

Research Article

A Novel Approach for Interest Point Detection via Laplacian-of-Bilateral Filter

Mingming Huang,¹ Zhichun Mu,¹ Hui Zeng,¹ and Hongbo Huang^{1,2}

¹*School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

²*Computing Center, Beijing Information Science and Technology University, Beijing 100192, China*

Correspondence should be addressed to Zhichun Mu; mu@ies.ustb.edu.cn

Received 7 September 2014; Revised 12 January 2015; Accepted 29 January 2015

Academic Editor: Chi Chiu Chan

Copyright © 2015 Mingming Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Scale-invariant feature transform (SIFT) algorithm, one of the most famous and popular interest point detectors, detects extrema by using difference-of-Gaussian (DoG) filter which is an approximation to the Laplacian-of-Gaussian (LoG) for improving speed. However, DoG filter has a strong response along edge, even if the location along the edge is poorly determined and therefore is unstable to small amounts of noise. In this paper, we propose a novel interest point detection algorithm, which detects scale space extrema by using a Laplacian-of-Bilateral (LoB) filter. The LoB filter, which is produced by Bilateral and Laplacian filter, can preserve edge characteristic by fully utilizing the information of intensity variety. Compared with the SIFT algorithm, our algorithm substantially improves the repeatability of detected interest points on a very challenging benchmark dataset, in which images were generated under different imaging conditions. Extensive experimental results show that the proposed approach is more robust to challenging problems such as illumination and viewpoint changes, especially when encountering large illumination change.

1. Introduction

Interest points (together with the small image patch around them) are local features for which the signal changes two-dimensionally. Due to their advantages of robustness, efficiency, and the ability of working without initialization, interest points have proven to be very successful in many applications such as image retrieval [1–4], object recognition [5–8], wide baseline matching [9–11], texture recognition [12], robot localization [13], and object categorization [14–18].

A desirable property of interest points is their robust repeatability [5, 18], which means that interest points should be repeatable and stable under both local and global perturbations. In the case of significant transformations, interest points have to be adapted to the transformations, as at least a subset of local features must be present in both images for the purpose of allowing for correspondence. Research efforts concerned in interest points can be divided into two categories, namely, detector and descriptor. Detector locates an interest point in the image, and descriptor designs features

to characterize the detected interest point. The most valuable property of an interest point detector is its repeatability, which expresses the reliability of a detector for finding the same physical interest points under different viewing conditions. Then, the neighborhood of every interest point is represented by a feature vector. This descriptor should be distinctive and at the same time robust to noise and a class of image transformations. Recently, a wide variety of interest point detectors have been proposed. Widely used interest point detectors include Harris-affine detector and its affine normalization [19, 20], maximally stable extremal regions (MSER) [21], features from accelerated segment test (FAST) [22], and the Hessian-affine detector [19]. All of these state-of-the-art interest point detectors have different strengths and weaknesses and yield different number of points depending on the image. They can be divided into three families: single-scale, direct-intensity, and multiscale [23]. The single-scale interest point detectors filter the image with a single-scale filter. The direct-intensity based methods compute a measure that indicates the presence of an interest point directly

from the gray values. The multiscale interest point detection algorithms detect the scale of the image features and they are strongly invariant to scale changes and other image artifacts such as illumination variation, noise, blur, rotation, and affine transformation.

Scale-invariant feature transform (SIFT) algorithm proposed by Lowe [5] is one of the most famous and popular interest point detectors and has been proven to be robust in many applications [24]. The SIFT interest points have a robust repeatability property against image translation, rotation, and scaling. Generally speaking, there are three major stages in a typical SIFT algorithm. The first stage is to detect extrema in DoG scale space. Secondly, interest points are filtered and located. The final stage is to assign orientation and generate descriptor for interest points. Even though a large number of works have been devoted to improving SIFT, much effort of most existing works in the literature has been spent in focusing on the second and third stage, that is, improving the description power of the descriptor for an interest point. Ke and Sukthankar [25] applied principal component analysis (PCA) to gradient patch of an interest point and reduced descriptor from 128D to 36D to form PCA-SIFT. The gradient location and orientation histogram (GLOH), proposed by Mikolajczyk and Schmid [24], replaced the Cartesian location grid used by the SIFT with a log-polar one and also applied PCA to reduce the dimension of the descriptor. Winder and Brown [26] made use of discriminative learning method to optimize local descriptors under matching constraints from a 3D construction. Abdel-Hakim and Farag [27] extended the SIFT algorithm to extract colored local invariant feature descriptor, named as Color-SIFT. Verma et al. [28] proposed new color-SIFT descriptors by extending the SIFT descriptor to different color space. Li et al. [29] proposed a new learning-to-rank framework to reject unstable extrema such as low contrast points and edge response points derived from the first stage of SIFT in order to improve interest point detection. Liao et al. [30] presented a modification to the SIFT in which the normalized elliptical neighboring region was instead of a rectangular region and a histogram was computed in a polar space. By contrast, only limited work has been devoted to solving problems in the first stage, that is, identifying locations and scales of potential interest points that can be repeatably assigned under differing views of the same object.

