IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 29, 2019, accepted December 20, 2019, date of publication January 9, 2020, date of current version January 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2965169

Keypoint Detection Using Higher Order Laplacian of Gaussian

YONGJU CHO[1,2], DOJIN KIM[3], SALEH SAEED[3], MUHAMMAD UMER KAKLI[2], SOON-HEUNG JUNG[1], JEONGIL SEO[1], AND UNSANG PARK[3]

[1]Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea
[2]University of Science and Technology, Daejeon 34113, South Korea
[3]Department of Computer Science and Engineering, Sogang University, Seoul 04107, South Korea

Corresponding author: Unsang Park (unsangpark@sogang.ac.kr)

This work was supported by the Korea Government (MSIT) (development of 3D spatial media core technology) under Grant 19ZR1120.

ABSTRACT This paper presents a keypoint detection method based on the Laplacian of Gaussian (LoG). In contrast to the Difference of Gaussian (DoG)-based keypoint detection method used in Scale Invariant Feature Transform (SIFT), we focus on the LoG operator and its higher order derivatives. We provide mathematical analogies between higher order DoG (HDoG) and higher order LoG (HLoG) and experimental results to show the effectiveness of the proposed HLoG-based keypoint detection method. The performance of the HLoG is evaluated with four different tests: i) a repeatability test of the keypoints detected across images under various transformations, ii) image retrieval, iii) panorama stitching and iv) 3D reconstruction. The proposed HLoG method provides comparable performance to HDoG and the combination of HLoG and HDoG provides significant improvements in various keypoint-related computer vision problems.

INDEX TERMS SIFT, DoG, LoG, higher order DoG, higher order LoG.

I. INTRODUCTION

Keypoint based image representation methods have shown their effectiveness for image search, registration and 2D to 3D reconstruction tasks [1]–[4]. One of the most successful and well-known keypoint-based image search approaches is the Scale Invariant Feature Transform (SIFT). SIFT detects keypoints using scale-space derivatives, a technique known as the Difference of Gaussian (DoG). The detected keypoints are represented as descriptors utilizing the combination of gradient orientation and magnitude around each keypoint. Images can then be compared based on their keypoints; a larger number of similar keypoints increases the similarity of the images being compared. If a query image is prepared to tightly contain a single object inside, the image search task can be effectively delegated as a known object search task.

Since the successful introduction of SIFT, many similar approaches have been developed to improve the robustness of keypoint detection [2], [5], [6], [15]–[18]. Among those, KAZE [15] detects features in a non-linear scale-space built by the Additive Operator Splitting (AOS) technique, WADE [17] finds repeatable keypoints based on salient symmetries at various scales by using Partial Differential Equations (PDEs), and MSD [18] uses a relatively wide range of image patches that are very dissimilar to detect keypoints.

On the other hand, many deep learning-based keypoint detection methods have been introduced to replace traditional hand-crafted keypoint detection methods over the past few years. TILDE [21] is a learning scheme for reliable detection of keypoints with drastic variations of weather and lighting conditions. It proposed an effective method of generating training data to learn a regressor. LIFT [10] is a deep neural network architecture that includes a feature detector, orientation estimator and descriptor constructor. It uses the Siamese architecture with four branches that takes a quadruplet of patches (i.e., base patch, positive patch, negative patch, and a patch that contains no distinctive feature point) as inputs to train keypoint detector and descriptor simultaneously. Covariant Feature Detectors [20] is a generic framework for detecting local viewpoint invariant features in images by casting detection as a regression problem. It can learn corner and orientation detectors by using a covariance constraint. Quad-networks [23] is a framework for training a neural network to rank transformation invariant interest points and only the top/bottom quantiles of the ranks are chosen as keypoints. SuperPoint [11] is a self-supervised framework for training

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

a feature detector and descriptor constructor. This network computes pixel-level feature locations and their associated descriptors on images in a single forward pass. It uses the domain adaption framework called Homographic Adaption and obtained state-of-the-art performance on HPatches when compared to LIFT, SIFT and ORB. LF-Net [12] is another deep architecture used to learn local features that contains the entire feature extraction pipeline and does not require a secondary hand-crafted detector to generate training data. Given an image pair, LF-Net extracts feature points from the first image, then it runs on the second image to produce a clean response map with sharp maxima at the right locations. Key.Net [22] is a shallow multi-scale architecture for keypoint detection that uses a combination of handcrafted and learned CNN filters. They designed a loss function to maximize the repeatability score and detect robust features across a range of scales. It is trained with synthetic data created from ImageNet and evaluated on an HPatches benchmark dataset. It outperformed the state-of-the-art detectors in terms of repeatability, matching performance, and complexity. R2D2 [24] is a framework for jointly learning keypoint detection and descriptor construction. It avoids ambiguous areas and increases the reliability of keypoint detections and descriptions. It outperformed the state-of-the-art detectors and descriptors on an HPatches dataset.

Even though the research trend in keypoint detection and descriptor construction is moving toward deep learning-based methods, traditional keypoint detection methods (e.g., SIFT) are still an important topic to study. Traditional keypoint detection methods still perform better than the deep learning-based methods, especially when there are large viewpoint or scale changes. Also, some of the traditional keypoint detectors are used to generate keypoint information that can be used to train neural networks. Some deep learning architectures try to mimic the scale-space operation of SIFT. Therefore, understanding and improving the performance of traditional keypoint detection methods can be beneficial to the study of deep learning-based keypoint detection.

In this paper, we propose an alternative method to SIFT or higher order Difference of Gaussian (HDoG) [7], which shows comparable performance to that of SIFT or HDoG. The proposed method is based on higher order spatial derivatives, which can be called higher order Laplacian of Gaussian (HLoG). Even though the proposed HLoG shows a close analogy with HDoG, the proposed HLoG scheme is significantly different from HDoG due to the difference in the base operators (DoG versus LoG). We briefly review the keypoint detection methods of SIFT and HDoG and explain the proposed method in the following sections.

