IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2019, accepted June 13, 2019, date of publication June 21, 2019, date of current version July 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924238

Interest Point Detection by Limiting Form of Median Log Filter

JUNZHANG CHEN[1], MENGYAO LYU[1], XING WANG[2], XIANGZHI BAI[1,3,4], CHAO YANG[1], MIAOMING LIU[1], AND FUGEN ZHOU[1]

[1]Image Processing Center, Beijing University of Aeronautics and Astronautics, Beijing 100191, China
[2]Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, Beijing 100074, China
[3]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[4]Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China

Corresponding authors: Xing Wang (wangxingtju@126.com) and Xiangzhi Bai (jackybxz@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1736217, in part by the Fundamental Research Funds for the Central Universities under Grant YWF-19-BJJ-43, and in part by the Program for New Century Excellent Talents in Universities under Grant NCET-13-0020.

ABSTRACT Interest point detection has been widely used in image analysis applications. However, some interest points, including small structures and large angle corners, could not be effectively extracted. This paper proposes a limiting form of median Laplacian of Gaussian (LMLG) filter, which combines the superiority of the traditional Laplacian of Gaussian (LoG) filter and a limiting form of the weighted median LoG filter. A detector is also proposed based on the LMLG filter. The LMLG filter aims to improve the detection of LoG-based methods for interest points, especially small structures and large angle corners. Also, it could detect blobs, edges, and local structures. We conduct the repeatability and discrimination experiments on the Oxford dataset. Moreover, we conduct the recall rate experiment on the DTU dataset. The experiments show that the proposed method achieves comparable performance with state-of-the-art methods. In order to verify the utility of the LMLG detector, we carry out a series of interest point detector-based applications: face recognition, infrared-visible image registration, and image classification. The results demonstrate that the LMLG detector performs better than the nine detectors in face recognition. The LMLG detector outperforms the nine detectors and Hrkać's, Han's and Liu's methods in infrared-visible image registration. Our method also gives a comparable result on image classification. The source code of the proposed LMLG detector is made publicly available at https://github.com/chenjzBUAA/LMLG-detector.

INDEX TERMS Feature extraction, image registration, interest point detection, Laplacian of Gaussian filter.

I. INTRODUCTION

In recent years, interest point detection has been widely used in various applications [1], [2], including object recognition [3], [4], object categorization [5]–[8], stereo matching [9], robot localization [10], image registration [11], [12], image retrieval [13], [14], keyframe selection [15], etc. Moreover, interest point detection is an important step in the task of feature extraction. Interest points can also be seen as points with robust saliency, thus can be used in saliency detection [16]–[19]. Besides, interest point detection can be used in multimodal image processing applications. For example, interest point detectors are able to deal with multispectral images such as hyperspectral classification [20] and target detection [21]. Many interest point detectors [22]–[29] have been proposed in the past. Prevalent corner detectors include Harris corner detector [22] and Harris-Laplace/affine detectors [23]. Well-known blob detectors include Hessian detector [25], Hessian-Laplace/affine detectors [23], Scale Invariant Feature Transform (SIFT) detector [30], Speeded-Up Robust Features (SURF) detector [26], the Accelerated KAZE (A-KAZE) detector [31] and Learned Invariant Feature Transform (LIFT) [32]. As for region detectors, popular detectors include Maximally Stable Extremal Regions (MSER) detector [28], edge-based region detector [27] and intensity-based region detector [27].

Corner detectors detect points with large intensity variations in multiple directions. Interest points detected by corner detectors are robust especially with viewpoint changes in image matching. However, corner detectors lack the

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

information of regions [33]. Blob detectors aim at detecting points or regions that are salient in surroundings. Compared to corner detectors, blob detectors are more robust to higher level of distortions, such as scale changes and rotations. But the detection performance of abrupt structures is poorer than corner detectors [33]. Region detectors are designed especially to achieve affine transform invariance. They provide high performance under image transformations such as shape changes and viewpoint changes. However, the shortcoming of region detectors is similar to blob detectors.

The Harris-Affine detector is extended from Harris-Laplace detector to deal with affine transformations. The basic idea of Harris-Affine detector is to select the characteristic scale and characteristic elliptic shape. Harris-Affine detector performs well in illumination and scale variations. However, the filtering operations might cause information loss, which would lead to a performance degradation [34]. Hessian-Affine is similar to Harris-Affine detector. Hessian-Affine detector uses a multiple scale iterative algorithm to select scale and affine invariant points. It is robust to the viewpoint and scale variations [34]. The widely used SIFT detector employs the difference of Gaussian (DoG) filter, which is a close approximation to the scale-normalized Laplacian of Gaussian (LoG) filter, to generate a blob map. However, SIFT detector is unstable for detecting image structures when other interference structures fall in the detection window of the DoG filter. A-KAZE detector is a multi-scale detector in nonlinear scale spaces. Keypoints are detected by A-KAZE detector with nonlinear diffusion filtering, which could reduce noise and obtain more object boundaries compared to LoG based filters. However, A-KAZE might also omits small scale keypoints. LIFT detector is trained in a deep network architecture. The performance of LIFT detector is limited by the training set. Rank order LoG filter [29] (ROLG) is proposed to deal with the problems of SIFT detector. However, there are still some problems left by ROLG detector. Firstly, the detector fails to detect small structures, such as small blobs. Secondly, corners with large angles would be regarded as edges which cannot be detected by ROLG. MSER is one of the most widely used region detectors. Watershed based segmentation algorithms extract regions with intensity inhomogeneities. However, the intensity distribution affects the performance of the extraction. MSER could find blob-like structures in images. It also detects other irregularly shaped structures [2]. SURF is a scale-invariant feature detector, which uses a blob detector based on Hessian matrix to detect interest points. Determinant of Hessian matrix is used for selecting the scale [2]. SURF is effective and convenient for interest point detection. However, missing points might be caused during detection.

In order to enhance LoG-based filters and solve the problems left by ROLG detector, we propose a limiting form of median Laplacian of Gaussian (LMLG) filter in this paper. LMLG reduces the size of the inner disk of ROLG filter to a point when approaching σ to zero. The proposed filter r_{lwm} can effectively detect small structures and corners with large angles in multiple scales. It also incorporates original LoG filter to remove false alarms and improves detection accuracy. Based on this filter, we propose a new interest point detector named LMLG detector. The detector is able to detect local image structures in multiple scales. The visual comparison experiments illustrate that LMLG detector performs better than ROLG, SIFT, SURF, MSER, HS-A, HR-A, FAST, A-KAZE and LIFT detectors in both single and multiple scales. The repeatability and discrimination experiments show that LMLG detector outperforms nine state-of-the-art detectors, which indicates that LMLG detector is more robust to image variations compared with the nine state-of-the-art detectors. In order to verify the practicability of LMLG detector, we employ LMLG detector in the applications of face recognition and infrared-visible image registration. The results demonstrate that LMLG detector performs better than other detectors. Our method also gives a convincing result on image classification.

The remainder of this paper is organized as follows. Section II introduces the traditional LoG filter. Section III-A presents the proposed method and analyses the properties of LMLG detector. In Section IV, visual comparison experiments are demonstrated. In addition, repeatability and discrimination experiments are performed. In Section V, the applications of LMLG detector for face recognition, image registration and image classification are provided. Section VI evaluates the computational cost of different methods. Section VII gives the conclusions.

II. BACKGROUND

The 2D LoG function with Gaussian standard deviation σ has the following form:

$$LoG(u, v, \sigma) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{u^2+v^2}{2\sigma^2}\right] e^{-\frac{u^2+v^2}{2\sigma^2}}. \quad (1)$$

Here u and v are the coordinate of a single pixel. For image I, $I(u, v)$ gives the intensity of (u, v). As described in [29], the output of LoG filter at (u, v) is calculated as follows:

$$\begin{aligned} r_{LoG}(u, v, \sigma) &= -\sum_{(p,q)\in S} LoG(p, q, \sigma) I(u-p, v-q) \\ &= \sum_{(p,q)\in S_2} LoG_-(p, q, \sigma) I(u-p, v-q) \\ &\quad - \sum_{(p,q)\in S_1} LoG_+(p, q, \sigma) I(u-p, v-q), \end{aligned} \quad (2)$$

where S represents the whole mask region; S_1, S_2 represent the mask region where the filter weights are positive and negative, respectively; LoG_+ and LoG_- represent the absolute values of filter weights. The shape of S is a large disk whose size is determined by σ. The shape of S_1 is a surrounding ring and the shape of S_2 is an inner small disk.

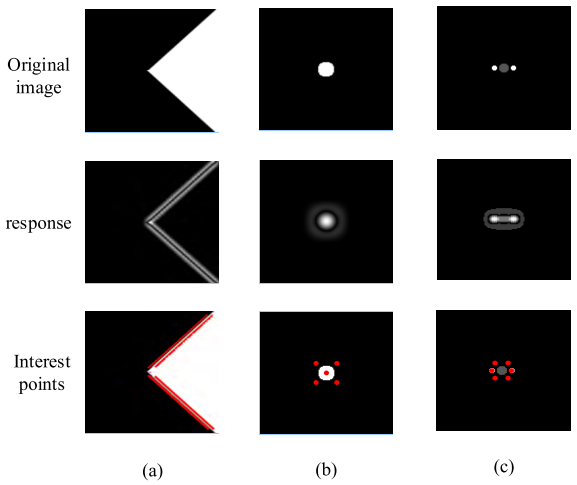


FIGURE 1. The responses and detected results of LoG filter in object structures. (a) Corner. (b) Blob. (c) Blob with abrupt structures.