In this paper, we apply Bilateral filter and Laplacian filter on SIFT algorithm and present a novel method to detect repeatable interest point with Laplacian-of-Bilateral (LoB) filter. We first smooth images by successively larger Bilateral filter and a series of smoothed images are called scale space or image pyramid. Compared with conventional Gaussian filter, the Bilateral filter can remove noise from images while preserving edges. Afterwards, we apply fixed window Laplacian filter on smoothed images in order to produce difference images called Laplacian-of-Bilateral (LoB) scale space. As a consequence, scale space extrema detection can be implemented efficiently in a series of LoB images by convoluting an image with LoB filter (with different scales). The next steps are similar to standard SIFT algorithm for location, orientation assignment, and descriptor generation.

The rest of the paper is organized as follows. In Section 2, we give a brief overview of the SIFT algorithm and explain its drawbacks. Then, our proposed interest point detection algorithm is presented in Section 3. Experiment results are discussed in Section 4. Finally, we conclude this paper in Section 5.

2. SIFT Algorithm Review

Before presenting in detail our approach, we briefly review the SIFT that forms the basis for our work. The SIFT algorithm proposed in [5] usually consists of three steps. First, a Gaussian scale space is constructed and candidate points are extracted by searching local extrema in a series of DoG images. Then all candidate points are localized to pixel-accuracy and unstable points of low contrast or strong edge response are eliminated. Finally, dominant orientation is identified for each survived point, and its descriptor is generated based on the image gradients in its local neighborhood.

In order to provide a clear background for further discussion, we give more insight into the details of the first step in SIFT. In the first step, the scale space or image pyramid is constructed by a variable-scale Gaussian, which is

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right), \quad (1)$$

where (x, y) denotes the coordinate of a point and the scale factor is σ .

Mikolajczyk found that the extrema of scale-normalized Laplacian-of-Gaussian (LoG), $\sigma^2\nabla^2G$, are the most stable local points on an image compared to a range of other image functions such as gradient, Hessian, and Harris. Moreover, the difference-of-Gaussian filter provides a close approximation to the scale-normalized LoG, which is

$$G(x, y, k\sigma) - G(x, y, \sigma) = (k - 1) \sigma^2 \nabla^2 G, \quad (2)$$

where k is a constant factor. So scale space extrema extracted in the DoG function convolved with the image, $D(x, y, \sigma)$, are regarded as candidate points, and

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \quad (3)$$

where $I(x, y)$ is an input image.

An effective way to construct $D(x, y, \sigma)$ and detect local extrema is shown in Figure 1. The initial image is incrementally blurred with successively larger Gaussian filters to produce images which are separated by a constant multiplicative factor in scale space, shown stacked in the left column. Afterwards, difference-of-Gaussian images are produced by subtracting each blurred image from the adjacent (more blurred) image, shown on the right. For detecting the local extrema (maxima or minima) of $D(x, y, \sigma)$, each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below. The sample point is selected as a candidate point if its value is larger or smaller than all of its neighbors.

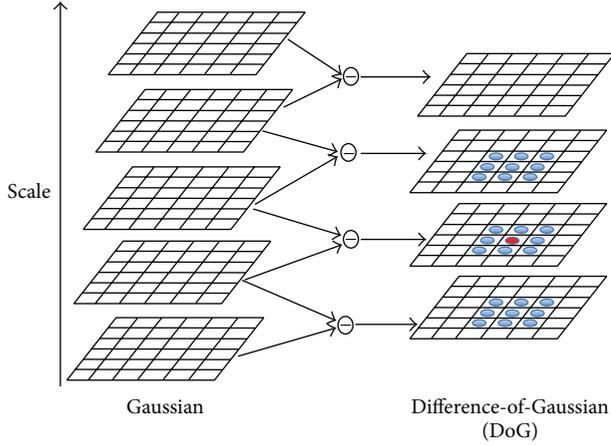


FIGURE 1: The process of local extrema detection based on Difference-of-Gaussian (DoG).

Discussions. Based upon our observations, the SIFT algorithm has two inevitable drawbacks.

The difference-of-Gaussian filter has a strong response along edge, even if the location along the edge is poorly determined and therefore is unstable to small amounts of noise. As a result, some interest points on the edge or in low contrast regions tend to be removed after detection, because two handcrafted rules: discarding low contrast points and eliminating edge response are used in the SIFT algorithm.

Moreover, in the SIFT algorithm, candidate points are extracted by searching local extremum in a series of DoG images. It has been shown by Lowe [5] that the more expensive operations in extracting features are applied only at locations that pass an initial test, which means that the major computational complexity occurs in filtering the image. Unfortunately, in order to detect extrema at scales of each octave, intervals images should be included in difference-of-Gaussian (DoG) scale space and intervals images are required for Gaussian scale space. It is completely disadvantageous to the cost of filtering the image with the increase of Gaussian scale space images. In the next section, we propose designing interest point detector based on both Bilateral filter and Laplacian filter.

