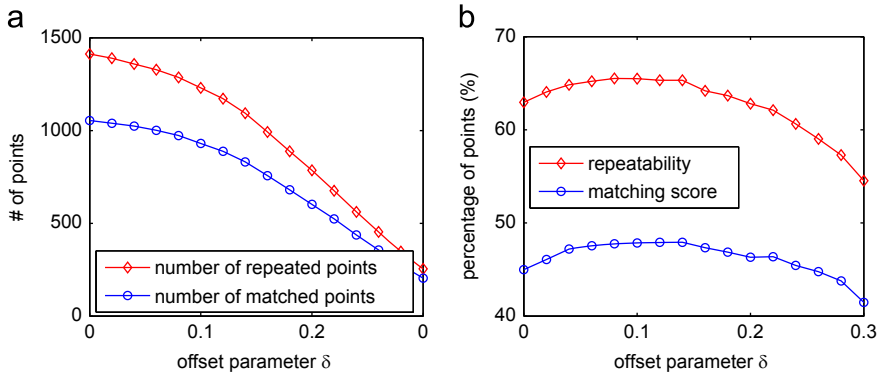

Fig. 10. Sensitivity of the offset parameter δ with respect to the detection results. (a) The average number of repeated points (top line) and matched points (lower line). (b) The average value of repeatability (top line) and matching score (lower line).

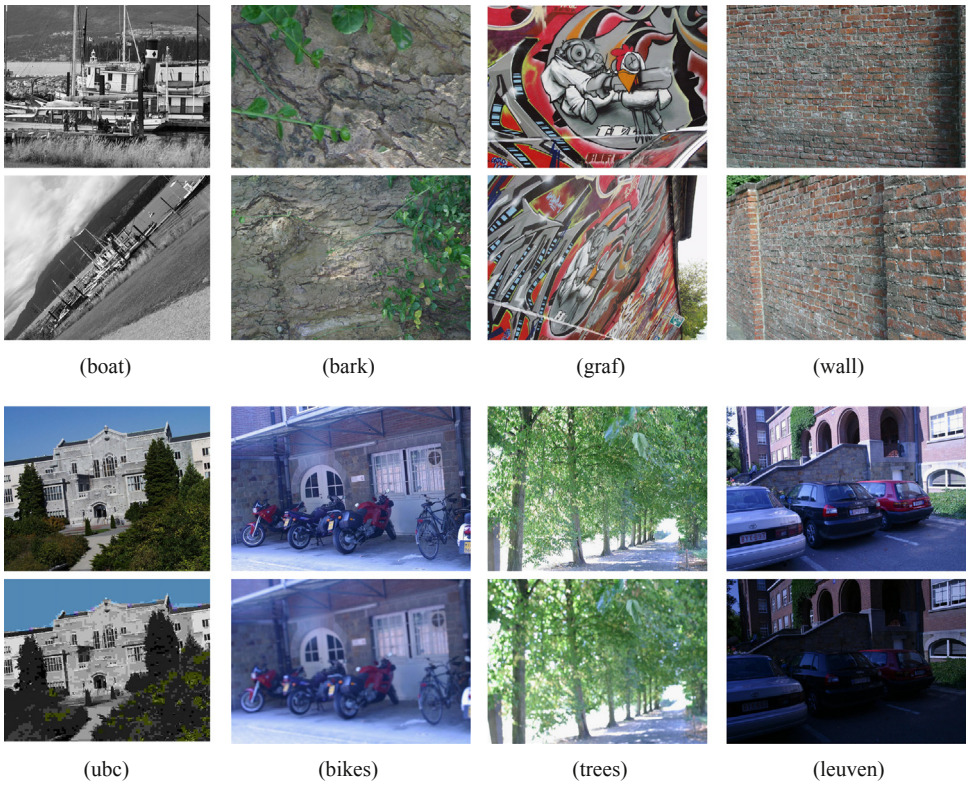

Fig. 11. Samples from the eight data sets of the Oxford database. 'boat' (of size 424×339) and 'bark' (382×255): scale and rotation change. 'graf' (399×319) and 'wall' (499×349): viewpoint change. 'ubc' (399×319): JPEG compression. 'bikes' (499×349) and 'trees'(499×349): image blur. 'leuven' (449×299): lighting change. Two (the 1st and the 6th) of the six images are shown for each data set. The top image in each set is used as the reference image in the experiments.