III. THE PROPOSED METHOD

A. THE PROPOSED LMLG FILTER

Performance of the LoG filter may be interfered by strong or abrupt structures. The filtering result would be affected by strong abrupt structures in the mask region. For example, two high intensity blobs are near the center blob in Fig. 1(c). Nearby blobs weaken the response of LoG filters, which might cause the missing detection of the center blob.

In this case, it is a good idea to replace the weighted mean values in regions S_1 and S_2 with the outputs of weighted rank order filters [29].

To improve the performance of LoG filter, we replace the weighted mean values with the weighted median values in regions S_1 and S_2. The output of weighted median LoG filter can be expressed as follows:

$$r_{wm}(u,v,\sigma) = Med(I_-(u,v,\sigma), LoG_-, r_w) - Med(I_+(u,v,\sigma), LoG_+, r_w). \quad (3)$$

Here $I_-(u,v,\sigma) = \{I(u-p, v-q)|(p,q) \in S_2\}$, $LoG_- = \{LoG_-(p,q,\sigma)|(p,q) \in S_2\}$,

$I_+(u,v,\sigma) = \{I(u-p, v-q)|(p,q) \in S_1\}$, $LoG_+ = \{LoG_+(p,q,\sigma)|(p,q) \in S_1\}$, the weight $LoG_-(p,q,\sigma)$ or $LoG_+(p,q,\sigma)$ corresponds to the input $I(u-p, v-q)$.

$Med(x,w,r_w)$ is a weighted median rank order filter, where $x = \{x_1, x_2, \ldots, x_n\}$, $w = \{w_1, w_2, \ldots, w_n\}$, $(0 \le w_i \le 1, \sum_{i=1}^{n} w_i = 1)$ and r_w $(0 \le r_w \le 1)$ represent the input series, the weights and the rank. The value of $Med(x,w,r_w)$ can be calculated as follows [35], [36].

Firstly, the input series are sorted in increasing order. The corresponding weights of the sorted input series $\tilde{x} = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$ are rearranged as $\tilde{w} = \{\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n\}$.

The output of the weighted median rank order filter is given by

$$Med(x,w,r_w) = Med(\tilde{x},\tilde{w},r_w) = \tilde{x}_i, \ \{i: \ b_{i-1} < r_w \le b_i\}, \quad (4)$$

where $b_i = \sum_{j=1}^{i} \tilde{w}_j$ is the cumulative sum of the sorted weights and $b_0 = 0$.

Besides, the output of $Med(x,w,r_w)$ should be the weighted median value of x, so we set $r_w = 0.5$.

For example, here is an input series $x = \{7,5,9,8\}$ with weights $w = \{0.1, 0.2, 0.3, 0.4\}$. The sorted input series $\tilde{x}$ and weights $\tilde{w}$ would be 5, 7, 8, 9, and 0.2, 0.1, 0.4, 0.3, respectively. The accumulation weights $\{b_0, b_1, b_2, b_3, b_4\} = \{0, 0.2, 0.3, 0.7, 1\}$, and $Med(x,w,r_w) = Med(\tilde{x},\tilde{w},0.5)$. Given $b_2 < 0.5 \le b_3$, $\tilde{x}_3 = 8$ is the weighted rank order filter output of the input series x.

However, the filter in Eq. (3) fails to detect large structures because the weighted median filter in the inner disk region cannot capture small target structures. Likewise, the filter in Eq. (3) cannot accurately locate the vertices with small angles.

In order to address these problems and further suppress the responses on edges of the detector, we proposed a limiting form of the median weighted LoG filter. After the scale of the filter is set, σ is set to approach zero. Then the limiting form of the weighted median LoG filter in Eq. (3) can be expressed as

$$r_{lwm}(u,v) = \lim_{\sigma \to 0} \{Med(I_-(u,v,\sigma), LoG_-, r_w) - Med(I_+(u,v,\sigma), LoG_+, r_w)\} \quad (5)$$

In practice, σ cannot turn to 0 for digital image processing. Thus, we use the center pixel of the LoG mask as the region of S_2. In this case, the proposed $r_{lwm}(u,v,\sigma)$ is given as follows:

$$r_{lwm}(u,v,\sigma) = \tilde{I}(u,v,\sigma) - Med(\tilde{I}(u,v,\sigma), W, 0.5), \quad (6)$$

where $\tilde{I}(u,v,\sigma)$ is the Gaussian filter output of $I(u,v,\sigma)$, $\tilde{I}(u,v,\sigma) = \left\{\tilde{I}(u-p, v-q, \sigma)|(p,q) \in S\right\}$, $W = \left\{\frac{1}{|S|}, \ldots, \frac{1}{|S|}\right\}$, S represents the whole mask region of the LoG filter, $|S|$ represents the cardinality of pixels in the whole mask which is determined by the parameter σ.

According to Eq. (1), value of $LoG(u,v,\sigma)$ equals zero when $r = \sqrt{2}\sigma$, where $r = \sqrt{u^2+v^2}$. Here r is the radius of S_2. S_2 of r_{lwm} is a pixel which can be seen as a disk with radius 0.5 ($r = 0.5$). In this case, the shape of filter r_{lwm} is the same as $LoG(u,v,\frac{1}{2\sqrt{2}})$. The difference of Eq. (5) and Eq. (6) can be defined as:

$$Diff(s) = \int_{-s}^{s}\int_{-s}^{s} \left(\lim_{\sigma \to 0} (LoG(u,v,\sigma)) - LoG(u,v,\frac{1}{2\sqrt{2}}) \right) dudv. \quad (7)$$

Here s is the scale of the mask S. According to the property of LoG function, $Diff(s)$ decreases and approaches zero with a growing s when $s > 0$.

Filter (6) can detect small structures because the center point in the mask of this filter can capture small structures. Moreover, filter (6) can suppress the responses of edges

because the center point and median filter catch the same side of edges. Due to the median filter, filter (6) may not be interfered by strong or abrupt structures.

Although filter (6) can solve the problems of the LoG filter and filter (3), its performance is not stable and its scope of application is less extensive than the LoG filter because it is a limiting form of the weighted median LoG filter. To get a stable and precise filter, we design our filter as follows:

$$r_{LMLG} = \begin{cases} r_{LoG} \times r_{lwm}, & if\ r_{LoG} > 0, r_{lwm} > 0 \\ -r_{LoG} \times r_{lwm}, & if\ r_{LoG} < 0, r_{lwm} < 0 \\ 0, & otherwise. \end{cases} \quad (8)$$

the design of r_{LMLG} is based on the following considerations.

Firstly, the values of $\tilde{I}(u, v, \sigma)$ and $Med(\tilde{I}(u, v, \sigma), W, 0.5)$ are both integers. If $\tilde{I}(u, v, \sigma) \neq Med(\tilde{I}(u, v, \sigma), W, 0.5)$, $r_{lwm}(u, v, \sigma) = \tilde{I}(u, v, \sigma) - Med(\tilde{I}(u, v, \sigma), W, 0.5) \geqslant 1$. Thus, the output of r_{LMLG} of the interest points can be expanded by multiplying the outputs of r_{LoG} and r_{lwm}.

Secondly, the output of r_{lwm} on edges is zero. In this case, multiplying r_{LoG} and r_{lwm} can make LMLG filter output zero on edges, which preserves the insensitivity of LMLG filters on edges.

In addition, instead of simply multiplying the two outputs, we value the output of LMLG filter based on the sign of r_{LoG} and r_{lwm}. Normally, the outputs of r_{LoG} and r_{lwm} should have the same sign while detecting a blob or a corner. So, if the outputs of r_{LoG} and r_{lwm} have different signs, the detected point will not be a correct interest point and the output of LMLG filter should be set to zero. Thus, formula (8) could remove some false interest points.

B. ANALYSIS OF LMLG FILTER

1) RESPONSES OF LMLG FILTER ON BLOBS

LoG filter produces some spurious local extrema around blobs. These spurious local extrema cause erroneous interest points detected by LoG filter based detectors. The first row of Fig. 2(b) shows that a ring is produced around the peak at the center of the blob. The first row of Fig. 2(c) shows that four erroneous points are detected around the blob by detecting the peaks of LoG filter responses.

The first row of Figs. 2(f) and (g) shows that LMLG filter produces only one peak at the center of the blob, thus only one point is detected by detecting the peaks of LMLG filter responses. This means that LMLG filter can avoid erroneous points around blobs. This can be explained by the filtering process of LMLG filter on a 1D blob shown in Fig. 3. When the 1D-LMLG-filter mask is on one side of the 1D blob, e.g. mask 1 in Fig. 3, inputs within mask 1 are monotonically increasing. LMLG filter will output zero because the value of the center point in mask 1 is equal to the output of median filter within mask 1. When the 1D-LMLG-filter mask is at the center of the 1D blob, e.g. mask 2 in Fig. 3, LMLG filter will output a positive value because the value of the center point in mask 2 is larger than the output of median filter within mask 2. Extending the 1D case to 2D case, we can know why LMLG filter does not generate a ring around the blob.