3. An Overview of the Proposed Approach

3.1. The Bilateral Filter. Tomasi and Manduchi [31] combined range filtering with domain filtering to produce bilateral filters. The Bilateral filtering can be defined as follows:

$$h(x, y) = \frac{1}{w_p} \sum_{i, j \in \Omega} w_d(i, j) w_r(i, j) f(i, j), \quad (4)$$

where (x, y) denotes the pixels of an image, Ω is the neighborhood of a center point (x, y) , $f(i, j)$ is defined as an input image, $h(x, y)$ is an output image, $w_d(i, j)$ denotes the weight of the domain filtering, $w_r(i, j)$ is the weight of the range filtering, and w_p is the normalization parameter: $w_p = \sum_{i, j \in \Omega} w_d(i, j) w_r(i, j)$. The size of Ω is $(2N+1) \times (2N+1)$ and N corresponds to the half-width of the Bilateral filter.

0	-1	0
-1	4	-1
0	-1	0

(a)

-1	-1	-1
-1	8	-1
-1	-1	-1

(b)

FIGURE 2: Two commonly used 3×3 sized discrete approximations to the Laplacian filter.

Because the Bilateral filter includes both a geometric distance in the domain component and a photometric distance in the range component, it can remove noise and preserve edges from images during smoothing.

3.2. The Laplacian Filter. A Laplacian filter is a derivative filter used to find areas of rapid change (edges) in images. As such, this filter type is commonly used in edge-detection applications. Since derivative filters are very sensitive to noise, it is common to smooth the image (e.g., using a Gaussian filter) before applying the Laplacian. The Laplacian $L(x, y)$ of an image with pixel intensity values $f(x, y)$ is given by

$$L(x, y) = \nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}. \quad (5)$$

Since the input image is represented as a set of discrete pixels, we have to find a discrete convolution kernel that can approximate the second derivatives in the definition of the Laplacian. Two commonly used small kernels are shown in Figure 2.

3.3. The Bilateral Scale Space Construction. Considering the drawbacks of the SIFT algorithm, some interest points on the edge easily removed after detection, we proposed the use of Bilateral filtering instead of Gaussian filtering in scale space construction.

Lowe [5] defined the scale space of an image as a function, $L(x, y, \sigma)$, that is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, described in (1), with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (6)$$

where $*$ is the convolution operation in x and y .

Similar to the Gaussian convolution operation, Bilateral filtering is also defined as a weighted average of the pixel values. Also, in contrast with traditional filter such as Gaussian filter, Bilateral filtering, which employs both geometric closeness and photometric similarity of neighboring pixel, can preserve edge characteristic by fully utilizing the information of intensity variety. As a consequence, we proposed constructing the scale space by the Bilateral filter named as Bilateral scale space. We define the scale space of an image as a function $B(\mathbf{u}, \sigma_d)$ as follows:

$$B(\mathbf{u}, \sigma_d) = \frac{1}{W_{\mathbf{u} \in \Omega}} \sum_{\mathbf{v} \in \Omega} I_{\mathbf{v}} G_{\sigma_d} G_{\sigma_r}, \quad (7)$$

where $\mathbf{u} = (x, y)$ denotes the pixels of an image, $\mathbf{v} = (u, v)$ is a nearby point of the neighborhood center $\mathbf{u} = (x, y)$, $I_{\mathbf{v}}$ is defined as an input image, and

$$W_{\mathbf{u}} = \sum_{\mathbf{v} \in \Omega} G_{\sigma_d} G_{\sigma_r}, \quad (8)$$

where

$$G_{\sigma_d} = \exp\left(-\frac{1}{2} \left(\frac{d(\mathbf{u}, \mathbf{v})}{\sigma_d}\right)^2\right), \quad (9)$$

$$G_{\sigma_r} = \exp\left(-\frac{1}{2} \left(\frac{\delta(I(\mathbf{u}), I(\mathbf{v}))}{\sigma_r}\right)^2\right),$$

where $d(\mathbf{u}, \mathbf{v})$ is the Euclidean distance between \mathbf{u} and \mathbf{v} , $\delta(\mathbf{u}, \mathbf{v})$ is a measure of distance between the two intensity values $I(\mathbf{u})$ and $I(\mathbf{v})$, σ_d denotes the domain parameter to control Gaussian shape in space, and σ_r is a range parameter to control influence of intensity change ($\sigma_r = 0.04$ in this paper).

3.4. The Laplacian-of-Bilateral Scale Space Construction. Standard SIFT algorithm detects candidate points using scale space extrema in the difference-of-Gaussian filter convolved with the image, $D(x, y, \sigma)$, described in (3). However, as discussed in Section 2, the major computational complexity occurs in filtering the image in extracting features. In order to detect extrema at s scales of each octave, there must be $s + 3$ Bilateral filter responses in Bilateral scale space. More specifically, it has been shown by Tomasi that the Bilateral filter is twice as expensive as a nonseparable domain filter such as Gaussian filter of the same size in terms of computational cost. If we adopt the method analogous to SIFT algorithm, shown in (3), to produce difference-of-Bilateral filter, the time cost will increase highly.