Four benchmark detectors, the MSER detector [25], the Harris-affine (HR-A) detector [15], the Hessian-affine (HS-A) detector [15], and the SIFT detector [19], are compared with the ROLG detector. The default parameters given by the authors are used for each detector.

The results of the absolute measures and relative measures are shown in Figs. 12 and 13, respectively. Each figure in Figs. 12 and 13 contains eight columns, corresponding to the eight data sets. In each column, horizontal axis represents the image index in the corresponding data set. From left to right of each figure shows the results on the scale change structured sequence 'boat', the scale change textured sequence 'bark', the viewpoint change structured sequence 'graf', the viewpoint change textured sequence 'wall', the JPEG compression sequence 'ubc', the blurring structured sequence 'bikes', the blurring textured sequence 'trees', and illumination change sequence 'leuven', respectively.

Results for the scale change and in-plane rotation are shown in the 1st and 2nd columns of Figs. 12 and 13. On both the structured scene (Fig. 11 boat) and the textured scene (Fig. 11 bark), the ROLG detector gives the best results.

Results for the viewpoint change are shown in the 3rd and 4th columns of Figs. 12 and 13. When the viewpoint change is small, the ROLG detector obtains higher repeatability score and larger numbers of repeated points and matched points than those of the other detectors on the structured scene (Fig. 11 graf). As the mask of the ROLG filter is not adapted to viewpoint change, the performance of the ROLG filter drops faster than that of the MSER detector under the viewpoint change increasing. The most stable one for the structured scene is the MSER detector, but its number of repeated points is small. On the textured scene (Fig. 11 wall), the ROLG detector outperforms the other detectors.

The 5th columns of Figs. 12 and 13 show the results for the JPEG compression sequence (Fig. 11 ubc). The Hessian-affine detector and the Harris-affine detector show best performance, but the number of repeated points is small. When the distortion under JPEG compression

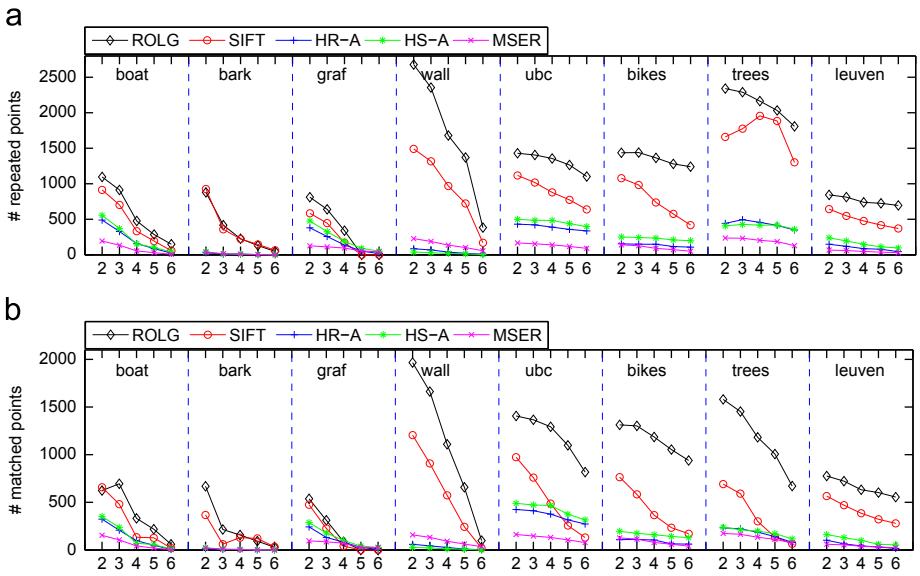

Fig. 12. Number of repeated interest points (a) and number of correct matched interest points (b) on the Oxford database. Each figure contains eight columns corresponding to the eight data sets. In each column, horizontal axis represents the image index in the corresponding data set. From left to right of (a) and (b) are the results on the image sequence of 'boat', 'bark', 'graf', 'wall', 'ubc', 'bikes', 'trees' and 'leuven', respectively.