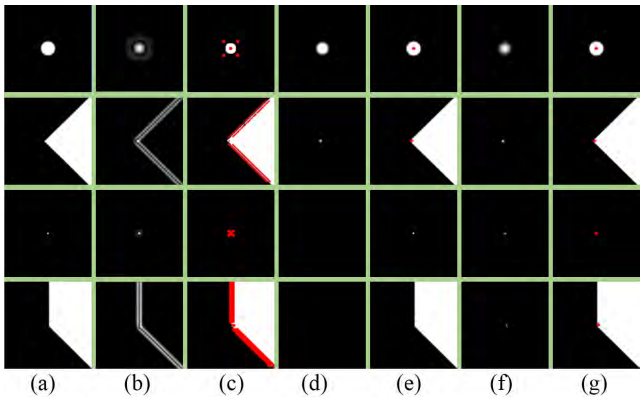


FIGURE 2. The responses and detected results of different filters. Red points are the detected interest points. (a) 2D structure. (b) Absolute values of the LoG responses. (c) Points detected by the LoG filter. (d) Absolute values of ROLG responses. (e) Points detected by ROLG filter. (f) Absolute values of LMLG responses. (g) Points detected by LMLG filter.

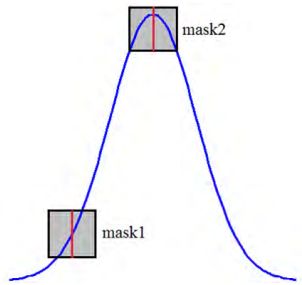


FIGURE 3. A 1D blob and two 1D-LMLG-filter masks. In each mask, the red line represents its center.

2) RESPONSES OF LMLG FILTER ON EDGES AND CORNERS

Another problem of LoG filter is that spurious local extrema are often produced along edges. These spurious local extrema will cause erroneous points detected by LoG based detectors. The second row of Fig. 2(b) shows that near the peak on the corner, strong responses are also produced along the edge. The second row of Fig. 2(c) shows that lots of erroneous points are detected along the edge by detecting the peaks of LoG filter responses.

The second row of Figs. 2(f) and (g) shows that LMLG filter produces only one peak on the corner and detects only one point by detecting the peaks of LMLG filter responses. This means that LMLG filter can avoid detecting erroneous points along edges. This can be explained by the filtering process of LMLG filter on a 1D edge shown in Fig. 4. When the 1D-LMLG-filter mask is on the edge, inputs within the mask are monotonically increasing. LMLG filter will output zero because the value of the center point in the mask is equal to the output of the median filter within the mask. Extending the 1D case to 2D case, LMLG filter does not generate spurious local extrema along edges.

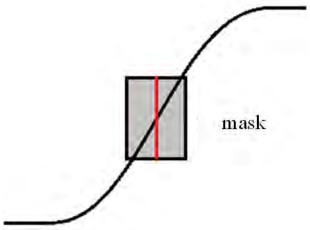


FIGURE 4. A 1D edge and a 1D-LMLG-filter mask. In the mask, the red line represents its center.

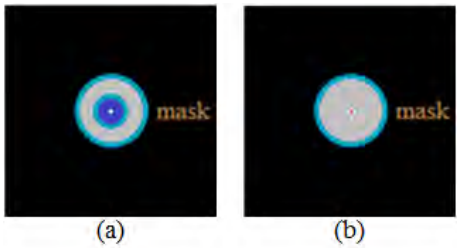


FIGURE 5. (a) A 2D small blob and a 2D-LoG-filter mask. In this mask, the blue region represents its inner part and the grey region represents its surrounding part. (b) A 2D small blob and a 2D-LMLG-filter mask. In this mask, the grey region represents its whole region and the red point represents its center point.

FIGURE 6. A comparison of interest points detected by two filters. (a) LoG filter. (b) LMLG filter. The red points are the detected interest points.

3) RESPONSES OF LMLG FILTER ON LOCAL STRUCTURES

Through above analysis, we find that LMLG filter has some advantages. LMLG filter can avoid erroneous points around blobs and suppress the responses of edges. Moreover, LMLG filter is able to detect small structures and corners with large angles.

Firstly, the third row of Figs. 2(f) and (g) shows that LMLG filter outputs a peak on the small blob and detects a point. This can be explained by the filtering process shown in Figs. 5(a) and (b). Fig. 5(b) shows that the center point in the mask of LMLG filter can always capture small blobs. Likewise, as shown in Fig. 6, LMLG filter can detect thin endpoints.

In addition, the last row of Figs. 2(f) and (g) shows that LMLG filter outputs a peak on the corner and detects a point. This can be explained by the filtering process shown in Figs. 7(a) and (b). Fig. 7(a) shows that the output condition cannot be satisfied when a significant majority of pixels (60% is used) in the inner disk are brighter than a significant majority of pixels (60% is used) in the surrounding ring. But Fig. 7(b) shows that the value of the center point in

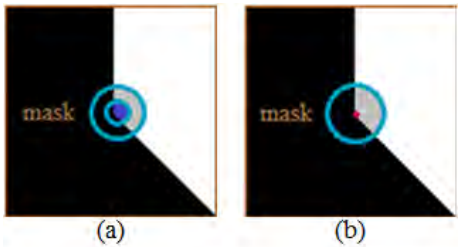


FIGURE 7. (a) A 2D corner and a 2D-LoG-filter mask. In this mask, the blue region represents its inner part and the grey region represents its surrounding part. (b) A 2D corner and a 2D-LMLG-filter mask. In this mask, the grey region represents its whole region and the red point represents its center point.

LMLG filter mask is always larger than the median value within the mask which satisfies the output condition of LMLG filter.

Based on the above three statements, we illustrated that LMLG can detect common interest points such as blobs, edges, corners and local structures.

C. PROPERTY ANALYSIS

From (8), four properties are given below.

Property 1. LMLG filter represent multi-scale interest points.

Since Gaussian kernel is a transformation kernel to achieve scale-space representation of signal, LMLG filter could represent multi-scale interest points through changing σ.

Property 2. Scales in LMLG filter are discretely continuous.

As for the number of image intervals s and constant factor $k = 2^{1/s}$, there are $s + 2$ images in LoG scale space. Their scales are $\sigma, k\sigma, k^2\sigma, k^3\sigma, \ldots, k^s\sigma, k^{s+1}\sigma$. Similarly to the scale of SIFT [30], these s scales $k\sigma, k^2\sigma, k^3\sigma, \ldots, k^s\sigma$ in the middle are chosen as the scales in the first octave. The images in the next octave is obtained by downsampling of the image in its previous octave. Then the scales in the second octave are $2\sigma, 2k\sigma, 2k^2\sigma, 2k^3\sigma, \ldots, 2k^s\sigma, 2k^{s+1}\sigma$. Likewise, the scales $2k\sigma, 2k^2\sigma, 2k^3\sigma, \ldots, 2k^s\sigma$ in the middle are chosen as the scales in the second octave. Obviously, the first scale $2k\sigma$ in the second octave is continuous with the last scale $k^s\sigma$ in the first octave. In the same way, the first scale in the current octave is continuous with the last scale in the previous octave. Thus, scales in LMLG detector are discretely continuous.

Property 3. LMLG filter is able to detect small structures, such as blobs and thin endpoints with single pixel width, in the smallest scale.

Since the mask of LMLG filter in the center is a point and can catch small interest structures, LMLG filter is able to detect blobs and thin endpoints with single pixel width. The single-pixel-width blobs detected by LMLG filter is shown in the third row of Fig. 2(g). The single-pixel-width thin endpoints detected by LMLG filter is shown in Fig. 6(b).

Property 4. LMLG filter is able to detect corners with large angles in the smallest scale.

As for the edges of the corners with large angles, it can be found that the center point and the median filter will catch

the same side of the edges from filter (6). Thus, the filter (6) output zero in the edges, that is, LMLG filter can suppress the responses of edges. Moreover, LMLG filter retains the advantage of the filter (6) and can catch the corner. That is to say, the response of LMLG filter in the corner is not zero. As a consequence, LMLG filter outputs a peak on the corner and detects a point. The corner with large angles detected by LMLG filter is shown in the fourth row of Fig. 2(g).

D. INTEREST POINT DETECTION BASED ON LMLG FILTER

Interest points are obtained by detecting the peaks which are on the corner/blob map generated by LMLG filter. Given one scale parameter σ, interest points can be detected in a single scale. To detect interest points in multiple scales, different interest points are detected in different scales by different scale parameters, as done in [29] and [37]. Since the responses of LMLG filter on a ridge are strong, some erroneous points on ridges are detected by the detector. We use the algorithm proposed in [30] to remove these erroneous points on ridges.

The main idea of LMLG detector can be summarized as follows:

Firstly, peaks are detected on the corner/blob map generated by LMLG filter. After removing the peaks which are on ridges, interest points are detected in a single scale. By using different scale parameters, interest points are detected in multiple scales.

Then the major steps of interest point detection using LMLG detector can be summarized as follows.

(1) The initial scale parameter σ and the number of scales are assigned firstly.

(2) LMLG filter is used to generate the corner/blob map by filtering the input image.