To considerably reduce the computation complexity and efficiently detect candidate locations in scale space, we proposed producing difference-of-Bilateral filter, $\text{LoB}(\mathbf{u}, \sigma_d)$, defined as follows:

$$\text{LoB}(\mathbf{u}, \sigma_d) = L(\mathbf{u}) \left(\frac{1}{W_{\mathbf{u}}} \sum_{\mathbf{v} \in \Omega} G_{\sigma_d} G_{\sigma_r} \right), \quad (10)$$

where $L(\mathbf{u})$ is a fixed window Laplacian filter and $W_{\mathbf{u}}$, G_{σ_d} , and G_{σ_r} are the same as parameters described in Section 3.3. And $L(\mathbf{u})$ adopted in this paper is the one shown in Figure 2(a). So we can use scale space extrema in the Laplacian-of-Bilateral (LoB) filter convolved with the image, $\text{LB}(\mathbf{u}, \sigma_d)$, which can be computed from the convolution of the fixed window Laplacian filter, $L(\mathbf{u})$, with each Bilateral image, $B(\mathbf{u}, \sigma_d)$:

$$\text{LB}(\mathbf{u}, \sigma_d) = L(\mathbf{u}) * B(\mathbf{u}, \sigma_d). \quad (11)$$

Figure 3 shows the construction of Laplacian-of-Bilateral (LoB) scale space, $\text{LB}(\mathbf{u}, \sigma_d)$, to detect local extrema in our proposed approach.

Apparently for the sake of detecting extrema at n scales of each octave, the Bilateral scale space consists of just $n + 2$ intervals images shown in Figure 3 in our proposed method.

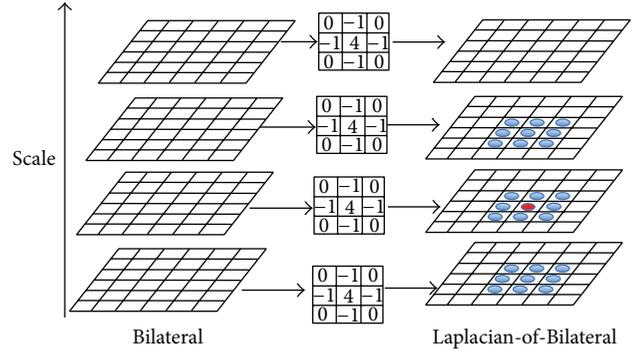


FIGURE 3: The process of local extrema detection based on Laplacian-of-Bilateral (LoB).

TABLE 1: Time cost comparison of LoB and DoB scale space construction.

Scale space	LoB	DoB
Times (ms)	3630	4385

However, if we construct Difference-of-Bilateral (DoB) scale space by simple image subtraction after forming Bilateral scale space rather than using the fixed window Laplacian filter, $n + 3$ intervals images have to be included in Bilateral scale space causing a heavy computation burden. In order to compare the time cost of LoB and DoB scale space construction, we conducted experiments on the ‘‘Graffiti’’ and ‘‘Leuven’’ sequences of Mikolajczyk dataset as described in Section 4.1. Table 1 shows the comparison of execution time in terms of LoB and DoB scale space construction. In this experiment, there are five parameters, including the octave (O), the scale per octave (S), the half-width of Bilateral filter (N), the domain parameter (σ_d), and the range parameter (σ_r). The related parameters, O , S , N , and σ_r , are fixed as 4, 3, 2, and 0.04, respectively. In addition, we select the value of domain parameter σ_d similar to the smoothing parameter (e.g., sigma of the Gaussian kernel) in SIFT algorithm proposed by Lowe because it controls Gaussian shape in space.

The different stages of the proposed interest point detection algorithm are graphically summarized in Algorithm 1.

4. Experiments and Results

We evaluate the results of our approach on the stand dataset commonly used to evaluate the detectors by most vision researchers.

4.1. Dataset. We evaluate the results of our approach on the stand dataset commonly used by most vision researchers. Figure 4 shows example images of Mikolajczyk dataset [24] used for the evaluation obtainable from Visual Geometry Group of Oxford University. This dataset contains 8 image sequences and 48 images in total. It also covers six different geometric and photometric transformations for different scene types, including viewpoint change, scale change, image rotation, image blur, illumination change, and JPEG

- (1) To detect extrema at n scales of each octave, compute Bilateral response $B(\mathbf{u}, \sigma_d)$ for all scales $\sigma_{d_0}, \dots, \sigma_{d_{n+1}}$, using (7) with $\sigma_r = 0.04$.
- (2) Compute the Laplacian-of-Bilateral response $LB(\mathbf{u}, \sigma_d)$ by using (11).
- (3) Extract the maximal/minimal point in scale-space of the $LB(\mathbf{u}, \sigma_d)$ response in a $3 \times 3 \times 3$ window as shown in Figure 3.
- (4) Remove edge responses and low contrast points using Hessian matrix similar to SIFT algorithm.
- (5) Assign orientation and compute the histogram of magnitude of first order gradients over different orientations resulting in a feature descriptor vector of dimension 128 analogous to SIFT algorithm.
- (6) Store the feature location, scale, orientation, and feature descriptor.

ALGORITHM 1: Proposed interest point detection algorithm.