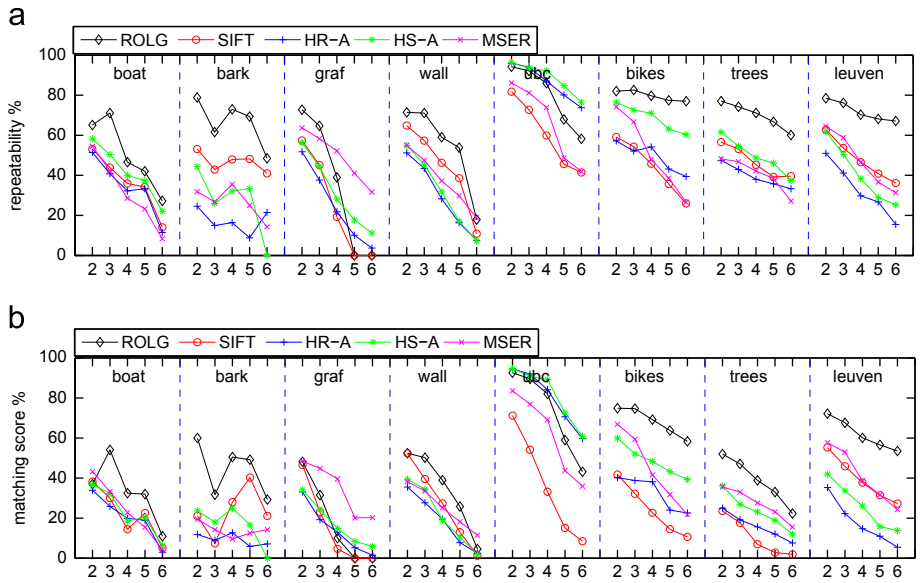

Fig. 13. Repeatability score (a) and matching score (b) on the Oxford database. Each figure contains eight columns corresponding to the eight data sets. In each column, horizontal axis represents the image index in the corresponding data set. From left to right of (a) and (b) are the results on the image sequence of 'boat', 'bark', 'graf', 'wall', 'ubc', 'bikes', 'trees' and 'leuven', respectively.

is low, the repeatability score and the matching score of the ROLG detector are as good as those of the Hessian-affine detector. Moreover, both the number of repeated points and the number of matched points of the ROLG detector are larger than those of the other detectors.

The 6th and 7th columns of Figs. 12 and 13 show the results for blur images. The ROLG detector outperforms the other detectors on both the structured scene (Fig. 11 bikes) and the textured scene (Fig. 11 trees).

Results for the illumination change (Fig. 11 leuven) are shown in the 8th column of Figs. 12 and 13. As the ROLG filter is robust to the variations caused by illumination changes, the ROLG detector outperforms the other detectors.

We compare the running time of the SIFT and ROLG detectors under the Window 7 system with the Intel Core i5 CPU 3.2 GHz and RAM 4 GB. The code of the SIFT detector is downloaded from http://www.vlfeat.org/~vedaldi/assets/sift/versions/sift-0.9.16.tar.gz. The ROLG detector is implemented by the MATLAB programming language. Both the image size and the number of detected interest points affect the running time. From Fig. 12(a) it is seen that the ROLG and SIFT detectors detect the similar number of repeated points on the 'bark' data set. Therefore, we test the running time of these two detectors on this data set. The average running time per image of size 382×255 for the ROLG detector is 4.1 s and that for the SIFT detector is 1.2 s. It is not a surprise that the SIFT detector is faster than the ROLG detector as the former optimizes its speed by utilizing the nice properties of the Gaussian linear filter.