II. PROPOSED METHOD

We first review the basic keypoint detection methods of DoG and LoG, and then the extended methods of DoG (HDoG), and the proposed method, which is the extended LoG or Higher Order LoG (HLoG). The analogy between HDoG and HLoG is shown using their mathematical derivations.

A. DIFFERENCE OF GAUSSIAN AND LAPLACIAN OF GAUSSIAN

The scale-space analysis for keypoint detection was extensively studied in [8] where it was proposed that the scale space extrema corresponds to local keypoints. However, the first practical application of the scale-space analysis-based keypoints for image processing tasks was introduced only after the invention of SIFT in which the DoG was used instead of the LoG. It is well known that DoG is an approximation of LoG as in Eq. (1):

$$\partial G/\partial\sigma \approx \sigma\nabla^2 G \tag{1}$$

We believe that the success of SIFT was induced due to the use of DoG as well as the subsequent operations, such as local extrema detection, keypoint removal on edge, and robust descriptor construction method. However, DoG or LoG-based keypoint detection is still the most important step in the entire keypoint detection process in SIFT. HDoG was recently introduced with superior performance to conventional DoG in terms of image retrieval [7].

B. HDOG AND HLOG

HDoG [7] was developed based on the observation that DoG only uses the first order derivative of scale-space (G) with respect to scale (σ). HDoG uses the 2nd, 3rd and higher order scale-space derivatives with respect to σ. Given

$$G(x, y) = exp(-(x^2 + y^2)/2\sigma^2) \tag{2}$$

as the Gaussian equation by disregarding the leading coefficient, HDoG up to the 4th degree can be expressed as Eqs. (3)-(6) by successively differentiating Eq. (2) with respect to σ as:

$$\partial G/\partial\sigma = (x^2 + y^2)/\sigma^3 \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{3}$$

$$\partial^2 G/\partial\sigma^2 = (x^2 + y^2)(x^2 + y^2 - 3\sigma^2)/\sigma^6 \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{4}$$

$$\partial^3 G/\partial\sigma^3 = (x^2 + y^2)(x^4 + y^4 + 2x^2y^2 - 9x^2\sigma^2 - 9y^2\sigma^2 + 12\sigma^4)/\sigma^9 \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{5}$$

$$\partial^4 G/\partial\sigma^4 = (x^2 + y^2)(x^6 + y^6 + 3x^4y^2 + 3x^2y^4 - 18x^4\sigma^2 - 18y^4\sigma^2 - 36x^2y^2\sigma^2 + 75x^2\sigma^4 + 75y^2\sigma^4 - 60\sigma^6)/\sigma^{12} \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{6}$$

On the other hand, if we take the derivatives of Eq. (2) with respect to x and y up to the 8th degree for even orders, we have Eqs. (7)-(10).

$$\nabla^2 G = \partial^2 G/\partial x^2 + \partial^2 G/\partial y^2 = (x^2 + y^2 - 2\sigma^2)/\sigma^4 \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{7}$$

$$\nabla^4 G = \partial^4 G/\partial x^4 + \partial^4 G/\partial y^4 = (x^4 + y^4 - 6x^2\sigma^2 - 6y^2\sigma^2 + 6\sigma^4)/\sigma^8 \cdot exp(-(x^2 + y^2)/2\sigma^2) \tag{8}$$

$$\nabla^6 G = \partial^6 G/\partial x^6 + \partial^6 G/\partial y^6 = (x^6 + y^6 - 15x^4\sigma^2 - 15y^4\sigma^2 + 45x^2\sigma^4 + 45y^2\sigma^4 - 30\sigma^6)/\sigma^{12} \cdot exp(-(x^2+y^2)/2\sigma^2) \quad (9)$$

$$\nabla^8 G = \partial^8 G/\partial x^8 + \partial^8 G/\partial y^8 = (x^8 + y^8 - 28x^6\sigma^2 - 28y^6\sigma^2 + 210x^4\sigma^4 + 210y^4\sigma^4 - 420x^2\sigma^6 - 420y^2\sigma^6 + 210\sigma^8)/\sigma^{16} \cdot exp(-(x^2+y^2)/2\sigma^2) \quad (10)$$

By comparing Eq. (3) and Eq. (7), we can verify that Eq. (1) is satisfied. Similarly, by comparing Eqs. (4), (5), and (6) with Eqs. (8), (9), and (10), respectively, we can find the relationships between the higher order scale-space derivatives and the spatial derivatives as Eq. (11)-(13):

$$\partial^2 G/\partial\sigma^2 \approx \sigma^2\nabla^4 G \quad (11)$$

$$\partial^3 G/\partial\sigma^3 \approx \sigma^3\nabla^6 G \quad (12)$$

$$\partial^4 G/\partial\sigma^4 \approx \sigma^4\nabla^8 G \quad (13)$$

This means that the σ-normalized HDoG approximates HLoG as DoG does LoG. Since the effectiveness of HDoG was already proven experimentally in [7], we can expect that HLoG can also be useful for keypoint detection. The oblique and top view of the graphical representation of HLoG filters of the 2nd, 4th, 6th and 8th orders are shown in Fig. 1. As the order increases, the shape of the filters become more complicated to better catch high frequency signals. Since the HDoG requires inter-scale subtraction operations that need to be performed sequentially to obtain higher order DoG spaces, it has an inherent drawback in parallel processing. However, HLoG filter kernels can be prepared beforehand and applied to the input image in parallel to obtain higher order HLoG spaces all at the same time. Therefore, HLoG provides much better parallel processing capability compared to HDoG.

