Robust Keypoint Detection Using Higher-Order Scale Space Derivatives: Application to Image Retrieval

Unsang Park, *Member, IEEE,* Jongseung Park, and Anil K. Jain, *Fellow, IEEE*

Abstract—**Image retrieval has been extensively studied over the last two decades due to the increasing demands for the effective use of multimedia data. Among various approaches to image retrieval, scale space representation and local keypoint descriptors have been shown to be a promising approach. Even though the concept of scale space representation has been known for a long time, it has now gained prominence as a powerful method for image retrieval mostly due to the invention of the Scale Invariant Feature Transform (SIFT). We will review the characteristics of the scale space operation and provide an extended method of scale space operation that significantly improves the image matching accuracy in the context of image retrieval. We use an operational tattoo image database containing 1,000 near duplicate images to show the superior retrieval performance of the proposed method compared to SIFT keypoints.**

Index Terms—**Image matching, image retrieval, keypoint, SIFT, scale space.**

I. INTRODUCTION

CTENT-based image retrieval (CBIR) is used as a representative term covering research on image retrieval, where the query is an image as opposed to keywords describing the image. CBIR has been extensively studied over the past couple of decades with significant progress [1]–[4]. A number of approaches have been proposed in CBIR adopting various techniques in image processing, computer vision, machine learning, and pattern recognition. However, it is well known that CBIR is still not a tractable problem due to many unresolved issues, including semantic gap [5].

Near Duplicate Image (NDI) retrieval [6] is a problem similar to CBIR but more tractable due to its limited goal of finding images that are visually similar (as opposed to semantically similar) to the query image. Even though NDI retrieval is a relatively easier problem than CBIR, it was only after the invention of SIFT [7] that an acceptable level of performance has been achieved. Following SIFT, there have been a number of similar approaches to extrcat local keypoints and descriptors such as SURF [8], ORB [9], and Fast SIFT [10]. We will first review the fundamentals and characteristics of SIFT, and then introduce our proposed method that extends the keypoint detection capability of SIFT.

U. Park and J. Park are with the Department of Computer Science and Engineering, Sogang University, Seoul, Korea. U. Park is the corresponding author. e-mail:{unsangpark,pleiadess07}@sogang.ac.kr

A. K. Jain is with the Department of Computer Science and Engineering, Michigan State University, E. Lansing, MI 48824, USA. e-mail:jain@cse.msu.edu.

The rest of the paper is composed as follows. In Sec. II, we will review the keypoint extraction method of SIFT in scale space. In Sec. III, we will propose an extended keypoint detection method using higher-order scale space derivatives. Sec. IV provides experimental results and Sec. V will conclude the paper.

II. REVIEW OF SIFT KEYPOINT DETECTION METHOD

While the scale space representation has been studied for over 30 years [11], [12], it has only recently been successfully applied to image matching tasks [7], [8]. The scale represents the size of an image or an object in an image. If an object appears at a different scale (size) in two different images (i.e., query and database), these two images do not match very well. We can consider a naive approach that represents the database or query image at a number of different scales and then perform matching. However, the large number of possible scales at which the object can appear makes this approach infeasible. Template matching based methods especially suffer from the scale change. Therefore, a number of approaches have been proposed to construct a scale space that consists of a set of images, where each image represents a specific scale [12]. This is followed by extraction of characteristic feature points (keypoints) that are invariant to the scale changes in the scale space.

The scale space is usually constructed by applying the Gaussian smoothing or blurring to the image [7]. Each blurring step is characterized by the standard deviation of the Gaussian kernel. To calculate the scale space derivatives, SIFT uses Difference of Gaussian (DoG) method that approximates the Laplacian of Gaussian [7]. In DoG images, pixels that are greater or less than their 26 neighboring pixel values are selected as local extrema, whose locations correspond to derivatives being equal to zero. This procedure can be essentially interpreted as finding the local maxima or minima using the first-order derivatives in the scale space $(\partial L(x, y, \sigma)/\partial \sigma = 0)$. We will show how to use this observation to extend the keypoint extraction method of SIFT in the following section.