5.2. Application to face recognition

Face recognition is an active research topic [36–40], and some works have been done to apply the SIFT detector and descriptor in

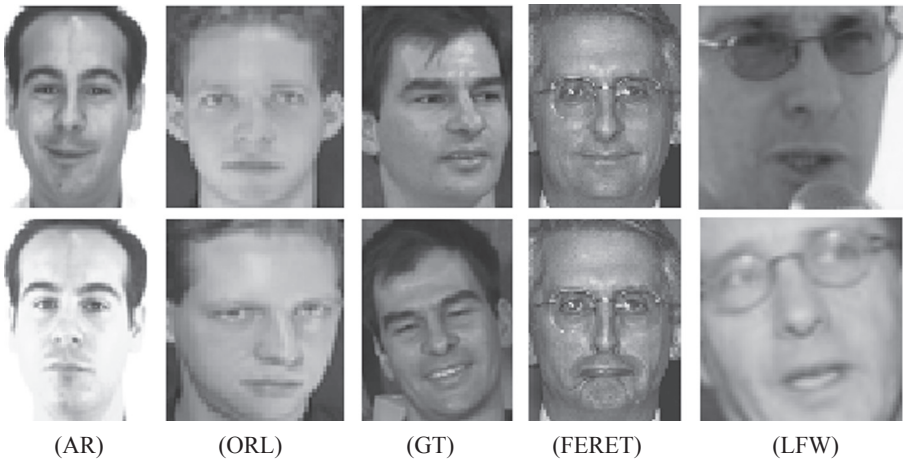

Fig. 14. Sample images in AR, ORL, GT, FERET and LFW databases. They show the typical image variations of the same persons in each database.

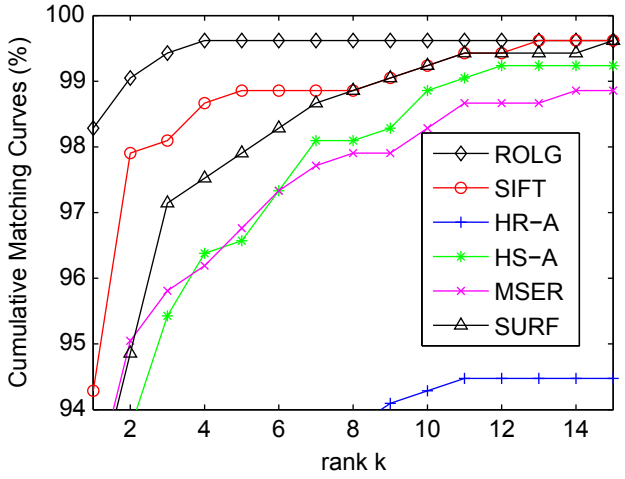

Fig. 15. Cumulative matching curves on AR database.

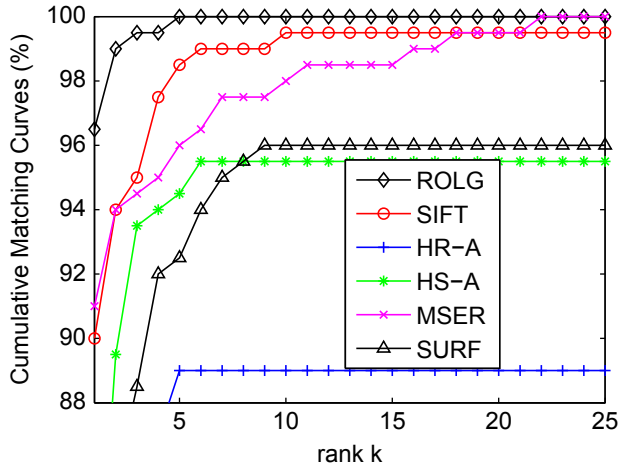

Fig. 16. Cumulative matching curves on ORL database.