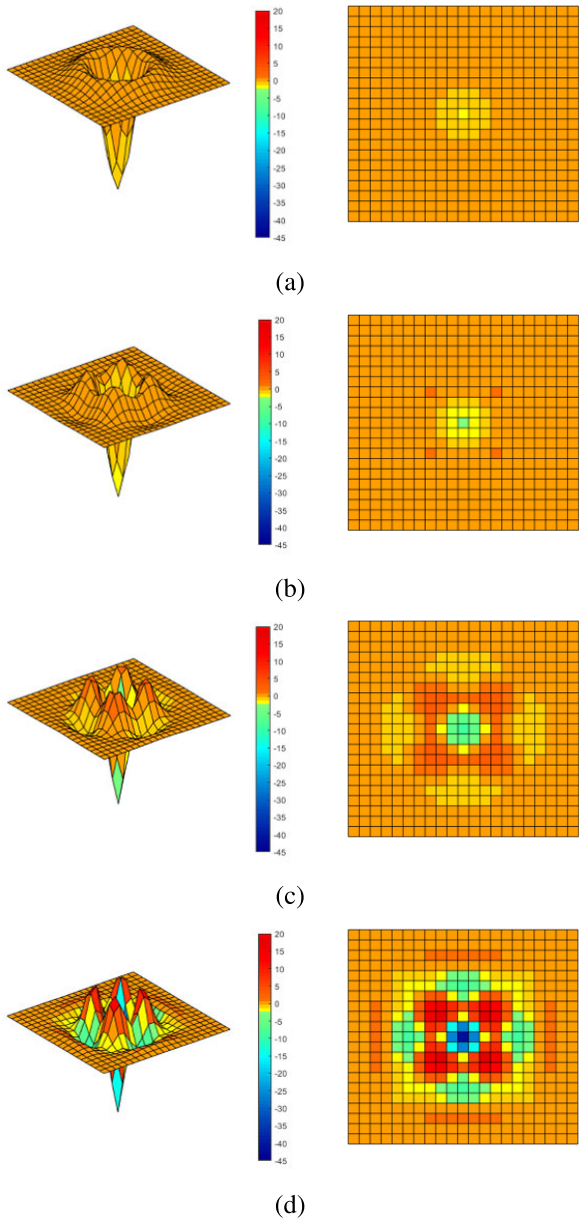


FIGURE 1. Oblique and top views of the HLoG filters for (a) 2nd, (b) 4th, (c) 6th, and (d) 8th orders.

III. EXPERIMENTAL RESULTS

We evaluated the proposed method based on four different tests: i) the repeatability of keypoint detection on duplicate images with various transformations, ii) image retrieval, iii) panorama stitching and iv) 2D to 3D reconstruction test. Detailed information on the databases and evaluation methodologies are presented in the following sections.

A. DATABASE

We used the Mikolajczyk database [9] for the keypoint's repeatability test, a subset of the Tattoo image data set obtained from Michigan State Forensics Laboratory for the image retrieval test, four sets of images with overlapping regions captured in our lab for the panorama stitching test, and a 2D image set for 3D reconstruction [14]. The Mikolajczyk database provides eight sets of images with five different image variations, that is, *blur* (two sets), *viewpoint* (two sets), *zoom+rotation* (two sets), *light* (one set), and *JPEG compression* (one set). Each set contains six images and five homographic transformation matrices. The homography matrices provide ground truth with which correspondences of keypoints across transformed images can be decided. The homography matrices are generated from the first image to all five other images in each of the eight sets. In total, there are 48 images and 40 homography matrices. We consider that a keypoint detected in the original image is repeatedly detected in the image after homography transformation, if the distance between two keypoints after canceling the homographic transformation is less than one pixel.

We used the same Tattoo image data set used in [7], where there are 1,000 images with 445 unique tattoos. Therefore, there are about 2.25 images for each tattoo. We performed the image retrieval test in leave-one-image-out fashion as in [7] and compared the performance between HDoG and HLoG. For the panoramic stitching test, we observed which of HDoG or HLoG succeeds more in stitching images that were

collected in our lab. Typically, a keypoint detector detecting keypoints with higher repeatability has higher accuracy in image retrieval and a greater chance to successfully stitch images with overlapping regions. For the 3D reconstruction test, we used a well-known 2D image sequence data set capturing a Hall [14]. The 3D reconstruction operation can generate a greater number of 3D points if it can find a greater number of matching keypoints across 2D images. The more 3D points we have, the more detailed information about the reconstructed 3D object we can generate. The repeatability test is done to evaluate the performance of the keypoint detectors directly, and the other three tests (i.e., image retrieval, panoramic stitching 3D reconstruction) evaluate performance on an application level. In all three major application problems used in the evaluation, we can say that a larger number of keypoints provides better performance.

B. PERFORMANCE EVALUATION

1) KEYPOINT REPEATABILITY TEST

We used both DoG and LoG for keypoint detection up to the 4th and 8th degrees of derivatives, respectively. As shown in Sec. II, the 1st, 2nd, 3rd, and 4th degrees of HDoG correspond to the 2nd, 4th, 6th, and 8th degrees of HLoG, respectively. Even though there is no upper limit on the available degrees of derivatives, we have considered only up to 4th order since we can obtain enough keypoints in such a configuration. Higher orders beyond 4 or 8 may not capture normally observable signals in images and may increase the computational time without much gain. The descriptor construction and matching scheme follow the conventional methods in SIFT [1]. For those matching keypoints, we verified whether they conform with the homographic relationship within a distance of less than one pixel. The number of ''genuine'' matching keypoints out of the total keypoints were then used as the measure of the repeatability of the keypoint detectors, HDoG or HLoG. Fig. 2 shows example keypoint detection results with an image from the Mikolajczyk database by HDoG and HLoG at various orders of derivatives. We can observe the following points in Fig. 2.

- The keypoints of HLoG appear at similar locations as those of HDoG at the first order (i.e., DoG and LoG).
- The number of keypoints of HDoG decrease as the order of derivative increases, whereas the keypoints of HLoG remain at similar numbers. The numbers of keypoints for HDoG are 1,260, 865, 224, and 99 at 1st, 2nd, 3rd, and 4th degrees, respectively. The numbers of keypoints for HLoG are 1,165, 1,419, 821, and 1,220 at 2nd, 4th, 6th, and 8th degrees, respectively.