(3) The peaks that have been detected on the corner/blob map are taken as the candidate interest points.

(4) The peaks on ridges are removed based on the algorithm proposed in [30]. Remaining peaks are the true interest points in this scale.

(5) After updating LMLG filter by increasing the scale parameter, the detection step goes back to step 2 in a larger scale until the maximum scale is reached.

IV. EXPERIMENTS

We used the experiment setting suggested in [30] to set the parameters. All scales were divided into octaves and each octave contains 3 scales $\left\{1.6 \times 2^{1/3},\ 1.6 \times 2^{2/3},\ 3.2\right\}$. After the image was down-sampled and the interest points were detected in a new octave. The number of the octaves was determined by image size.

Nine comparison methods are used in our experiment, including ROLG [29], SIFT [30], SURF [26], MSER [28], Hessian-affine (HS-A) detector [23], Harris-affine (HR-A) detector [23], FAST [38], A-KAZE [31] and LIFT detectors [32]. We use the same parameters of HS-A and HR-A detectors as in [39]: thresholds of HS-A and HR-A are 500 and 1000 respectively. SIFT, SURF, MSER, FAST and A-KAZE are implemented in OpenCV 3.0 using default parameters. SIFT: contrast threshold $=$ 0.04, $\sigma = 1.6$; SURF: octave $=$ 4, threshold $=$ 500; MSER: $\delta = 5$, area threshold $=$ 1.01; FAST: threshold $=$ 10; A-KAZE: octave $=$ 4, threshold $=$ 0.001. ROLG is also implemented in the same settings as mentioned in [29]: $\sigma = 1.6$. Model of LIFT is trained as suggested in [32] with Piccadilly Circus dataset. Based on our experiment, all of the comparison methods obtain the best performance in these settings.

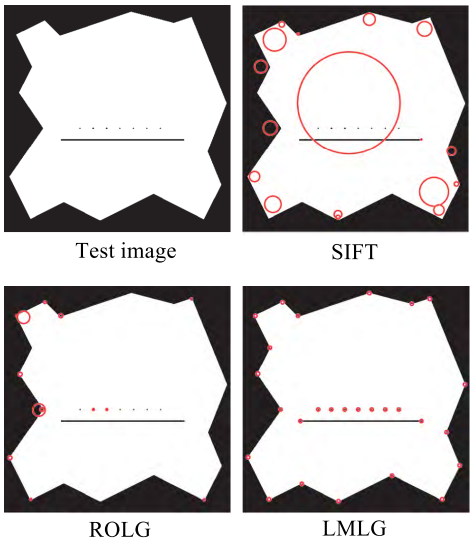


FIGURE 8. The detection results of SIFT, ROLG and LMLG on a test image. The test image contains varied structures, including sharp angles, obtuse angles, small size points, and a straight line.

A. VISUAL COMPARISON EXPERIMENTS

We carried out some visual experiments by comparing interest points detected by LMLG detector and the comparison methods. Fig. 9 shows that ROLG, SIFT, SURF, HS-A, HR-A, A-KAZE and LIFT detectors omitted some obvious blobs such as the mole near the corner of the mouth. Meanwhile, ROLG and SIFT detectors detected some negligible and erroneous points such as the points below the eyes. Region detector MSER missed the mole near the mouth. MSER also detected less interest points than LMLG. LMLG detector detects the obvious blobs omitted by ROLG, SIFT, SURF, A-KAZE and LIFT detectors such as the mole. Simultaneously, LMLG detector avoid some negligible and erroneous points. Compared with ROLG, SIFT, SURF, MSER, HS-A, HR-A, FAST, A-KAZE and LIFT detectors, LMLG captures more blobs and corners. This is mainly because the product of the two filters widens the gap between interest points and others. Thus, the interest points are easier to be detected and the negligible points are easier to be avoided.

In [40], a simple test pattern is used to evaluate the method. The input image contains circular patterns with increasing scales. The detection results of the input image demonstrate the ability of their method to detect multi scale circles. Inspired by [40], we evaluated LMLG detector on an artificial

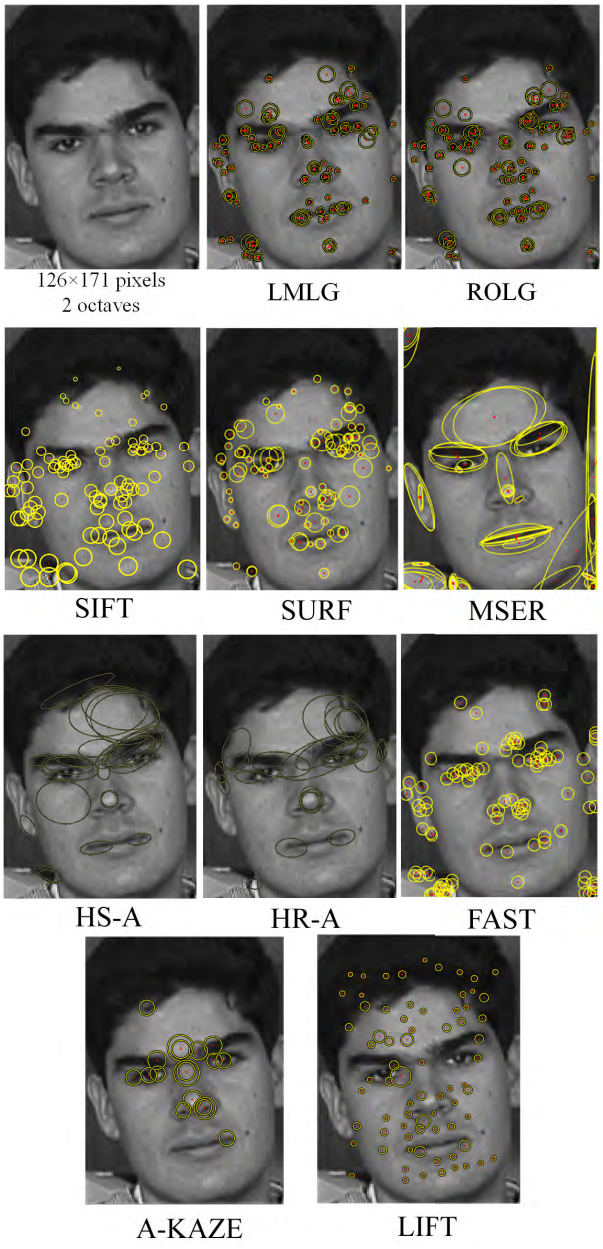


FIGURE 9. The comparison results of interest points location and scales. The red points are the detected interest points and the radius of the yellow circle represents the scale of the interest point at its center.

image. As shown in Fig. 8, the test image contains varied structures, including sharp angles, obtuse angles, small size points, and a straight line. Fig. 8 gives the results obtained by SIFT, ROLG and the proposed LMLG. SIFT fails to detect small size points. Also, there are false alarms around blobs and corners. ROLG and LMLG avoids that using a rank order filter instead of a weighted mean filter. However, ROLG omits obvious corners with large angles, small structures and endpoints. We can find that proposed LMLG can successfully meet all above-mentioned challenges and avoid erroneous points around blobs, which validates the analysis we carried out in III-C.

B. REPEATABILITY AND DISCRIMINATION EXPERIMENTS

In order to evaluate LMLG detector under different image variations, we carried out the repeatability and discrimination experiments. We used the evaluation suggested in [39] to make comparisons with nine detectors, including ROLG detector, SIFT detector, SURF detector, MSER detector, Hessian-affine (HS-A) detector, Harris-affine (HR-A) detector, FAST detector, A-KAZE detector and LIFT detector. The repeatability measures include the number of repeated interest points and the repeatability score. Each point corresponds to a detected region. Two regions are repeated if their overlap error is less than a threshold (set to 40% in our experiments). The repeatability score was computed as the ratio between the number of the repeated points and the smaller number of the detected points in the pair of images. The discrimination measures include the number of matched points and matching score. Two regions are matched if they are repeated and their SIFT descriptors are the nearest-neighbor in descriptor space. The matching score was computed as the ratio between the number of matched points and the smaller number of detected points in the pair of images. Larger values of repeatability and matching score indicate good performance of corresponding method.

In our experiments, the database provided by [39] was used to evaluate detectors. This database contains eight datasets. Some examples of the database are shown in Fig. 11. These datasets contain five different image variation conditions: JPEG compression, image blur, illumination change, scale and rotation change, and view point change. Each dataset consists of six images with five homographies between the first image and the other five images. The sizes of images range from 765×512 pixels to 1000×700 pixels.

For all the images, the interest points were detected by LMLG detector in 5 octaves. By comparing the interest points in the first image and the other five images, the evaluation measures between the first image and the other five images can be calculated. From the average evaluation measures in the eight datasets for each detector in Fig. 10, we can find that LMLG detector obtained higher average values of repeatability and matching score than the other detectors. What is more, LMLG detector obtained larger numbers of repeated points and matched points than the other detectors. From the number of repeated points and matched points, we can find that LMLG detector detected much more interest points than the other detectors. Experimental results show that LMLG detector is robust to image variations and can detect interest points with high repeatability and discrimination. Besides, the experiment database contains images with distortion and quality variation. The result of this experiment show that LMLG outperforms other methods, with shows its stability and robustness.