III. EXTENDED KEYPOINT DETECTION METHOD IN SCALE SPACE

The keypoint extraction method of SIFT has been shown to be effective in image matching and retrieval in a number of studies [7]–[10]. We have also observed that the underlying principle is to find the local extrema in scale space for
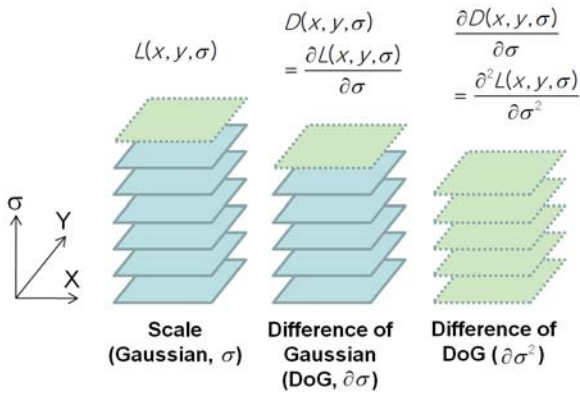
Fig. 1. Schematic of scale space construction for higher-order derivatives. The scale images shown with different color and dotted lines represent the additional scales considered for the higher-order scale derivative.

the keypoint detection. More sepcifically, SIFT uses first-order scale space derivative ($\partial L(x,y,\sigma)/\partial\sigma = 0$), which is approximated by taking the difference of successive scale images to obtain DoG images. Therefore, it is natural to explore the possibility of using higher-order scale space derivatives, i.e., $\partial^2 L(x,y,\sigma)/\partial\sigma^2 = 0$, $\partial^3 L(x,y,\sigma)/\partial\sigma^3 = 0$, $\partial^4 L(x,y,\sigma)/\partial\sigma^4 = 0$, etc., to extract more keypoints than SIFT. Fig. 1 shows the schematic of constructing higher-order scale space for the proposed method. Higher-order scale spaces are obtained by successively subtracting previous scale spaces of lower order, similar to how DoG space (D) is obtained by subtracting consecutive scale images of the previous order (L). These scale spaces are labeled as σ, $\partial\sigma$, and $\partial\sigma^2$ acoording to the order of derivatives as shown in Fig. 1. Scale spaces higher than the order of two can also be similarly labeled. Scale space of $\partial\sigma^i$ is constructed by subtracting consecutive scales in $\partial\sigma^{i-1}$. The local Extrema in the second-order scale space corresponds to the inflection point in the 1st-order scale space derivative.

In the conventional SIFT, keypoint extraction from s different scales requires $(s + 3)$ scales constructed to generate $(s+2)$ different DoG images. The additional two DoG images ensure all pixles in s DoG images have 26 neighbors. As the order of scale space derivative increases, there needs to be additional number of scale spaces. To generalize this relationship, keypoint extraction from s scale spaces using i^{th}-order of scale space derivative ($\partial\sigma^i$) requires $(s+2+i)$ scale spaces. Fig. 2 shows example keypoint extraction results up to the 4th-order scale space derivative. Since the proposed mehtod extracts more keypoints, it requires additional time both in keypoint extraction and matching.

IV. EXPERIMENTAL RESULTS

A. Database

We use a subset of an operational Tattoo image dataset obtained from Michigan State Forensics Laboratory [13]. The entire database includes 63,593 color images of size 640×480 pixels. We have manually identified 445 tattoos that have duplicate images (1,000 total images). These 445 tattoos have about 2.25 duplicate images on average in the database of

(a) Original image

(b) $\partial\sigma$ (SIFT, 16)

(c) $\partial\sigma^2$ (80)

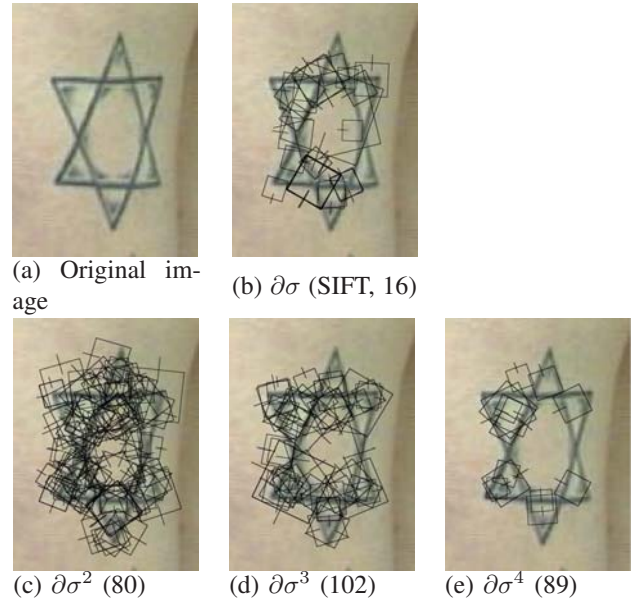(d) $\partial\sigma^3$ (102)

(e) $\partial\sigma^4$ (89)

Fig. 2. Keypoint extraction from various orders of scale space derivatives. Each keypoint is marked by a box whose size represents the scale of a keypoint. The short line on one edge of a box represents the major orientation. The numbers in parentheses represent the numbers of extracted keypoints.