  In general, the number of keypoints are supposed to decrease at higher derivative orders because the high frequency components are fewer than the low frequency components in regular images. However, we believe that the HLoG filters shown in Fig. 1 somehow interact with the high frequency components sensitively to generate a large number of keypoints. We will evaluate the effectiveness of the HLoG keypoints in various experimental setups later in this paper.
- Overall, a substantial number of keypoints were differently detected between HLoG and HDoG.

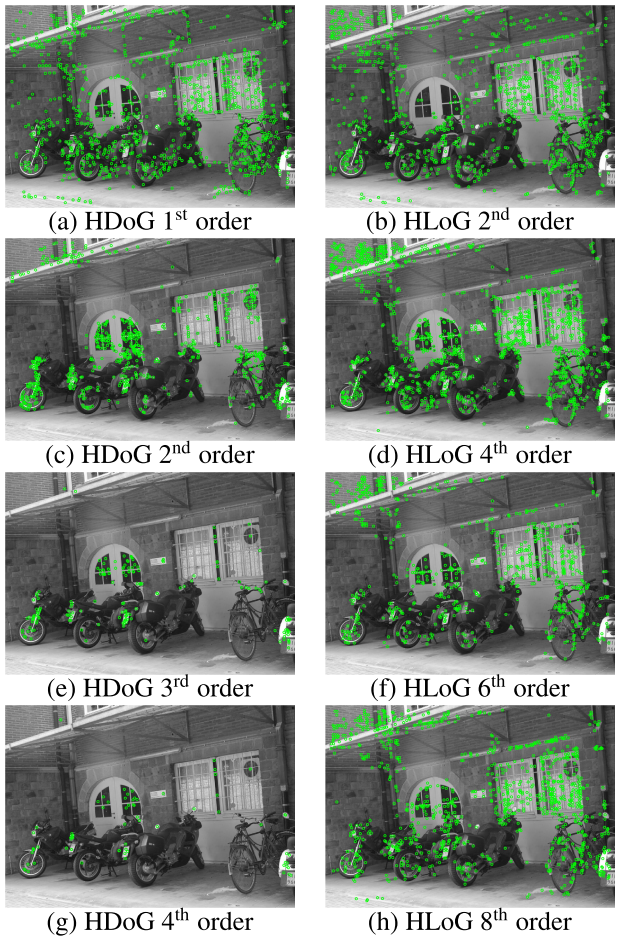


FIGURE 2. Example keypoint detection results using HDoG and HLoG at four different orders.

Experimental results of the repeatability tests are summarized in Fig. 3, where the keypoint repeatabilities are evaluated in five different image transformations as well as their average. The x and y axes of the figures in Fig. 3 are the average number of keypoints for each derivation order and number of matching keypoints, respectively. Note that we used an average number of keypoints for each derivation order in Fig. 3 because we plotted the graphs of the fusions of multiple derivation orders together with those graphs from each derivation order. We also showed the fusion of HDoG and HLoG in the same graph with each derivation order. We limited the number of detected keypoints according to the tic values of the x axis in Fig. 3; the keypoints are selected by the strength of the keypoints in descending order. The strength of a keypoint is decided by the value of a keypoint in DoG or LoG space. The number of matching keypoints is proportional to the total number of detected keypoints. Therefore, we showed the number of matching keypoints