C. RECALL RATE EXPERIMENT ON DTU DATASET

To better evaluate the performance of different detectors under illumination changes and scene variations, we conduct an experiment on the DTU dataset [41]. The DTU dataset

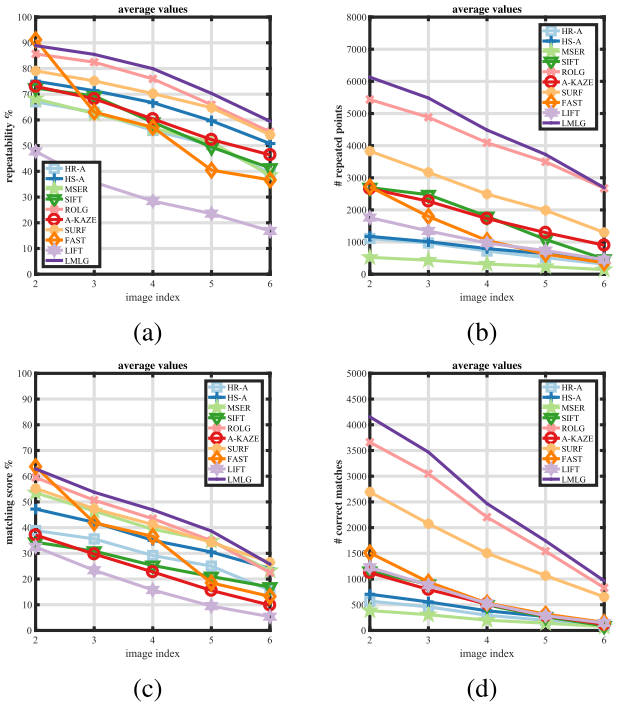


FIGURE 10. The average values of evaluation measures in 8 experiments of different filters. There are 6 images in each group, the horizontal coordinator image index indicates the index of the image been tested with the first image in one experiment. For example, the node of the 4th image index in LMLG curve in image (a) gives the average value of repeatability of the first and 4th image. (a) Repeatability. (b) Number of repeated points. (c) Matching score. (d) Number of matched points.

contains 60 scenes, such as model houses, fruit and vegetables, fabric. Images of each scene are acquired in 119 camera positions. And, 19 individual LED illuminations are used for each camera position.

The 119 positions are grouped into three horizontal arcs and a straight path away from the scene. The three arcs are distributed in circular paths with radii of 0.5 m (Arc 1, positions 1-49), 0.65 m (Arc 2, positions 65-94) and 0.8 m (Arc 3, positions 95-119). The straight path ranges from $0.5 - 0.8$ m (positions 50-64). Ranges of the angles of the arcs from the key frame are: $\pm 40°$, $\pm 25°$, $\pm 20°$. Performance of interest points detection on viewpoint and distance changes can be evaluated on the frames taken in these positions.

We adopted the experiment setting as in [41], where images are down-sampled to 600×800 in grayscale. Also, we use the recall rate, which is suggested in [41], to evaluate the performance of the proposed LMLG detector and eight other detectors. The recall rate here is the ratio of potential matches to total interest points of the key frame (frame 25). The match we used in our experiment is the same as in [41], where interest points of two images are matched only when they fulfill all three criteria below:

1. Corresponding interest points should be within 2.5 pixels from the epipolar line of the camera positions;

2. Corresponding interest points should be within a window of interest with a radius of 5 pixels;

3. Corresponding interest points are within a scale factor of 2 from each other.

To evaluate the performance of interest point detection on viewpoints and distance changes, we evaluate the recall rate on 118 frames of the DTU dataset. Scale of each interest point is essential in this experiment. To be fair, we only choose multi-scale detectors.

However, due to the properties of interest point detectors, interest points detected cannot be restrained to exact the same number by setting parameters. Thus, to make a fair comparison, we propose an experiment using the same number of detected interest points. The experiment is named random-200 experiment. For every detector, firstly, the interest points of images in the DTU dataset are detected, and from which we randomly select 200 interest points for every image. This could ensure the number of interest points used for each method is equal. Finally, the recall rate of each method for images is obtained by calculating the number of matches between keyframe and every image. Images in the DTU dataset are acquired in various camera positions and lighting conditions. Thus, recall rate can be used as the evaluation metric of the performance of each detection method. The number 200 is chosen because the minimum number of interest points detected on the DTU dataset by all detectors is 254. We carried out the experiment 5 times, and calculate the mean value of recall rates as given in Fig. 12 and table 1. We can see that our method gives a comparable performance on viewpoint changes (Fig. 12 a,b,c) and distance changes (Fig. 12 d).

Moreover, to evaluate the performance on lighting changes, we obtain the recall rate on 19 illumination conditions. Mean value of the recall rate of 19 illumination conditions of DTU dataset is given in Fig. 13. The result also gives a comparable performance with state of the art methods.

V. APPLICATIONS

In order to verify the performance of LMLG detector in real tasks, we carried out some applications based on LMLG detector including face recognition, infrared-visible image registration and image classification.

A. FACE RECOGNITION

Face recognition is a challenging and active research area. In recent years, various methods are proposed [42], [43]. It is known that face recognition is an important application of interest point detectors, which can be used to evaluate the effectiveness of interest point detectors.

We performed the nearest-neighbour matching procedures in [30] for face recognition to compare the discriminative power of interest point detection by eight detectors, including LMLG, ROLG, SIFT, SURF, MSER, Hessian-affine (HS-A) detector, Harris-affine (HR-A) detector, FAST, A-KAZE and LIFT. Firstly, interest points were detected by all the detectors and then their SIFT descriptors were computed. Secondly, the minimum Euclidean distance of SIFT descriptors between the image pair from the gallery set and the probe set was found. Finally, the image class of the nearest neighbor

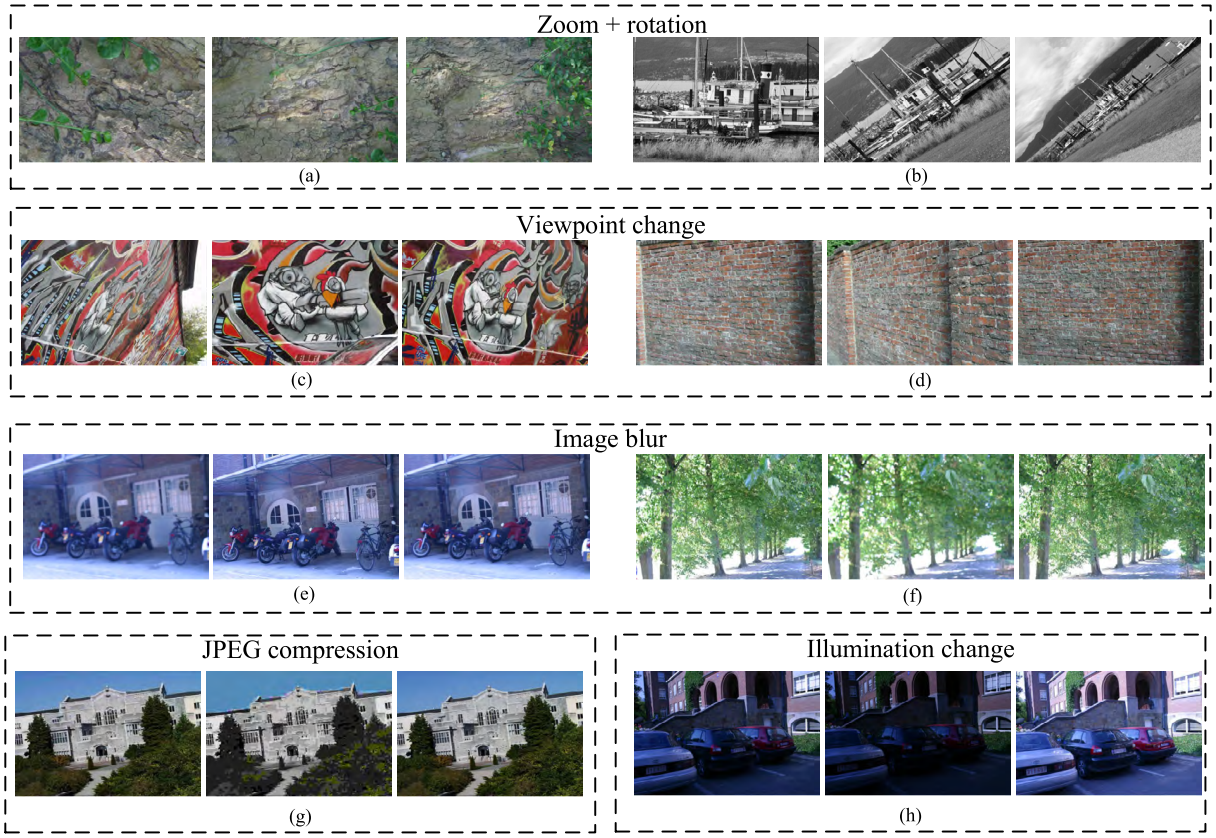


FIGURE 11. Dataset of the repeatability and discrimination experiments. Zoom+rotation: (a) bark, (b) boat. View point change: (c) graf, (d) wall. Image blur: (e) bikes, (f) trees. JPEG compression: (g) ubc. Illumination change: (h) leuven.