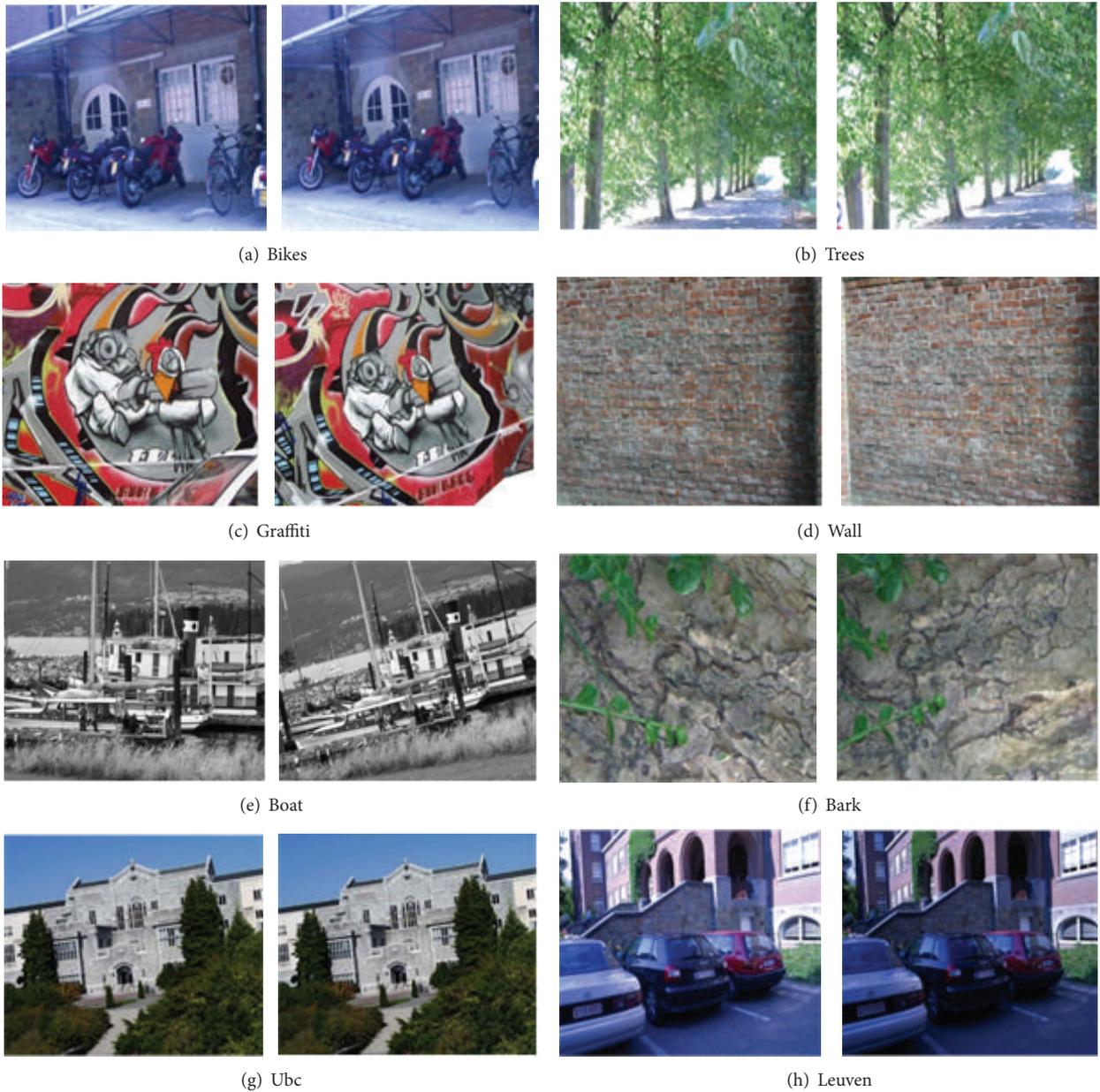


FIGURE 4: Dataset. Examples of images used for the evaluation: (a) and (b) image blur, (c) and (d) viewpoint change, (e) and (f) zoom + rotation changes, (g) JPEG compression, and (h) illumination change.

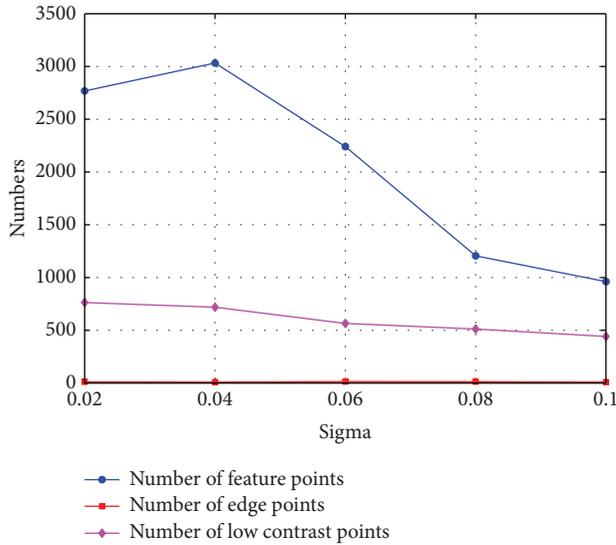


FIGURE 5: Parameters evaluation of the proposed detection algorithm.

compression. Meanwhile, this dataset has the ground-truth matches through estimated homography.