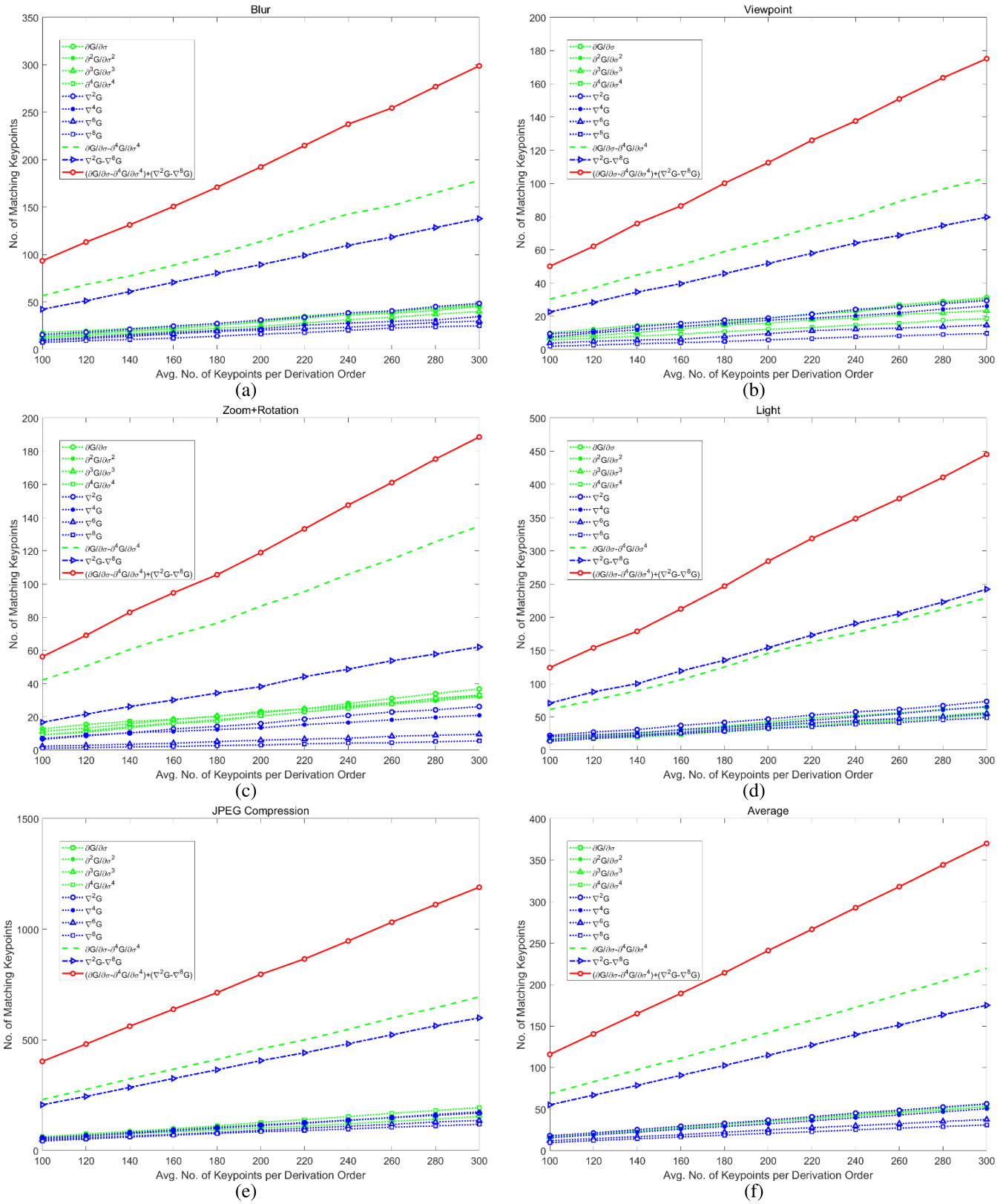


FIGURE 3. Comparisons of the repeatabilities of HDoG and HLoG keypoints with respect to various image transformations. (a) blur, (b) viewpoint, (c) zoom+rotation, (d) light, (e) JPEG compression, and (f) average.

with respect to the number of detected keypoints up to 300; a keypoint detector that generates more matching keypoints out of the same number of keypoints can be considered to be more robust. Transformations of *blur*, *viewpoint*, and *zoom+rotation* contain two sets of images in each. Therefore, we provided the average number of matching keypoints for those categories (Fig. 3(a), (b), and (c)).

This keypoint repeatability test is comparable to the image matching test performed in [7]. However, we used the number of matching keypoints across images under homographic transformation as the measure of the performance of a keypoint detector rather than image matching accuracy to focus more on the robustness of a keypoint detector itself. The image matcher can show different performances depending on the matching scheme even when the same keypoint detector is used. For example, the geometric configuration of the matching keypoints can be used in the image matching process rather than the simple number of matching keypoints. We also provide the image retrieval test results in this paper.

The overall performances of HLoG are shown to be slightly lower than those of HDoG, except for *Light* (Fig. 3(d)). The performance of HLoG is lower than that of HDoG with a large margin for *zoom+rotation* variation. The performance of HLoG is slightly better than that of HDoG in the *light* variation. In accordance with [7], the HDoG using multiple derivation orders (i.e., 1^{st} to 4^{th} degrees) significantly increases the number of matching keypoints compared to those of DoG (i.e., 1^{st} order derivative). Similar behavior is also seen in HLoG. Moreover, the fusion of HDoG and HLoG with multiple derivation orders significantly increases the number of total matching keypoints. Therefore, we can expect that better performance will be achieved in various keypoint-related tasks, for example, image matching, panoramic stitching, or 3D reconstruction, by using the proposed method.

2) IMAGE RETRIEVAL TEST

For the image retrieval test, we followed the same experimental setup as that of [7]. We chose the number of octaves (O), number of scales (S), and cutoff threshold values (Tx) as {3}, {3, 4, 5}, and {0.180, 0.085, 0.070, 0.060}, respectively for the keypoint detection using HLoG. As in [7], the parameter T_x plays an important role in deciding the number of keypoints. The parameter T_x is inversely proportional to the number of keypoints detected. Even though HDoG and HLoG approximate each other, their overall values from HDoG or HLoG pyramid appears significantly different. Therefore, we did not use the same values of T_x's between HDoG and HLoG for the comparison. We used the T_x's of HLoG to control the number of detected keypoints to a level similar to that of HDoG. Consequently, we found the series of T_x's of HLoG as 0.180, 0.085, 0.070, and 0.060, corresponding to the T_x's of HDoG, 0.020, 0.010, 0.007, and 0.005. We followed the same strategies for descriptor construction and matching as those in [7]. Therefore, we provide the image retrieval performance in terms of rank-1 matching accuracies in four different schemes: 2^{nd} order (i.e., traditional LoG),

FIGURE 4. Example matching results: the first, second, and third columns correspond to image queries, false mates, and true mates, respectively. (a)-(c) are examples where HDoG failed, but HLoG succeeded, (d) is an example where HDoG succeeded, but HLoG failed, and (e) and (f) are examples where both the HDoG and HLoG succeeded. The values of parameters O, S, and T_s were chosen as 3, 3, and 0.02 for HDoG and 3, 3, and 0.18 for HLoG, respectively.

diagonal, all, and off-diagonal. The 2^{nd} order scheme considers the matching of the keypoints only from LoG filter. The other schemes, diag., all, or off-diag., differs depending on whether the keypoints are allowed to match across only the same orders, all orders, or only different orders of derivatives. More details of the four different matching schemes are given in [7].

The average number of keypoints and the rank-1 matching accuracies for four different matching schemes at a few different parameter settings for HDoG and HLoG are summarized in Table 1 and 2, respectively. The performance of HDoG is the same as that in [7] and is provided here as

TABLE 1. Summary of the performance of HDOG.

Cutoff D (T_x)	#Octaves (O)	#Scales (S)	Average #keypoints/image[(i)]				Rank-1 accuracy[(ii)] (%)			
			$\partial\sigma$	$\partial\sigma^2$	$\partial\sigma^3$	$\partial\sigma^4$	$\partial\sigma$	Diag.	All	Off-diag.
0.020	3	3	139.08	4.77	0.14	0.03	86.5	**86.8**	**86.7**	1.4
		4	86.47	0.03	0	0	74.3	74.3	74.2	0.1
		5	48.38	0	0	0	59.4	59.4	59.4	0
0.010	3	3	228.05	138.28	24.03	9.66	90.6	**92.9**	**93.3**	25.2
		4	205.91	119.34	0.03	0	90.5	**91.3**	**91.4**	0.3
		5	178.75	0.41	0	0	89.4	89.4	89.3	0.1
0.007	3	3	253.10	327.84	118.75	68.49	87.4	**93.4**	**94.2**	60.6
		4	241.73	71.03	1.30	0.03	91.0	**92.8**	**92.4**	5.3
		5	228.69	7.19	0	0	91.5	**91.7**	**91.8**	0.9
0.005	3	3	272.85	533.43	329.90	242.31	85.0	**94.8**	**94.7**	78.4
		4	266.52	210.80	12.92	0.89	89.9	**93.1**	**93.7**	22.3
		5	260.54	47.36	0.08	0	91.0	**92.6**	92.8	3.0
Mean							85.5	**87.7**	**87.8**	16.5
Max							91.5	**94.8**	**94.7**	78.4

(i) $\partial\sigma^i$ represents extracting keypoints using the i$^{\text{th}}$ order scale-space derivative.