face recognition [13]. In the following experiments, we compare the ROLG detector with five state-of-the-art detectors, the SIFT detector [19], the MSER detector [25], the Harris-affine (HR-A) detector [15], the Hessian-affine (HS-A) detector [15], and the SURF detector [20]. Using the default parameters given by their respective authors, all of these detectors detect too few points and lead to very poor performance. Thus, we decrease the contrast threshold and find that zero is the best for all detectors. Therefore, the thresholds used to remove low contrast interest points are set to zero for all detectors. The MSER detector is controlled by several parameters. Even if we set the contrast threshold to the smallest zero, there is no point detected on the ORL database. In order to make MSER detector workable for the face recognition experiments, we reduce the minimum size of output region of the MSER detector. The minimum size of output region is set to 1/4 of its default setting to make it workable on all face databases as its recognition performance is better than those with 1/2 and 1/8 on the ORL database. This optimal parameter setting is further confirmed on the AR database. The matching procedures described in [19] are employed in these experiments.

AR [41], ORL [42], Georgia Tech (GT) [43], FERET [44], and labeled faces in the wild (LFW) [45] databases are chosen to test the discriminative power of the interest points in face recognition. Some sample images of these databases are shown in Fig. 14. Before the interest point detection, images are resized to those commonly used in most other face recognition approaches. The rank 1 recognition rates and the cumulative matching curves are used to evaluate the detectors. The cumulative matching curve of the Harris-affine detector is not drawn in Fig. 17, Figs. 18 and 19, because it is drastically lower than that of the other detectors in these three figures.

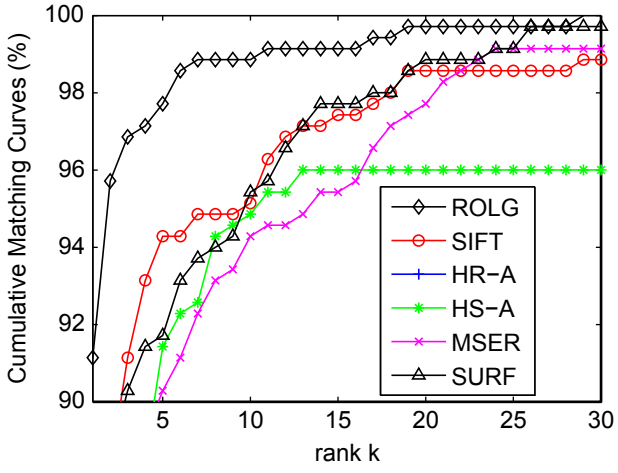

Fig. 17. Cumulative matching curves on GT database.

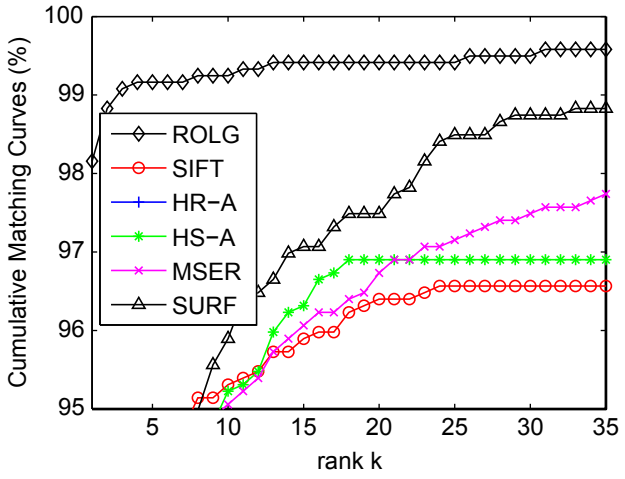

Fig. 18. Cumulative matching curves on FERET database.