4.2. Repeatability. The repeatability criterion described in [32, 33] is calculated as the ratio between the corresponding interest points and the minimum total number of interest points detected in both images. By the repeatability, we evaluate whether the same physical location in the image (i.e., localization in space) under different viewing conditions is detected with the interest point detection algorithm and whether the detected scale in each view overlaps over identical image surfaces around the feature regions (i.e., localization in scale) in both images, making use of ground-truth data. The features are considered repeated if they are detected within $\epsilon_{loc} = \|\mathbf{x}_1 - \mathbf{H}\mathbf{x}_2\| < 1.5$ pixels and the scale overlap error $\epsilon_s < 40\%$ in our experiments, where \mathbf{x}_1 and \mathbf{x}_2 are feature points and \mathbf{H} is the ground-truth homography. The number of correspondences is defined as the number of pairs of regions if the distance between their descriptors is below an overlap error [24, 34], where we use $\epsilon_s = 40\%$ in our experiments.

4.3. Experiment Setup

4.3.1. Parameter Selection. In this section, we performed simulations to select the appropriate parameter demonstrated in Section 3. There are three parameters in the proposed LoB algorithm: the domain parameter σ_d , the range filter parameter σ_r , and the half-width of Bilateral filter N . The domain parameter σ_d is to control Gaussian shape in space, so we select its value similar to the smoothing parameter (e.g., sigma of the Gaussian kernel) in SIFT algorithm proposed by Lowe. In order to evaluate the influence of various values of the range filter parameter on the proposed interest point detection algorithm, we have conducted experiments on 48 images of benchmark dataset as described in Section 4.1. And the half-width of Bilateral filter, N , is fixed

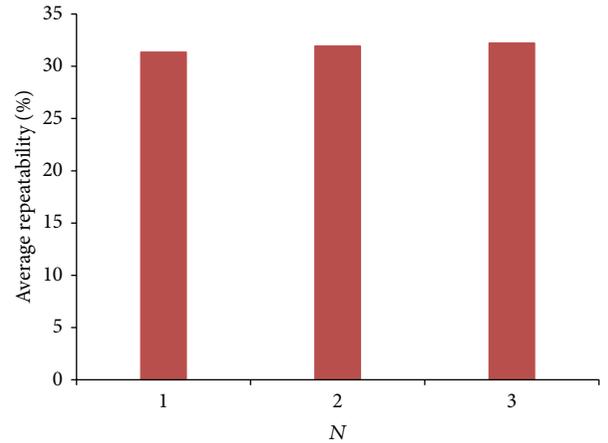


FIGURE 6: The performance of the proposed algorithm with different parameter settings.

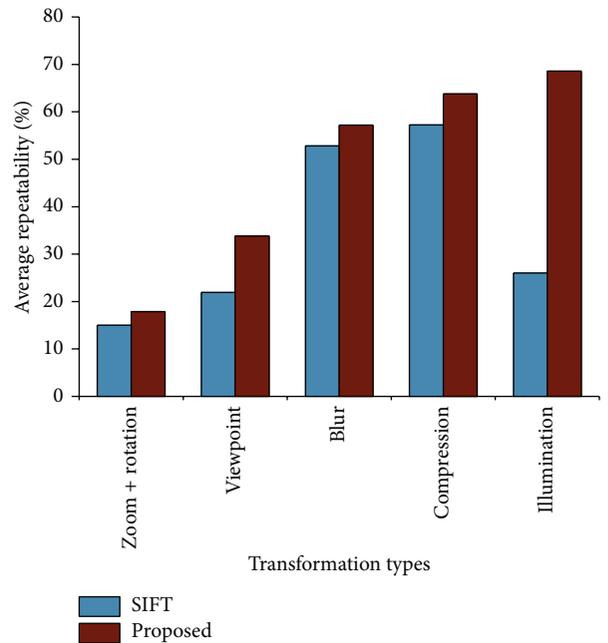


FIGURE 7: Average repeatability score of our proposed approach compared with the SIFT.

as 2 in this experiment. The performance evaluations varying σ_r (0.02, 0.04, 0.06, 0.08, 0.1, and 0.2) are shown in Figure 5. It is clear that the number of edge points eliminated is not sensitive to changes of the value of σ_r . And the number of feature points and low contrast points eliminated gradually decreased with the increase of the value of σ_r . Table 2 shows the comparison of execution time in terms of LoB scale space construction (for LoB scale space $O = 4$ octaves, $S = 3$ scales per octave, and $N = 2$). From the experimental results given in Figure 5 and Table 2, setting the value of σ_r to 0.04, the construction of Laplacian-of-Bilateral (LoB) scale space is at a dramatically lower computational cost when relatively high numbers of feature points are detected.