(ii) "$\partial\sigma$" uses keypoints from the 1$^{\text{st}}$ order scale-space derivative, "Diag." uses keypoints from the same order of scale-space derivatives, "All" uses all keypoints, and "Off-diag" allows keypoints only from different orders of derivatives to match. Bold numbers represent HDoG's performances that are better than those of SIFT.

TABLE 2. Summary of the performance of HLOG.

Cutoff D (T_x)	#Octaves (O)	#Scales (S)	Average #keypoints/image[(i)]				Rank-1 accuracy[(ii)] (%)			
			$\partial(x^2y^2)$	$\partial(x^4y^4)$	$\partial(x^6y^6)$	$\partial(x^8y^8)$	$\partial(x^2y^2)$	Diag.	All	Off-diag.
0.180	3	3	142.82	138.62	79.44	111.50	78.2	**86.9***	**89.1***	72.4
		4	125.87	68.57	28.93	23.14	**82.5***	**86.2***	**87.3***	52.8
		5	103.50	27.87	7.17	193.50	**81.0***	**83.8***	**84.7***	36.0
0.085	3	3	234.73	319.48	240.03	303.00	74.2	89.3	**91.1**	85.5
		4	228.79	281.96	180.35	111.39	80.9	89.2	**91.3**	82.0
		5	215.25	222.21	120.27	536.17	80.3	**91.7***	**92.8***	80.6
0.070	3	3	259.52	361.36	279.94	350.79	72.6	**89.9**	**91.6**	85.5
		4	258.75	336.67	231.78	140.27	77.8	88.7	**91.0**	85.2
		5	247.06	278.32	181.82	658.64	79.0	**91.6**	**93.6***	84.6
0.060	3	3	274.81	394.20	309.03	384.05	73.1	**90.2**	**91.5**	**85.5**
		4	282.45	378.94	271.97	162.41	77.4	89.2	**91.5**	85.3
		5	273.72	337.61	237.95	755.63	78.4	**91.2**	**93.2***	87.8
Mean							78.0	**89.0***	**90.7***	76.9
Max							82.5	**91.7**	**93.6**	87.8

(i) $\partial x^i y^i$ represents extracting keypoints using the i$^{\text{th}}$ order spatial derivative.

(ii) "$\partial(x^2y^2)$" uses keypoints from the 2$^{\text{nd}}$ order spatial derivative, "Diag." uses keypoints from the same order of spatial derivatives, "All" uses all keypoints, and "Off-diag" allows keypoints only from different orders of derivatives to match. Bold numbers represent HLoG's performances that are better than those of SIFT. Bold numbers marked with $*$ represent HLoG's performances that are better than those of SIFT and HDoG.

the baseline. Based on the experimental results summarized in Table 1 and 2, we can make the following observations: (i) The best performance of the 1$^{\text{st}}$ order derivative of HDoG (SIFT) and 2$^{\text{nd}}$ order derivative of HLoG (LoG) are 91.5% and 82.5%, respectively, (ii) the use of higher-order derivatives show significantly better matching accuracy compared to the SIFT or LoG in most of the parameter settings (bold numbers), where the best matching accuracies of HDoG and HLG are 94.8% (Diag.) and 93.6% (All), respectively, (iii) the differences between the best matching accuracies between SIFT and HDoG, and LoG and HLoG are 3.3% and 11.1%, respectively, and (iv) there are many cases where the performance of HLoG is better than that of HDoG (marked with *).

Since the superiority of the performances of HDoG and HLoG differ for different parameter settings, we performed

TABLE 3. Summary of the performance of the fusion of HDoG and HLOG.

Cutoff D (T_x, HDoG/HLoG)	#Octaves (O)	#Scales (S)	Rank-1 match accuracy (%)						
			SIFT	HDoG		HLoG		HDoG + HLoG	
				Diag.	All	Diag.	All	Diag.	All
0.020/0.180	3	3	86.5	**86.8**	**86.7**	**86.9***	**89.1***	**87.0***	**89.2***
		4	74.3	74.3	74.2	**86.2***	**87.3***	**88.7***	**89.0***
		5	59.4	59.4	59.4	**83.8***	**84.7***	**84.7***	**85.8***
0.010/0.085	3	3	90.6	**92.9**	**93.3**	89.3	**91.1**	**93.7***	**93.7***
		4	90.5	**91.3**	**91.4**	89.2	91.3	**92.5***	**93.1***
		5	89.4	89.4	89.3	**91.7***	**92.8***	**93.0***	**94.2***
0.007/0.070	3	3	87.4	**93.4**	**94.2**	**89.9**	**91.6**	**94.2***	**94.7***
		4	91.0	**92.8**	**92.4**	88.7	**91.0**	**93.3***	**93.7***
		5	91.5	**91.7**	**91.8**	**91.6**	**93.6***	**93.4***	**94.2***
0.005/0.060	3	3	85.0	**94.8**	**94.7**	**90.2**	**91.5**	**94.7**	**95.3***
		4	89.9	**93.1**	**93.7**	89.2	**91.5**	**94.2***	**94.3***
		5	91.0	**92.6**	**92.8**	**91.2**	**93.2***	**93.9***	**94.1***
Mean			85.5	**87.7**	**87.8**	**89.0***	**90.7***	**91.9***	**92.6***
Max			91.5	**94.8**	**94.7**	**91.7**	**93.6**	**94.7**	**95.3***

Bold numbers represent performances that are better than those of SIFT. Bold numbers marked with $*$ in HLoG and HDoG+HLoG represent performances that are better than those of HDoG and both HDoG and HLoG, respectively.

fusion with HDoG and HLoG. The fusion of multiple detectors is performed at the keypoint level, i.e., we simply combined keypoints from multiple detectors and used them for the matching. Table 3 shows rank-1 matching accuracies of the fusions of HDoG and HLoG using Diag. and All matching schemes. It can be seen that the fusion of HDoG and HLoG shows consistent performance improvement from either HDoG or HLoG. Example matching results are shown in Fig. 4 with a few different cases where HDoG failed but HLoG succeeded, HDoG succeeded but HLoG failed, and both HDoG and HLoG succeeded.