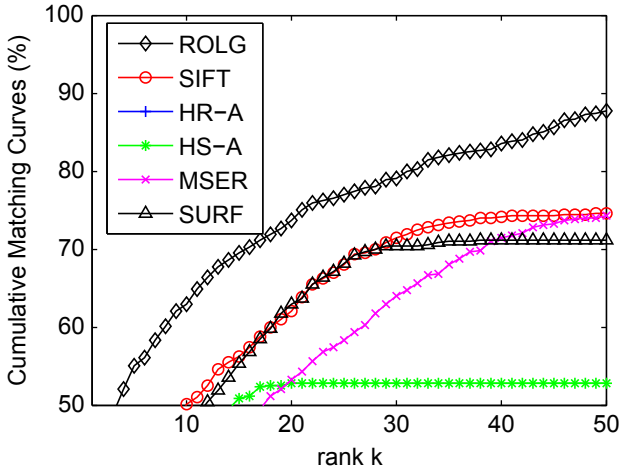

Fig. 19. Cumulative matching curves on LFW database.

Table 1
Rank 1 recognition rate on AR database.

	ROLG	SIFT	SURF	MSER	HS-A	HR-A
AR (%)	98.3	94.3	92.6	92.7	88.6	74.5

Table 2
Rank 1 recognition rate on ORL database.

	ROLG	SIFT	SURF	MSER	HS-A	HR-A
ORL (%)	96.5	90.0	78.5	91.0	80.0	66.5

5.2.1. Results on AR database

Color images of the AR database are converted to gray images and normalized into the size of 60×85. In total, 75 subjects with 14 nonoccluded images per subject are selected. The first seven images of all subjects are chosen as gallery set, and the remaining seven images as probe set.

Table 1 gives the rank 1 recognition rates. Fig. 15 shows the cumulative matching curves. From Table 1, it is clear to see that the ROLG detector outperforms the other detectors. Images in the AR database are taken under controlled conditions of the illumination and viewpoints [41]. The variations of the test images are well represented by the gallery images. Hence, the ROLG detector, the SIFT detector, the SURF detector and the MSER detector achieve high recognition rates. The Harris-affine detector gives the worst performance because human face is a non-rigid surface, and there are few sharp corners in a face image.

5.2.2. Results on ORL database

Images of the ORL database are normalized into the size of 50×57. The first five images of all 40 subjects are chosen as gallery set, and the remaining five images as probe set.

The rank 1 recognition rates are shown in Table 2 and the cumulative matching curves are shown in Fig. 16. Although the ORL database is smaller than the AR database, the performance of all the detectors here is poor compared to that on the AR database. The main reason could be the smaller image size, and the information captured by the detectors on the ORL database is less than that on the AR database. Nevertheless, as shown in Table 2 and Fig. 16, the ROLG detector still outperforms the other detectors.

5.2.3. Results on GT database

Color images of the GT database are converted to gray images and normalized into the size of 60×80. The first eight images of all 50 subjects are chosen as gallery set, and the remaining seven images as probe set.

Rank 1 recognition rates and cumulative matching curves are shown in Table 3 and Fig. 17, respectively. Images in the GT database have large variations in expression, pose and illuminations. Hence, the performance of all the detectors is poor compared with that of the AR database. However, the ROLG detector still outperforms the other detectors, and its rank 1 recognition rate is still larger than 90%.

5.2.4. Results on FERET database

Images in the FERET database are cropped into the size of 60×80. In total, 1194 subjects with two images per person are selected. The first image of all subjects is chosen as gallery set, and the second image as probe set.

The experiment results are shown in Table 4 and Fig. 18. Although the number of subjects of the FERET database is drastically larger than that of the GT database, the rank 1 recognition rates of all detectors on the FERET database are higher than those on the GT database. The reason is that the variation between the gallery set and the test set is small for the FERET

Table 3
Rank 1 recognition rate on GT database.

	ROLG	SIFT	SURF	MSER	HS-A	HR-A
GT (%)	91.1	84.0	84.6	81.1	74.0	47.4

Table 4
Rank 1 recognition rate on FERET database.

	ROLG	SIFT	SURF	MSER	HS-A	HR-A
FERET (%)	98.2	89.9	89.6	89.3	85.3	49.7

Table 5
Rank 1 recognition rate on LFW database.