Furthermore, we conducted experiments on the ‘‘Graffiti’’ sequences of Mikolajczyk dataset for evaluating the effects of

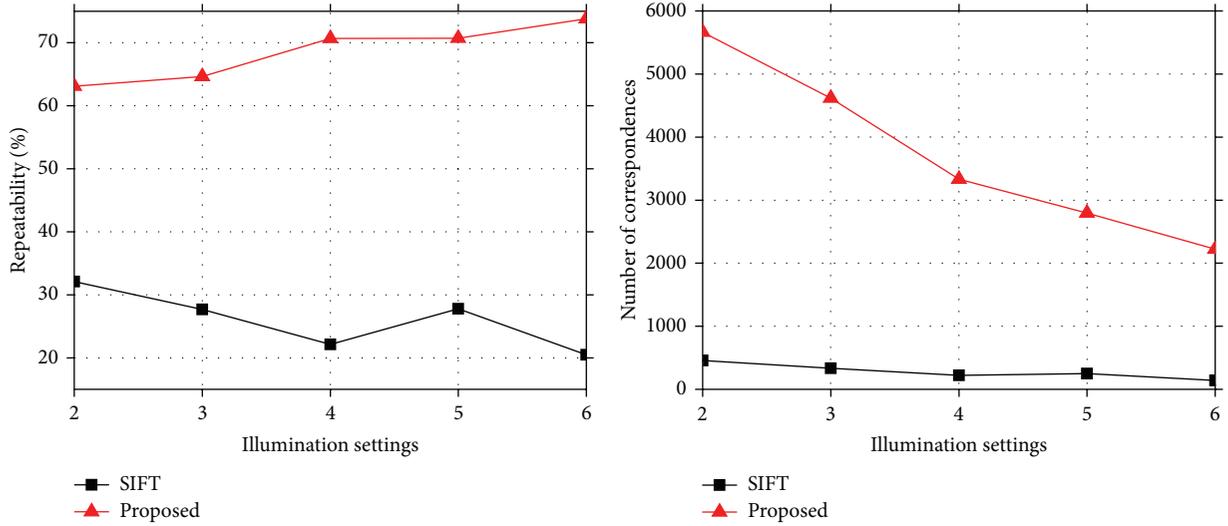


FIGURE 8: Repeatability score and number of correspondences for the illumination variation (Leuven).

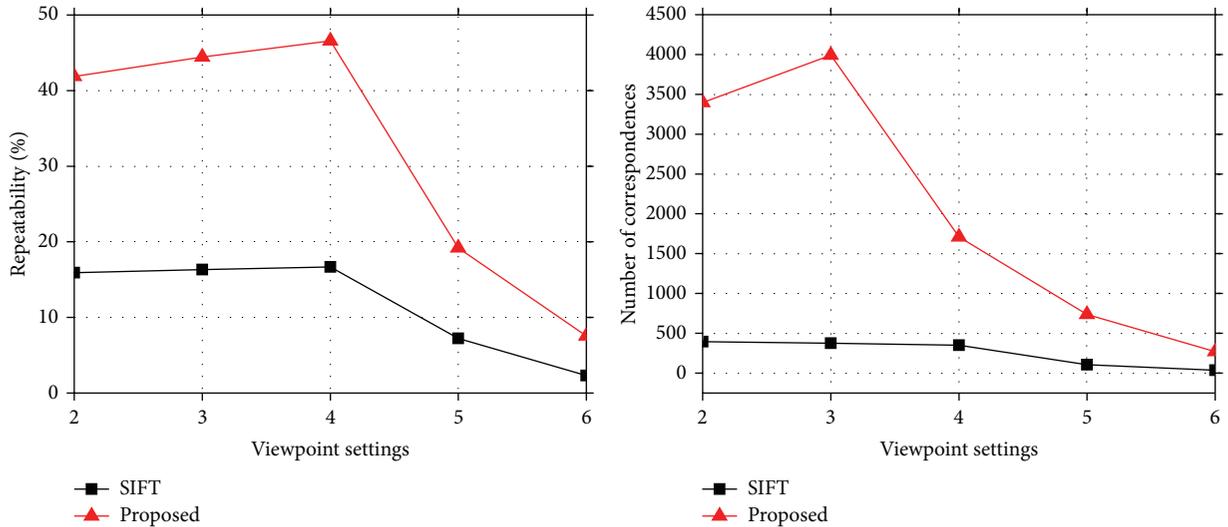


FIGURE 9: Repeatability score and number of correspondences for the viewpoint variation (Graffiti).

TABLE 2: Time cost comparison of LoB scale space construction for different values of σ_r , 2.40 GHz, 12 GB RAM CPU.

σ_r	0.02	0.04	0.06	0.08	0.1	0.2
Times (ms)	3906	3750	3838	3755	3744	3500

the Bilateral filter half-width, N , on our proposed detection algorithm. We tried three sizes: 1, 2, and 3. And the corresponding Bilateral filter window sizes are 3×3 , 5×5 , and 7×7 . The results are shown in Figure 6. From Figure 6, we can see that the performance of our proposed algorithm is gradually getting better with the increase of the value of N . However, more filter time will be consumed when the value of N increases. In terms of performance and filter time cost, we use the parameter setting of $N = 2$.

4.3.2. Repeatability Evaluation. In this section, we evaluated the repeatability and the number of correspondences to

measure the quality of our proposed interest point detection algorithm and SIFT. For each image sequence, its first image is deemed as a reference image, and other images conjuncted as the reference image construct some image pairs. Repeatability score is computed based on these image pairs.