The reason for performance improvement with the fusion of multiple derivation orders in image matching can be explained in Fig.3; the fusion of multiple derivation orders increases the number of matching keypoints and hence improves the image matching accuracy. Also, the fusion of HDoG and HLoG with multiple derivation orders shows performance improvement due to the increase in the number of matching keypoints. More importantly, the performance improvement of the image retrieval test proves that there are additional keypoints detected by HLoG and matched correctly, which were not detected by HDoG. Therefore, we can expect to achieve better performance in various keypoint-related tasks, for example, image registration or 3D reconstruction, with the proposed method.

3) PANORAMIC STITCHING TEST

We performed panoramic stitching experiments to further support the effectiveness of the proposed method. We fed the sets of matching keypoints from HDoG, HLoG, and the fusion of HDoG and HLoG up to the 4th degree. Out of the four image pairs, we found one pair that could not be stitched by either HDoG or HLoG, but was successfully stitched by using the fusion of HDoG and HLoG. The stitching results of Fig. 5(a) and (b) are shown in Fig. 5(c). We used the Open Source Computer Vision Library (OpenCV, version 2.4.9) [13] for the panoramic stitching operation.

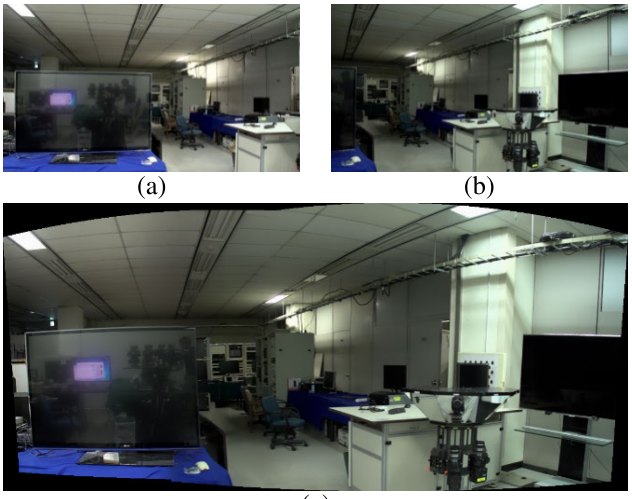
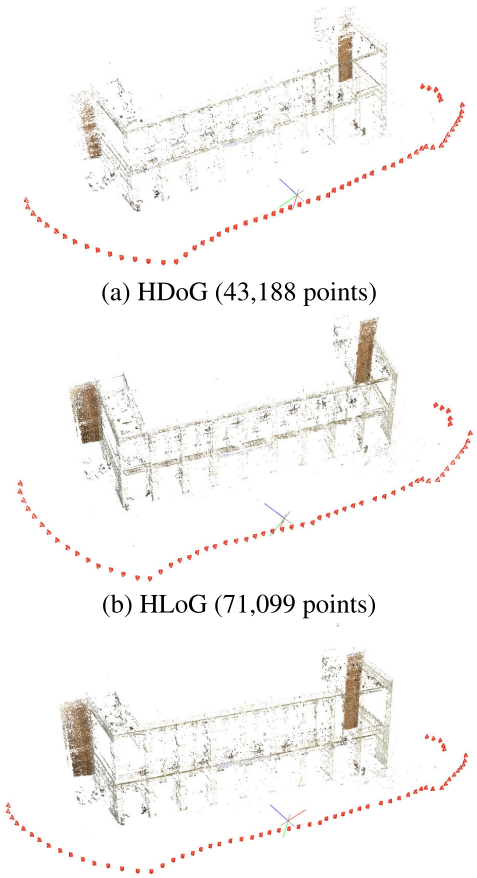


FIGURE 5. Example images of panoramic stitching: (a), (b) a pair of images with overlapping regions and (c) stitching results of (a) and (b).

4) 2D TO 3D RECONSTRUCTION TEST

Finally, we evaluated the application of the proposed method for 3D reconstruction, which is a well-known problem that involves keypoint matching of multiple 2D images. We used a 2D image sequence data set capturing a Hall [14] for the evaluation. Fig. 6 shows the reconstruction results using only HDoG or HLoG and both HDoG and HLoG. The number of reconstructed 3D points are 43,188, 71,099, and 82,612 from



(a) HDoG (43,188 points)

(b) HLoG (71,099 points)

(c) HDoG + HLoG (82,612 points)

FIGURE 6. Example results of 3D reconstruction using keypoints from HDoG, HLoG, and both of HDoG and HLoG.

HDoG, HLoG, and HDoG+HLoG, respectively. As shown in the previous tests, the increased number of keypoints by using HLoG also helps in 2D to 3D reconstruction tasks.

IV. CONCLUSION AND FUTURE WORK

In conclusion, we proposed an alternative keypoint detector based on higher order spatial derivatives, HLoG. We showed the effectiveness of the proposed method based on four different experimental evaluations. Firstly, a repeatability test was performed using a small set of images with known homography information. Secondly, an image retrieval test was performed using a data set with a larger number of images. Thirdly, a panoramic stitching test was performed to show the effectiveness of the proposed method in the image alignment problem. Finally, we performed a 3D reconstruction test and verified that the proposed method helped in generating a larger number of 3D points. We have successfully shown that the proposed HLoG-based keypoint detection method is advantages when it is used together with HDoG for various computer vision problems. We believe that the improvement in performance with the fusion of HDoG and HLoG is due to the difference between the base operators (i.e., DoG vs. LoG). The differences between HDoG and HLoG add up in each derivation order and become significant when multiple derivation orders are considered.