	ROLG	SIFT	SURF	MSER	HS-A	HR-A
LFW (%)	36.4	27.6	20.9	16.1	16.0	10.4

database. For this high quality database, the ROLG detector significantly outperforms the other detectors over all ranks.

5.2.5. Results on LFW database

Color images are converted to gray images and cropped into the size of 64×64. In total, 134 subjects with 10 images per person are selected. It is seen that these samples have large variations. The first five images of all subjects are chosen as gallery set, and the remaining five images as probe set.

Table 5 and Fig. 19 show the experiment results. For the LFW database, significant variations in face expression, pose, illumination and occlusion exist. These variations result in very poor performance for all detectors in this experiment. For this very difficult database, the ROLG detector also significantly outperforms the other five detectors over all ranks.

6. Conclusions

A novel nonlinear filter named rank order Laplacian of Gaussian (ROLG) filter is proposed, based on which a new interest point detector called ROLG detector is developed in this paper. The proposed ROLG filter is a weighted rank order filter. Compared to the SIFT detector, the ROLG detector detects less spurious and unstable points, and is more robust to abrupt variations of images caused by illumination and geometric changes. Experiment results demonstrate that its performance is better compared to four state-of-the-art detectors in terms of the repeatability and the discrimination of the interest points. The application of interest point detectors to face recognition on five databases further verifies the superiority of the proposed ROLG detector.

Conflict of interest statement

None declared.

References

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, A comparison of affine region detectors, International Journal of Computer Vision 65 (1–2) (2005) 43–72.

[2] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors, International Journal of Computer Vision 37 (2) (2000) 151–172.

[3] T. Tuytelaars, K. Mikolajczyk, Local invariant feature detectors: a survey, Foundations and Trends in Computer Graphics and Vision 3 (3) (2008) 177–280.

[4] M. Brown, D. Lowe, Automatic panoramic image stitching using invariant features, International Journal of Computer Vision 74 (2007) 59–73.

[5] C. Schmid, R. Mohr, Local grayvalue invariants for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (5) (1997) 530–535.

[6] G. Yang, C. Stewart, M. Sofka, C. Tsai, Registration of challenging image pairs: initialization, estimation, and decision, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (11) (2007) 1973–1989.

[7] S. Lazebnik, C. Schmid, J. Ponce, Affine-invariant local descriptors and neighborhood statistics for texture recognition, in: IEEE International Conference on Computer Vision, vol. 1, 2003, pp. 649–655.

[8] R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning object categories from Google's image search, in: IEEE International Conference on Computer Vision, 2005, pp. 1816–1823.

[9] D. Lowe, Object recognition from local scale-invariant features, in: IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157.

[10] H. Riemenschneider, M. Donoser, H. Bischof, Online object recognition by MSER trajectories, in: International Conference on Pattern Recognition, 2008, pp. 1–4.

[11] F. Rothganger, S. Lazebnik, C. Schmid, J. Ponce, 3D object modeling and recognition using affine-invariant patches and multi-view spatial constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. II-272–277.

[12] J. Sivic, F. Schaffalitzky, A. Zisserman, Object level grouping for video shots, International Journal of Computer Vision 67 (2) (2006) 189–210.

[13] C. Geng, X. Jiang, Face recognition based on the multi-scale local image structures, Pattern Recognition 44 (10–11) (2011) 2565–2575.

[14] C. Harris, M. Stephens, A combined corner and edge detector, in: Alvey Vision Conference, 1988, pp. 147–151.

[15] K. Mikolajczyk, C. Schmid, Scale and affine invariant interest point detectors, International Journal of Computer Vision 60 (1) (2004) 63–86.

[16] S. Smith, J. Brady, SUSAN: a new approach to low level image processing, International Journal of Computer Vision 23 (1) (1997) 45–78.

[17] X. Zhang, H. Wang, A. Smith, X. Ling, B. Lovell, D. Yang, Corner detection based on gradient correlation matrices of planar curves, Pattern Recognition 43 (4) (2010) 1207–1223.