1,000 images. In the matching experiments, each image in the database serves as a query and the rest of the database is considered as the gallery.

B. Image Retrieval Using Higher Oder Scale Space Derivatives

The set of parameters we have tested for the general SIFT framework consists of number of octaves (O), number of scales (S), and cutoff threshold values (T_x), which are chosen as {3}, {3, 4, 5}, and {0.02, 0.01, 0.007, 0.005}, respectively. In general, the numbers of octaves and scales are proportional to the number of keypoints extracted because there will be more number of scales and DoG images; the larger the number of octaves and scales, the larger the number of extrcated keypoints. However, the increase in scale space reduces the differences of pixel values across successive scales, which may reduce the number of pixels that are larger or smaller than their 26 neighbors, resulting in the reduction in the number of keypoints. The parameter T_x is inversly proportional to the number of keypoints because a small value of T_x increases the number of candidate keypoints. All the extracted keypoints are input to the subsequent descriptor construction and keypoint matching processes which are the same as those in the traditional SIFT [7].

Once keypoints are extracted from higher-order scale space derivatives, a number of different matching schemes can be considered that utilize the information about derivative orders. The first scheme allows keypoint matches within the same order of scale space derivatives, that is, keypoints from $\partial\sigma^i$ can only match to the keypoints from the same order. This is based on the intuition that the keypoints from the same derivative order are more likely to match each other. Restricting the matching candidates to the same derivative

TABLE I
A COMPARISON OF RANK-1 MATCH ACCURACIES USING HIGHER-ORDER SCALE SPACE DERIVATIVES AND A FEW REPRESENTATIVE MATCHING SCHEMES.

Cutoff D (T_x)	#Octaves (O)	#Scales (S)	Average #keypoints/image*				Rank-1 accuracy** (%)			
			$\partial\sigma$	$\partial\sigma^2$	$\partial\sigma^3$	$\partial\sigma^4$	$\partial\sigma$	Diag.	All	Off-diag.
0.02	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
0.01	3	3	228.05	138.28	240.30	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	12.90	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

∗ $\partial\sigma^i$ represents extracting keypoints using i^{th}-order scale space derivative.

∗∗ "$\partial\sigma$" uses keypoints from 1st-order scale space derivative, "Diag." uses keypoints from same order of scale space derivatives, "All" uses all keypoints, and "Off-diag" allows keypoints only from different order of derivative to match.

order can reduce the computation time while still yielding a good matching accuracy. The second scheme allows keypoints to match across different derivative orders. We name the first scheme as *diagonal match*, and the second as *all match*. One additional matching scheme is to allow keypoints to match only across different derivative orders. This scheme is named as *off-diagonal match*. The off-diagonal match scheme is evaluated to analyze the fraction of keypoints matched across different derivative orders. We have also calculated the matching accuracies using the keypoints only from $\partial\sigma$ (i.e., SIFT) as the baseline performance.

The average number of keypoints and the rank-1 matching accuracies for four different matching schemes are summarized in Table I. Based on the experimental results summarized in Table I we can make the following observations: (i) The best performance of the SIFT in our experimental setup is 91.5%, (ii) the use of higher-order derivatives show significantly better matching accuracy compared to the SIFT for all parameter settings, (iii) the best matching accuracy is observed in either diagonal match or all match schemes depending on different parameter settings, and (iv) the difference between the best matching accuracy of SIFT and the proposed method is 3.3%. Fig. 3 shows example matching results using conventional SIFT and the proposed higher-order scale space derivative. Since the image matching time is proportional to the number of keypoints, the matching process takes longer when more keypoints are detected in the proposed method compared to SIFT. It is expected that the matching time can be significantly reduced when the Bag-of-Words matching scheme is used [13].