TABLE 1. Mean value of the recall rates in random-200 experiment.

Detector	LMLG	ROLG	SIFT	SURF	MSER	HS-A	HR-A	A-KAZE	LIFT
Recall rate (%)	18.84	14.46	15.62	17.88	14.94	7.56	7.90	22.37	15.78

TABLE 2. The recognition rate on face databases.

	LMLG	ROLG	SIFT	SURF	MSER	HS-A	HR-A	FAST	A-KAZE	LIFT
AR(%)	**89.2**	81.8	82.5	42.8	33.38	39.0	39.5	54.5	43.5	62.5
ORL(%)	**91.5**	83.7	70.8	63.0	60.4	46.1	36.0	78.8	54.2	64.3
GT(%)	**84.3**	78.8	72.8	62.0	34.1	40.4	36.2	70.0	34.1	52.8

in the gallery set was assigned to the image in the probe set.

We conducted face recognition procedure in images of three face databases, including AR [44], ORL [45], GT [46] database. There are 100 subjects in AR [44] database which includes 14 nonoccluded images per subject. Color images are converted to gray images and then normalized into the size of 60 × 80 pixels. There are 40 subjects in ORL [45] database and the size of each image is 92 × 112 pixels. For Georgia Tech (GT) [46] database, color images are converted to gray images and the size of images ranges from 50 × 75 pixels to 91 × 140 pixels.

Table 2 gives the recognition rates of LMLG, ROLG, SIFT, SURF, MSER, HS-A, HR-A, FAST, A-KAZE and LIFT based methods. As depicted in Table 2, the performances of HS-A, HR-A and MSER detectors in AR, ORL and GT databases are lower than the other detectors. The results of LMLG, ROLG, FAST and LIFT are comparable in AR database. For GT and ORL database, result of LMLG detector is better than the other detectors. From Table 2, we can find that LMLG detector outperforms the other detectors in all three face databases. This indicates that LMLG detector can detect accurate interest points.

B. INFRARED-VISIBLE IMAGE REGISTRATION

Image registration is one of the important image processing applications [47]. As a kind of image registration, infrared-visible image registration is a challenging task. It is difficult

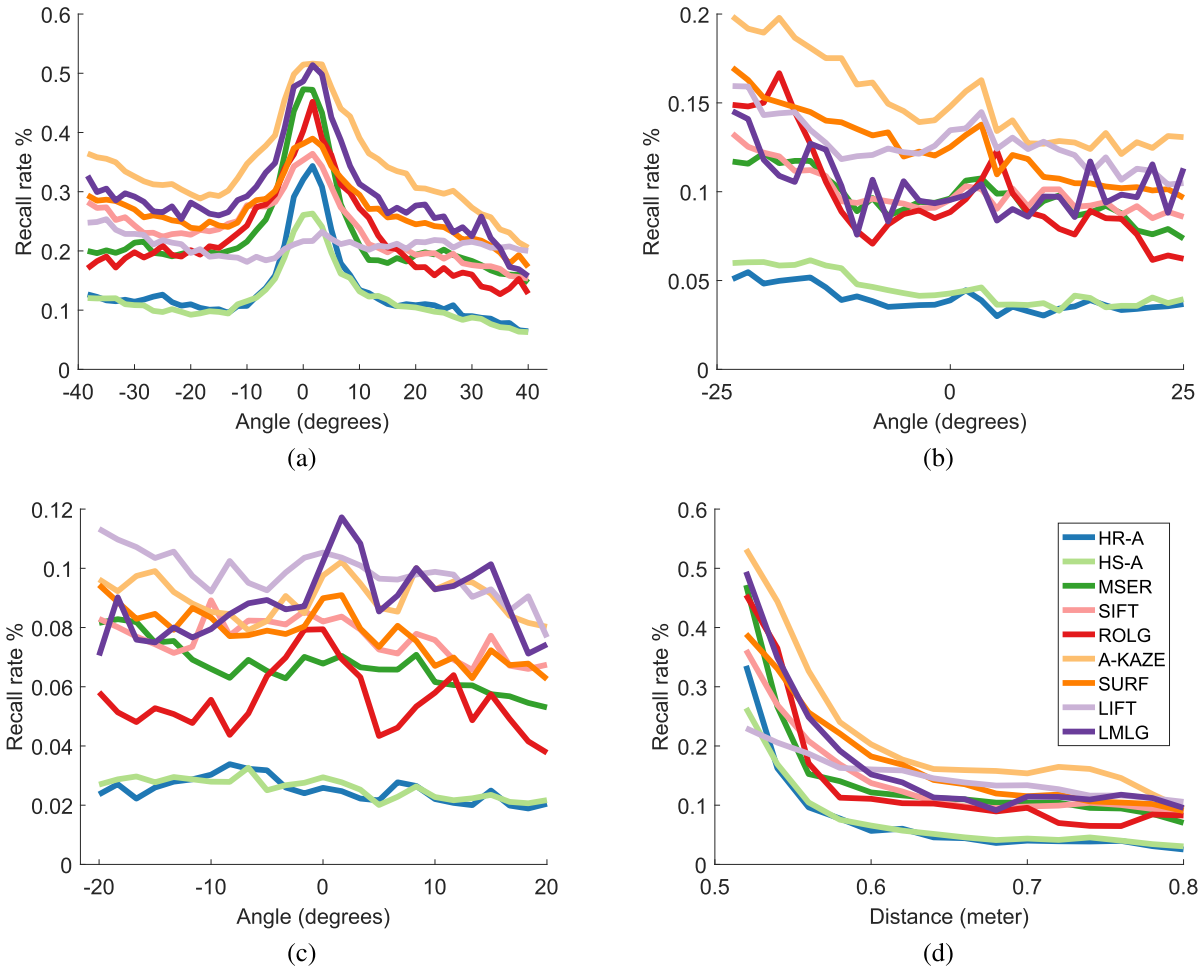


FIGURE 12. Mean values of the recall rate in random-200 experiment. The graphs show the recall rate of all camera positions. (a) Arc 1, ±40°. (b) Arc 2, ±25°. (c) Arc 3, ±20°. (d) Straight path, distance: 0.5m-0.8m.

to construct effective algorithms for this application. In this section, we proposed a method for infrared-visible image registration to verify the practicability of LMLG detector. This image registration method employed geometric structure features including interest points and image edges, which adopted a coarse-to-fine procedure. This registration method was carried out in three main stages: feature extraction stage, coarse registration stage based on edge alignment and fine registration stage based on interest point matching. In the feature extraction stage, interest points were detected by an interest point detector. In the coarse registration stage, an initial approximate transform was computed by Canny [48], [49] edge alignment through searching for the transform which can make the edges of the two images overlap. In the fine registration stage, candidate matching region of each interest point was reduced to the neighborhood region of the transformed point and the false interest point matching pairs are removed. Then, the more accurate transform was computed to refine the registration.

In order to better illustrate the performance of LMLG detector, we compared LMLG based registration method with the nine detectors and three other registration algorithms on a same dataset in visual and quantitative ways. The three other algorithms used for comparison are Hrkać's method based on corners and hausdorff distance [11], Han's method based on line-based geometric analysis [50] and Liu's method based on SIFT flow [51]. All of the three methods are representative feature-based methods for infrared-visible image registration.

The dataset includes 20 infrared-visible image pairs, which are from the literatures [52]–[54] and our cooperators. The dataset includes image pairs taken in outdoor and indoor scenes, in natural and man-made environments, under various illumination conditions, from different directions, and at different times of a day. The scenes of the images include indoor furniture, urban buildings, landscape, persons and military targets. The illumination conditions of the images include brightness, dark, obscuration and faint lamplight. The shooting times of the images include dawn, daytime, evening and night. The image size ranges from 224 × 106 pixels to 700×556 pixels. All the image pairs differ greatly in intensity and there is contrast reversal in many image regions. Some difficult image pairs differ greatly in scale and orientation.

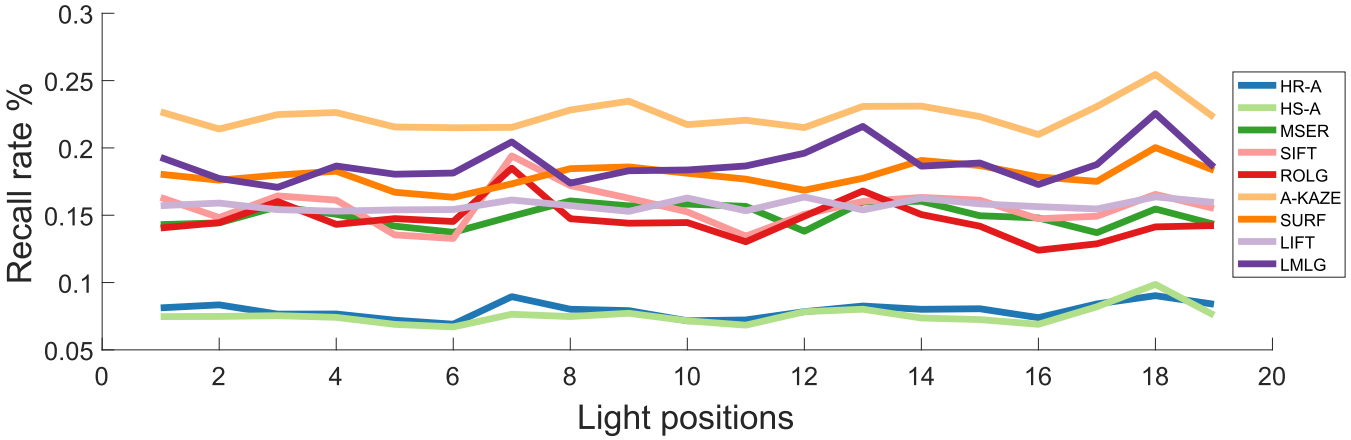


FIGURE 13. Mean values of the recall rate in 19 lighting conditions in random-200 experiment.