[18] P. Beaudet, Rotationally invariant image operators, in: International Conference on Pattern Recognition, 1978, pp. 579–583.

[19] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[20] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF), Computer Vision and Image Understanding 110 (3) (2008) 346–359.

[21] T. Kadir, M. Brady, Saliency, scale and image description, International Journal of Computer Vision 45 (2) (2001) 83–105.

[22] J. Maver, Self-similarity and points of interest, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (7) (2010) 1211–1226.

[23] W. Lee, H. Chen, Histogram-based interest point detectors, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1590–1596.

[24] T. Tuytelaars, L. Van Gool, Matching widely separated views based on affine invariant regions, International Journal of Computer Vision 59 (1) (2004) 61–85.

[25] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, Image and Vision Computing 22 (10) (2004) 761–767.

[26] Q. Li, J. Ye, C. Kambbamettu, Interest point detection using imbalance oriented selection, Pattern Recognition 41 (2) (2008) 672–688.

[27] B. Josef, Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision, Springer, Sweden, 2006.

[28] G. Dorkó, C. Schmid, Maximally stable local description for scale selection, in: European Conference on Computer Vision, 2006, pp. IV: 504–516.

[29] T. Lindeberg, Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention, International Journal of Computer Vision 11 (3) (1993) 283–318.

[30] G. Arce, Nonlinear Signal Processing: A Statistical Approach, Wiley, New York, 2004.

[31] D. Charalampidis, Steerable weighted median filters, IEEE Transactions on Image Processing 19 (4) (2010) 882–894.

[32] K. Choi, A. Morales, S. Ko, Design of linear combination of weighted medians, IEEE Transactions on Signal Processing 49 (9) (2001) 1940–1952.

[33] K. Paler, J. Fölein, J. Illingworth, J. Kittler, Local ordered grey levels as an aid to corner detection, Pattern Recognition 17 (5) (1984) 535–543.

[34] J. Ren, X. Jiang, Eye detection based on rank order filter, in: IEEE International Conference on Information, Communications and Signal Processing, 2009, pp. 1–4.

[35] T. Lindeberg, Feature detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 79–116.

[36] X. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (5) (2009) 931–937.

[37] X. Jiang, B. Mandal, A. Kot, Eigenfeature regularization and extraction in face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3) (2008) 383–394.

[38] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.

[39] X. Jiang, Linear subspace learning-based dimensionality reduction, IEEE Signal Processing Magazine 28 (2) (2011) 16–26.

[40] H. Yu, H. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2070.

[41] A. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 748–763.

[42] F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: Second IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.

[43] Georgia Tech Face Database ⟨http://www.anefian.com/face_reco.htm⟩ (2007).

[44] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

[45] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report 07-49, University of Massachusetts, Amherst (2007).

Zhenwei Miao received the B.Sc. degree in acoustics from the Ocean University of China in 2005, and the M.Sc. degree in acoustics from the Institute of Acoustics, Chinese Academy of Sciences in 2008. He is currently a Ph.D. candidate in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include interest point detection, feature extraction, pattern recognition, computer vision, image processing, biometrics, and acoustic signal processing. He was the recipient of the best paper award at the National Acoustic Conference of Young Scientists 2007.

Xudong Jiang received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China (UESTC) in 1983 and 1986, respectively, and the Ph. D. degree from Helmut Schmidt University Hamburg, Germany, in 1997, all in electrical engineering. From 1986 to 1993, he was a lecturer at UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1993 to 1997, he was with Helmut Schmidt University Hamburg, as a scientific assistant. From 1998 to 2004, he was with the Institute for Infocomm Research, A*STAR, Singapore, as a lead scientist and the head of the Biometrics Laboratory where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition (FVC00). Since 2003, he has been a faculty member in Nanyang Technological University, Singapore. Currently, he is a tenured associate professor. Dr. Jiang has published over 100 papers and holds seven patents. His research interests include signal/image processing, pattern recognition, computer vision, machine learning and biometrics.