For the sake of comparing the overall performance, we computed an average repeatability score over image pairs of this sequence. Figure 7 illustrates average repeatability score of our proposed approach and SIFT. The experimental results illustrate that our proposed interest point detection algorithm performs better than SIFT with respect to all types of imaging conditions. In the figure, the repeatability increases from left to right, which indicates the difficulties of different geometry and photometric changes. Zoom and rotation changes are apparently the most difficult ones for our work.

In Figure 7, we tend to conclude that when images exhibit illumination or viewpoint change, our proposed approach is better, especially when encountering large illumination

change. The repeatability and the number of feature correspondences evaluations on the images exhibiting illumination and viewpoint changes can be seen from Figures 8 and 9.

5. Conclusions

In this paper, a new Laplacian-of-Bilateral (LoB) filter is proposed to improve interest point detection. We provided the theoretical background and experimental results showing that the proposed interest point detection algorithm achieves better performance. In terms of interest point detection algorithm quality, our approach substantially improves the repeatability of detected interest point and the number of correspondences. Additionally, our proposed approach is found to be more robust to images exhibiting illumination and viewpoint changes, especially when encountering large illumination change. The proposed LoB filter is general and can be flexibly extended to other based on DoG filter interest point detection algorithms. This is one of our future work directions.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (Grant nos. 61170116, 61375010, 61300075, and 61472031), Beijing Higher Education Young Elite Teacher Project (Grant no. YETP0375), and the Fundamental Research Funds for the Central Universities under the Grant no. FRF-TP-14-120A2. The authors would like to express their sincere appreciation to the anonymous reviewers for their insightful comments, which greatly helped them to improve the quality of the paper.

References

- [1] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, IEEE, Minneapolis, Minn, USA, June 2007.
- [2] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proceedings of the 8th International Conference on Computer Vision*, vol. 1, pp. 525–531, July 2001.
- [3] J. Li, N. Allinson, D. Tao, and X. Li, "Multitraining support vector machine for image retrieval," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3597–3601, 2006.
- [4] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1088–1099, 2006.
- [5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [6] V. Ferrari, T. Tuytelaars, and L. van Gool, "Simultaneous object recognition and segmentation by image exploration," in *Computer Vision—ECCV 2004*, pp. 40–54, Springer, Berlin, Germany, 2004.
- [7] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 26–33, June 2005.
- [8] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.
- [9] T. Tuytelaars and L. van Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.
- [10] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "how do I organize my holiday snaps?," in *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part I*, vol. 2350 of *Lecture Notes in Computer Science*, pp. 414–431, Springer, Berlin, Germany, 2002.
- [11] D. Tell and S. Carlsson, "Wide baseline point matching using affine invariants computed from intensity profiles," in *Computer Vision—ECCV 2000*, vol. 1842 of *Lecture Notes in Computer Science*, pp. 814–828, Springer, Berlin, Germany, 2000.
- [12] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, 2005.
- [13] S. Se, D. Lowe, and J. Little, "Global localization using distinctive visual features," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 226–231, IEEE, 2002.
- [14] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.
- [15] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II/264–II/271, IEEE, June 2003.
- [16] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1792–1799, IEEE, October 2005.
- [17] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3d object modeling and recognition using affine-invariant patches and multi-view spatial constraints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-272–II-277, June 2003.
- [18] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.
- [19] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Computer Vision—ECCV 2002*, vol. 2350 of *Lecture Notes in Computer Science*, pp. 128–142, Springer, Berlin, Germany, 2002.
- [20] F. Schaffalitzky and A. Zisserman, "Viewpoint invariant texture matching and wide baseline stereo," in *Proceedings of the 8th International Conference on Computer Vision*, vol. 2, pp. 636–643, July 2001.

- [21] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [22] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.
- [23] P. Mainali, G. Lafruit, Q. Yang, B. Geelen, L. V. Gool, and R. Lauwereins, "SIFER: scale-invariant feature detector with error resilience," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 172–197, 2013.
- [24] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [25] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. II506–II513, IEEE, July 2004.
- [26] S. A. J. Winder and M. Brown, "Learning local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
- [27] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: a SIFT descriptor with color invariant characteristics," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 2, pp. 1978–1983, IEEE, June 2006.
- [28] A. Verma, C. Liu, and J. Jia, "New colour SIFT descriptors for image classification with applications to biometrics," *International Journal of Biometrics*, vol. 3, no. 1, pp. 56–75, 2011.
- [29] B. Li, R. Xiao, Z. Li, R. Cai, B.-L. Lu, and L. Zhang, "Rank-SIFT: learning to rank repeatable local interest points," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1737–1744, Providence, RI, USA, June 2011.
- [30] K. Liao, G. Liu, and Y. Hui, "An improvement to the SIFT descriptor for image representation and matching," *Pattern Recognition Letters*, vol. 34, no. 11, pp. 1211–1220, 2013.
- [31] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE 6th International Conference on Computer Vision*, pp. 839–846, IEEE, January 1998.
- [32] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.
- [33] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [34] F. Tang, S. H. Lim, N. L. Chang, and H. Tao, "A novel feature descriptor invariant to complex brightness changes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 2631–2638, IEEE, June 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