We will further evaluate HDoG and HLoG to understand their characteristics and optimal parameter settings. We are optimizing and trying to apply a normalization technique to the HLoG filters to stabilize the filter response to high frequency components. We will also apply HLoG to more computer vision problems such as 2D or 3D alignments and evaluate the performance.

ACKNOWLEDGMENT

The authors thank Professor Anil K. Jain at Michigan State University for providing the tattoo image database.

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *Comput. Vis. Image Understand*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[3] J. M. Seok and Y. Lee, "Visual-attention-aware progressive RoI trick mode streaming in interactive panoramic video service," *ETRI J.*, vol. 36, no. 2, pp. 253–263, Apr. 2014.
[4] S. Saeed, R. Hafiz, A. Rasul, M. M. Khan, Y. Cho, U. Park, and J. Cha, "A unified panoramic stitching and multi-projector rendering scheme for immersive panoramic displays," *Displays*, vol. 40, pp. 78–87, Dec. 2015.
[5] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555.
[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
[7] U. Park, J. Park, and A. K. Jain, "Robust keypoint detection using higher-order scale space derivatives: Application to image retrieval," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 962–965, Aug. 2014.
[8] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, Nov. 1998.
[9] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Nov. 2005.
[10] 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.
[11] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2018.
[12] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "LF-Net: Learning local features from images," in *Proc. Conf. Neural Inform. Process. Syst.*, Nov. 2018, pp. 1–13.
[13] *Open Source Computer Vision Library*. Accessed: Dec. 10, 2019. [Online]. Available: http://sourceforge.net/projects/opencvlibrary
[14] *Patch-based Multi-view Stereo Software*. Accessed: Dec. 10, 2019. [Online]. Available: https://www.di.ens.fr/pmvs/
[15] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 214–227.
[16] 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.
[17] S. Salti, A. Lanza, and L. D. Stefano, "Keypoints from symmetries by wave propagation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 2898–2905.
[18] F. Tombari and L. D. Stefano, "Interest points via maximal self-dissimilarities," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 586–600.
[19] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler, "D2-Net: A trainable CNN for joint description and detection of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 8092–8101.
[20] K. Lenc and A. Vedaldi, "Learning covariant feature detectors," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 100–117.

[21] Y. Verdie, K.-M. Yi, P. Fua, and V. Lepetit, "TILDE: A temporally invariant learned detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5279–5288.
[22] A. Barroso-Laguna, E. Riba, D. Ponsa, and K. Mikolajczyk, "Key. Net: Keypoint detection by handcrafted and learned CNN filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 1–9.
[23] N. Savinov, A. Seki, L. Ladicky, T. Sattler, and M. Pollefeys, "Quad-networks: Unsupervised learning to rank for interest point detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1822–1830.
[24] J. Revaud, P. Weinzaepfel, C. De Souza, N. Pion, G. Csurka, Y. Cabon, and M. Humenberger, "R2D2: Repeatable and reliable detector and descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 1–12.

YONGJU CHO received the B.S. and M.S. degrees from Iowa State University, in 1997 and 1999, respectively, and the Ph.D. degree from Michigan State University, in 2009, all in electrical and computer engineering. He joined the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, in 2001. He is also a Professor with the University of Science and Technology, Daejeon. He has been involved in the development of a data broadcasting system, multimedia networking, and immersive media. His research interests include digital signal processing, adaptive wireless video communications, and computer vision.

DOJIN KIM received the B.S. and M.S. degrees in computer science from Sogang University, Seoul, South Korea, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include pattern recognition, computer vision, and machine learning.

SALEH SAEED received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2011, and the M.Sc. degree in electrical engineering from the National University of Sciences and Technology, Pakistan, in 2015. He is currently pursuing the Ph.D. degree with Sogang University, Seoul, South Korea. From 2014 to 2015, he was a Research Assistant with the Vision, Imaging, and Signal Processing Laboratory, National University of Sciences and Technology (NUST). He was with Mentor Graphics, Palmchip, and Ebryx. His research interests include image processing, computer vision, and machine learning.

MUHAMMAD UMER KAKLI received the B.S. degree in electronic engineering from International Islamic University, Pakistan, in 2009, and the M.S. degree in electrical engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2013. He is currently pursuing the Ph.D. degree with the University of Science and Technology, Daejeon, South Korea. From 2013 to 2015, he was a Research Associate with the Vision, Imaging, and Signal Processing Laboratory, NUST. His research interests include image and video processing, multiview imaging, and computational photography.

SOON-HEUNG JUNG received the B.S. degree in electronics from Pusan National University, South Korea, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2003 and 2016, respectively. From March 2003 to March 2005, he was with LG Electronics. Since April 2005, he has been a Principal Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His research interests include immersive media, video coding, video signal processing, realistic broadcasting systems, and visual communication.

JEONGIL SEO was born in Goryoung, South Korea, in 1971. He received the M.S. and Ph.D. degrees in electronics from Kyoungpook National University (KNU), Daegu, South Korea, in 1996 and 2005, respectively. He was a Senior Member Engineering Staff with the Laboratory of Semiconductors, LG Semicon, Cheongju, South Korea, from 1998 to 2000. He has been a Managing Director of the Tera-Media Research Group, Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, since 2000. His recent research interests include immersive video (panorama, 360 video, and light field) processing, immersive audio processing, and realistic broadcasting systems.

UNSANG PARK received the B.S. and M.S. degrees in materials science from Hanyang University, Seoul, South Korea, in 1998 and 2000, respectively, and the M.S. and Ph.D. degrees in computer science from Michigan State University, MI, USA, in 2004 and 2009, respectively. He is currently an Associate Professor with the Department of Computer Science and Engineering, Sogang University. His research interests include pattern recognition, image processing, computer vision, and machine learning.