It is apparant the the performance gain of the higher-order SIFT comes from the additional keypoints extracted. In our experiments, SIFT failed to extract keypoints more than 314 even when the T_x is set to 0, but higher-order SIFT succeeded to extract more keypoints and provided improved matching accuracy. We consider the higher-order SIFT as a supportive method to improve the performance of the SIFT.
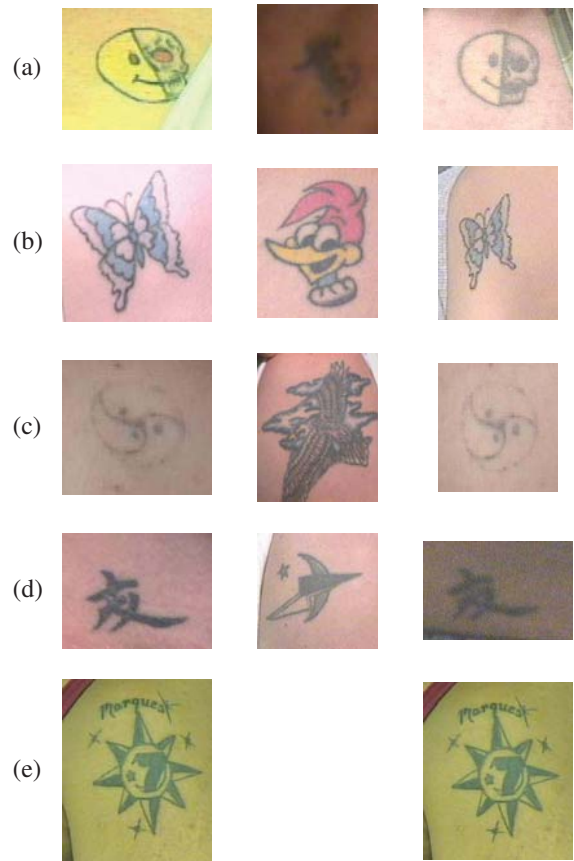

Fig. 3. Example matching results; first, second, and third columns correspond to image queries, false mates, and true mates, respectively: (a) - (c) examples where first-order scale derivative failed, but higher-order scale derivative succeeded, (d) example where first-order scale derivative succeeded, but higher-order scale derivative failed (e) example where both the first-order and higher-order scale derivatives succeeded. Values of parameters O, S, and T_x were chosen as 3, 3, and 0.005 to obtain these results.

We believe that this performance improvement should be

advantages for some applications where the accuracy is more important compared to the computational cost.

V. Conclusions and Future Work

We have proposed an extended keypoint detection method that significantly improved the image matching accuracy on an operational tattoo image database. The core of the method is extending the basic property of keypoint extraction in SIFT from first-order scale space derivative to higher-order scale space derivatives. Experimental results showed that the proposed method significantly outperforms the conventional SIFT in a number of different parameter setups. As an example, when values of parameters T_x, O, and S are 0.005, 3, and 3, respectively, rank-1 retrieval accuracy of the proposed method increased from 85.0% for SIFT to 94.8%. Future work includes utilizing adaptive cutoff threshold value (T_x) to detect a smaller but more informative set of keypoints, and reducing the overall computation time in image matching.

References

[1] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation and psychophysical experiments," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 20–37, 2000.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, 2002.

[4] M. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 2, no. 1, pp. 1–19, 2006.

[5] N. Rasiwasia, P. J. Moreno, and N. Vasconcelos, "Bridging the gap: Query by semantic example," *IEEE Trans. on Multimedia*, vol. 9, no. 5, pp. 923–938, 2007.

[6] J. J. Foo, J. Zobel, and R. Sinha, "Clustering near-duplicate images in large collections," in *Proc. of the international workshop on multimedia information retrieval*, 2007, pp. 21–30.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[8] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. International Conf. Computer Vision*, 2011, pp. 2564–2571.

[10] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated sift," in *Proc. Asian Conf. Computer Vision*, 2006, pp. 918–927.

[11] B. P. J and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. COM-31, no. 4, pp. 532–540, 1983.

[12] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 1573–1405, 1998.

[13] J.-E. Lee, R. Jin, and A. K. Jain, "Unsupervised ensemble ranking: Application to large-scale image retrieval," in *Proc. Int. Conf. Pattern Recognition*, 2010, pp. 3902–3906.