TABLE 3. The average transform errors for 20 image pairs.

	LMLG	ROLG	Hrkać's [11]	Han's [50]	Liu's [51]	SIFT	SURF	FAST	MSER	HS-A	HR-A	A-KAZE	LIFT
AVE	**34.21**	35.50	101.11	109.02	47.32	128.62	48.58	97.37	115.23	163.74	168.46	56.98	40.71
STD	**1.65**	1.88	45.15	40.41	11.24	22.53	3.09	8.78	18.60	43.23	18.74	5.18	1.89

Some of them have complex and irregular backgrounds, such as the waves. Thus, the registration task on the dataset is very challenging.

The registered result images were employed for visual comparison, which was constructed by overlaying the Canny edges of the transformed visible image onto the corresponding infrared image. If the transformed visible edges and infrared edges overlap very well in the result image, the image pair is registered accurately. Some of the result images are shown in Fig. 14. From Fig. 14, it can be found that Hrkać's method, Han's method and Liu's method cannot make the visible and infrared edges overlap very well. It can also be found that the scenes in the result images obtained by Liu's method were twisted in different degrees because the computed dense correspondences between images were not accurate enough. ROLG based method could register some of the image pairs, but some of them were not accurate enough. SIFT and FAST are good at matching indoor image pairs, however they failed at some outdoor image pairs. A-KAZE can match the outdoor images precisely, however the result seems a little worse on the indoor images. The matching results of SURF and LIFT are fine in both indoor and outdoor images. However, the results were a little worse than LMLG. MSER, HS-A and HR-A do not perform well for infrared-visible image registration. Most of their matching results were far from the correct matching. Moreover, MSER, HS-A and HR-A were all failed at giving a matching result of some image pairs.

As shown in Fig. 14, our method based on LMLG detector can register the image pairs with high accuracy and high success rate, while the other algorithms can only register parts of the image pairs successfully and some registration results of them were far from satisfactory. The matching results indicate the interest points detected by LMLG detector is more accurate than other detectors and algorithms.

In order to evaluate the accuracy of different registration algorithms quantitatively, we measured the transform errors of different algorithms. The transform error is defined by the distance between a point in an infrared image and its transformed point in the corresponding visible image, which is given by

$$error = \|H \cdot P_{Vis} - P_{IR}\| . \tag{9}$$

Here P_{Vis} and P_{IR} are the coordinates of two matching points in the visible and infrared images, H represents the transform matrix.

For each image pair, we selected 20 matching pairs manually and calculated their transform errors. Then we calculated their average and standard deviation values for comparison, which were represented by *AVE* and *STD* respectively. The smaller the value of *AVE* is, the more accurate the transform is calculated. The smaller the value of *STD* is, the more uniformly an image pair is registered. For example, *AVE* value of Han's method is 109.02. It means the corresponding pair points are with a distance of average 109.02 pixels. As the second column images shown in Fig. 14, Han's method gives a large mistake of registration. The average *AVE* and *STD* values of our algorithm for 20 image pairs are only 34.2 and 1.65 pixels as shown in Table 3, which are smaller than the values of the other twelve methods. In other words, our results are very satisfactory, which demonstrates the high practicability of LMLG detector in infrared-visible image registration.

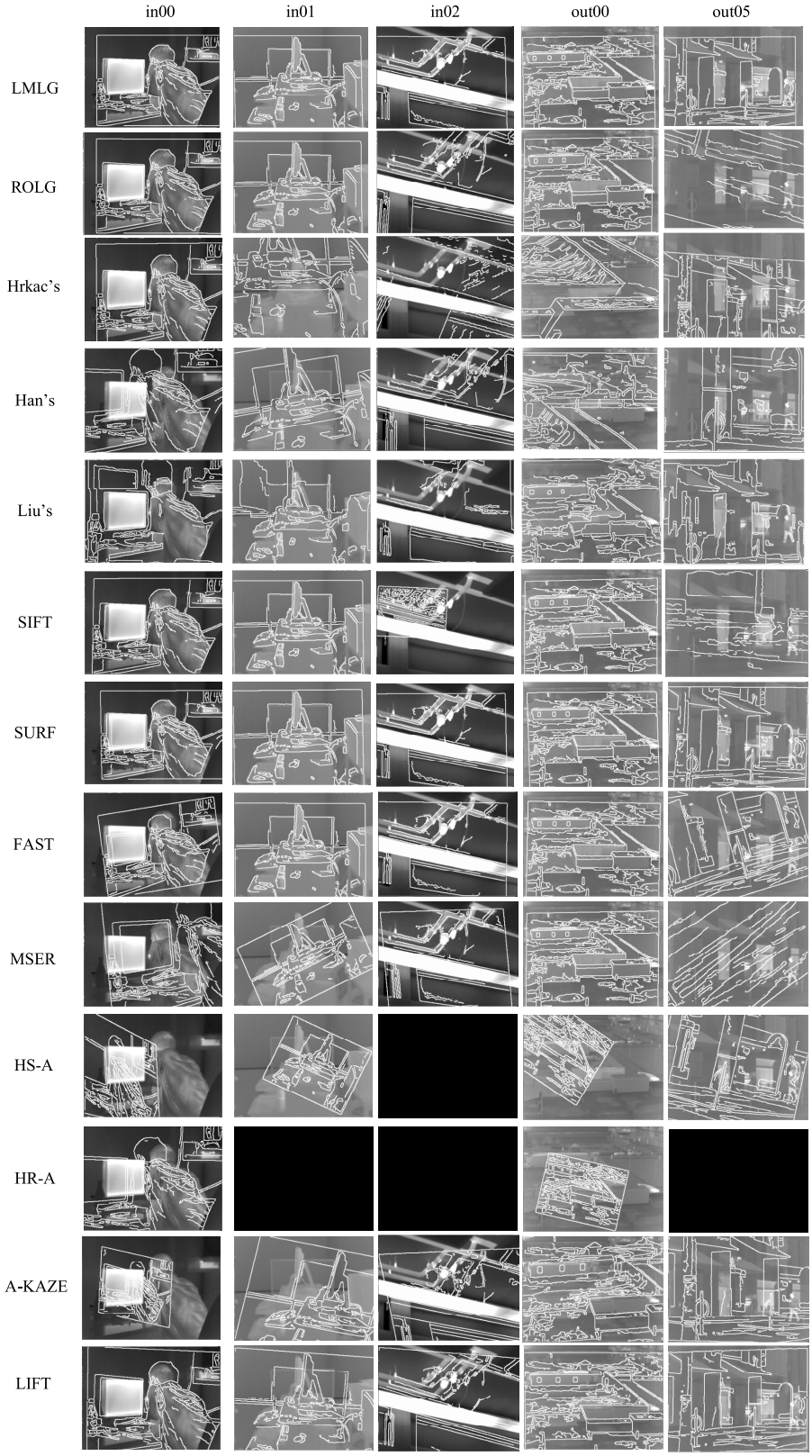


FIGURE 14. Some matching results obtained by different algorithms. The black image represents that the method cannot work while registering an image pair.

TABLE 4. The classification result of all methods.

	LMLG	ROLG	SIFT	SURF	MSER	HS-A	HR-A	FAST	A-KAZE	LIFT
Accuracy(%)	**83.73**	81.19	76.95	82.88	76.27	77.46	79.32	73.73	74.58	87.12

TABLE 5. Running time of all methods.

	LMLG	ROLG	SIFT	SURF	MSER	HS-A	HR-A	FAST	A-KAZE	LIFT
time(s)	**3.481**	5.352	0.143	0.025	0.162	0.271	0.630	0.002	0.121	0.923

C. IMAGE CLASSIFICATION BASED ON INTEREST POINT DETECTION

Image classification is to categorize images into different classes. It plays an essential role in many applications. Both state-of-the-art convolution neural networks and interest point detectors provide features for classifiers. Therefore, image classification can also evaluate hand-crafted features.

In the experiment, we performed image category classification using a bag of features based approach. Among the database provided by VOC2012, two categories (bus and horse) were chosen to classify for their preponderance of train/validation images. We adjusted the number of images per category in the dataset to make a balance. Then we split the datasets into 70% training and 30% testing sets.

First, interest points of the images in the training set were detected by eight detectors, including LMLG, ROLG, SIFT, SURF, MSER, HS-A, HR-A, FAST, A-KAZE and LIFT. Then, their SIFT descriptors and feature metric were computed. We kept 80% of the strongest features from each category and balanced the number of features across all image categories. After that, K-Means clustering was applied to create a 500 word visual vocabulary, which used the squared Euclidean distance measure and the k-means algorithm for the initialization of the cluster center. Finally, images were encoded into feature vectors by visual vocabulary, which were then fed into a support vector machine (SVM) multi-class classifier using framework of the error correcting output codes (ECOC).

The classification result is given in Table 4. As shown in Table 4, LIFT gives the highest accuracy. This is mainly due to the fact that the training set contains VOC2012. Our method achieves the second high precision in this task. The results of ROLG and SURF are close to LMLG. The classification rate of SIFT, MSER, HS-A, HR-A, FAST and A-KAZE are lower than the other methods, which indicate their detection result of images are less accurate than the other methods.

The precision of LMLG is the second high among all the methods, which indicates the detection performance of LMLG is high when applied in image classification.

VI. COMPARISON OF RUNNING TIME

For the comparison of running time with other methods, we detect interest points on the "bark" dataset provided by [39]. LMLG, ROLG, MSER, SURF and FAST detectors are implemented by MATLAB. SIFT, HS-A, HR-A and A-KAZE are implemented by C++. LIFT detector is implemented in Python. The experiments are conducted under Inter(R) Core(TM) i7-5960X CPU, 3.00GHz, and 32.00 GB RAM. The test images are with the size of 765 × 512. As Table 5 shows, LMLG and ROLG have higher computational cost than the other methods. This is because the computational complexity of median filters are higher than linear filters. However, the computational cost of LMLG is lower than ROLG, which can be attributed to the structure of LMLG filter. LMLG filter only contains one median filtering, which is simpler than ROLG. In the future, we consider improving efficiency of our method by designing a more effective median rank order filter used in our detector. Therefore, the time complexity of LMLG can be significantly decreased with the improvement, creating the opportunity for real-time applications.

VII. CONCLUSIONS

In this paper, the limiting form of median Laplacian of Gaussian (LMLG) filter is proposed. After that, a new interest point detector named LMLG detector is developed. Comparing with other detectors in the visual comparison experiments, LMLG detector can detect more interest points and less erroneous points. In the repeatability and discrimination experiments on Oxford dataset and recall rate experiment on DTU dataset, LMLG gives the comparable performance with state of the art methods. The results show LMLG is more robust in image variations and distortion. Finally, the applications in face recognition, infrared-visible image registration, image classification verify the effectiveness of LMLG detector.

ACKNOWLEGMENT

(Junzhang Chen and Mengyao Lyu contributed equally to this work.)

REFERENCES

[1] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, "Survey of spatio-temporal interest point detection algorithms in video," *IEEE Access*, vol. 5, pp. 10323–10331, 2017.
[2] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Fundations Trends Comput. Graph. Vis.*, vol. 3, no. 3, pp. 177–280, Jan. 2008.
[3] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.

[4] H. Riemenschneider, M. Donoser, and H. Bischof, "Online object recognition by MSER trajectories," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[5] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from Google's image search," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Oct. 2005, pp. 1816–1823.

[6] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 634–639.

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2003, pp. 264–271.

[8] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, "Weak hypotheses and boosting for generic object detection and recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 71–84.

[9] J.-S. Park, H.-E. Kim, H.-Y. Kim, J. Lee, and L.-S. Kim, "A vision processor with a unified interest-point detection and matching hardware for accelerating a stereo-matching algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 12, pp. 2328–2343, Dec. 2016.

[10] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Robot. Res.*, vol. 21, no. 8, pp. 735–758, 2002.

[11] T. Hrkać, Z. Kalafatić, and J. Krapac, "Infrared-visual image registration based on corners and Hausdorff distance," in *Proc. Scand. Conf. Image Anal.*, vol. 4522, 2007, pp. 383–392.

[12] P. Mainali, Q. Yang, G. Lafruit, L. Van Gool, and R. Lauwereins, "Robust low complexity corner detector," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 435–445, Apr. 2011.

[13] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.

[14] T. Tuytelaars and L. Van Gool, "Content-based image retrieval based on local affinely invariant regions," in *Proc. Int. Conf. Vis. Inf. Syst.*, vol. 1614, 1999, pp. 493–500.

[15] G. Guan, Z. Wang, S. Lu, J. D. Deng, and D. D. Feng, "Keypoint-based keyframe selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 729–734, Apr. 2013.

[16] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.

[17] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, 2016.

[18] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.

[19] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 141–145, Jan. 2006.

[20] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[21] Y. Dong, B. Du, L. Zhang, and X. Hu, "Hyperspectral target detection via adaptive information—Theoretic metric learning with local constraints," *Remote Sens.*, vol. 10, no. 9, p. 1415, 2018.

[22] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, vol. 15, 1988, pp. 147–151.

[23] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.

[24] S. M. Smith and J. M. Brady, "SUSAN—A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, no. 1, pp. 45–78, 1997.

[25] P. Beaudet, "Rotationally invariant image operators," in *Proc. Int. Joint Conf. Pattern Recognit.*, vol. 579, 1978, pp. 579–583.

[26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[27] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, 2004.

[28] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 384–393.

[29] Z. Miao and X. Jiang, "Interest point detection using rank order LoG filter," *Pattern Recognit.*, vol. 46, no. 11, pp. 2890–2901, Nov. 2013.

[30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[31] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. Brit. Mach. Vis. Conf.*, 2013, pp. 1–11.

[32] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "LIFT: Learned invariant feature transform," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 467–483.

[33] S. Krig, "Interest point detector and feature descriptor survey," in *Computer Vision Metrics*. Springer, 2016, pp. 187–246.

[34] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3D objects," *Int. J. Comput. Vis.*, vol. 73, no. 3, pp. 263–284, 2007.

[35] G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach*. Hoboken, NJ, USA: Wiley, 2005.

[36] D. Charalampidis, "Steerable weighted median filters," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 882–894, Apr. 2010.

[37] W.-T. Lee and H.-T. Chen, "Histogram-based interest point detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1590–1596.

[38] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2006, pp. 430–443.

[39] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, 2005.

[40] K. Gou and K. Uchimura, "Scale-space processing using polynomial representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2744–2751.

[41] H. Aanæs, A. L. Dahl, and K. S. Pedersen, "Interesting interest points," *Int. J. Comput. Vis.*, vol. 97, no. 1, pp. 18–35, 2012.

[42] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G. Aachen, "SURF-Face: Face recognition under viewpoint consistency constraints," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–11.

[43] C. Geng and X. Jiang, "Fully automatic face recognition framework based on local and global features," *Mach. Vis. Appl.*, vol. 24, no. 3, pp. 537–549, 2013.

[44] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

[45] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Dec. 1994, pp. 138–142.

[46] L. Chen, H. Man, and A. V. Nefian, "Face recognition based on multi-class mapping of Fisher scores," *Pattern Recognit.*, vol. 38, no. 6, pp. 799–811, 2005.

[47] M. N. Haque, M. Biswas, M. R. Pickering, and M. R. Frater, "A low-complexity image registration algorithm for global motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 426–433, Mar. 2012.

[48] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[49] Y. Zhang and P. I. Rockett, "The Bayesian operating point of the canny edge detector," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3409–3416, Nov. 2006.

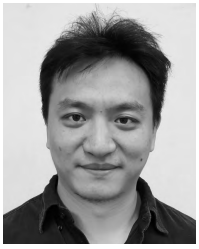
[50] J. Han, E. Pauwels, and P. de Zeeuw, "Visible and infrared image registration employing line-based geometric analysis," in *Proc. Int. Workshop Comput. Intell. Multimedia Understand.*, vol. 7252, 2012, pp. 114–125.

[51] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 978–994, May 2011.

[52] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–7.

[53] G. H. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, Nov. 2007.

[54] J. Kang, K. Gajera, I. Cohen, and G. Medioni, "Detection and tracking of moving objects from overlapping EO and IR sensors," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun./Jul. 2004, p. 123.



JUNZHANG CHEN received the B.S. degree from the Huazhong University of Science and Technology (HUST), in 2009. He is currently pursuing the Ph.D. degree with the Beijing University of Aeronautics and Astronautics (BUAA).

MENGYAO LYU received the B.S. degree from Beijing Information Science and Technology University (BISTU), in 2018. She is currently pursuing the master's degree with the Beijing University of Aeronautics and Astronautics (BUAA).

XING WANG received the B.S. and master's degrees from Tianjin University (TJU), in 2010 and 2013, respectively. His research interests include scene matching, image registration, and visual navigation.

XIANGZHI BAI received the B.S. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), in 2003 and 2009, respectively, where he is currently a Full Professor with the Image Processing Center. He holds ten national invention patents and has published more than 100 international journal and conference papers in the field of mathematical morphology, image analysis, pattern recognition, and bioinformatics. He also acts as an Active Reviewer for around 60 international journals and conferences.

CHAO YANG received the B.S. degree from the Beijing University of Aeronautics and Astronautics (BUAA), in 2012. His research interest includes point detection.

MIAOMING LIU received the B.S. degree from the Beijing University of Posts and Telecommunications (BUPT), in 2014. Her research interest includes point detection.

FUGEN ZHOU received the B.S. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), in 1986 and 2006, respectively. His research interests include object recognition, image compressing, and restoration.