A ROBUST FEATURE DESCRIPTOR BASED ON MULTIPLE GRADIENT-RELATED FEATURES

Zhaomang Sun, Fei Zhou, Qingmin Liao*

Shenzhen Key Lab of Information Science and Technology/ Shenzhen Engineering Lab of IS&DRM
Department of Electronics Engineering/ Graduate School at Shenzhen, Tsinghua University, China

ABSTRACT

In this paper, we propose a robust descriptor named as multiple gradient-related features (MGRF) in virtue of local and overall order encoding. Specifically, three types of features are introduced, including multidirectional gradient, gradient orientation, and first derivative of gradient orientation, each of which represents different aspect of region of interest (ROI). To extract these features, we also propose a novel sampling pattern of tree structure. Furthermore, each gradient-related feature is encoded with both local and overall order information of ROI, and the encoding results are respectively called local and overall gradient order code (GOC). Finally, our descriptor is formed by concatenating the respective feature vector of each type of feature, which is computed as a 2-D joint histogram of GOC and ordinal bin. The experiments conducted on Oxford dataset demonstrate that the proposed descriptor significantly outperforms other state-of-the-art descriptors.

***Index Terms*—** feature description, ordinal information, multiple gradient-related features, gradient order code

1. INTRODUCTION

Local image descriptors lay important foundations on many visual applications in computer vision and pattern recognition, involving image retrieval, image matching, 3D reconstruction, panoramic stitching, object tracking and recognition, etc. In many applications, it has been shown that local image description have more remarkable influences on the performance of local features compared with the different choices of detection methods [1]. Therefore, current researches on local features are mainly devoted to computing a more discriminative and robust descriptor for region of interest (ROI).

Recently, large numbers of descriptors have been proposed in the literature. The most typical descriptor is SIFT [2], which is computed as a histogram of gradient orientation on 4×4 location cells. Many other similar feature descriptors such as GLOH [1], SURF [3], and DAISY [4], also have been introduced, encouraged by the success of the SIFT description. Unfortunately, although these descriptors are fully or partially robust to many of the geometric image transformations such as rotation, scale, occlusions, etc., they cannot deal with complex brightness changes. To cope with this problem, an increasing number of methods, which use orders of intensities rather than the raw intensity values, are proposed, such as HRI-CSLTP [5], LIOP [6], MROGH [7], MRRID [8], and MIOP [9]. In [5], the HRI descriptor is constructed as a histogram of overall intensity order on a 4×4 spatial cells similar to the location cells used by SIFT. The CS-LTP descriptor captures local gradient properties, which is an improvement of CS-LBP by replacing a binary code with a ternary code. As for [6, 7, 8, 9], the only difference lies in different encoding schemes: LIOP uses a permutation-based encoding scheme [6], MROGH adopts a gradient-based encoding method [7], and MRRID utilizes an intensity comparison approach similar to CS-LBP [8]. An interesting descriptor, known as MIOP, is presented in [9], which is the first to integrate local and overall intensity order information. The MIOP descriptor is obtained by concatenating LIOP [6] and OIOP descriptor in [9], which encodes each sampling point with the quantized overall intensities order.

The excellent performances of these descriptors [5, 6, 8, 9] have demonstrated the effectiveness of using ordinal information to describe ROI. However, all of these descriptors only use intensities to encode local information and ignore use of the gradient, which can capture the structural and texture information in images in detail. Meanwhile, the gradient is less susceptible to lighting and camera changes. Therefore, these descriptors [5, 6, 8, 9] cannot preserve local contrast and edge information in detail and suffer degradation in the discriminative ability.

To make full use of the informative gradient as well as the effective ordinal information, we introduce three gradient-related features and aggregate order into gradient domain in this paper. Our contributions are threefold: 1) We establish a novel sampling pattern of tree structure to extract our proposed features. 2) We propose three types of gradient-related

*corresponding author:flying.zhou@163.com. This work was supported by the National Natural Science Foundation of China under Grant No. 61271393 and 61301183, and the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under JCYJ20150331151358138.

features and aggregate ordinal information into them. 3) We compute a 2-D joint histogram of gradient order code (GOC) and ordinal bin for each type of features.

2. OUR PROPOSED DESCRIPTOR

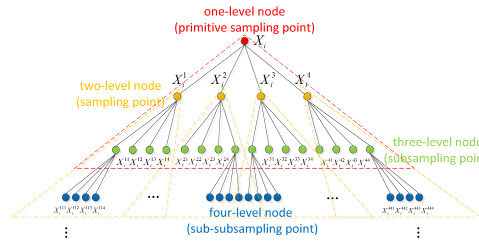
To give an overall impression of the proposed descriptor, its construction framework is briefly stated here. First, we preprocess the image with Gaussian blur, detect ROI using Hessen-Affine detector [10], and normalize ROI. Then, we make use of the overall intensity order to divide a certain ROI into sub-regions called ordinal bins, which is similar to region division in [6]. Next, multiple gradient-related features (MGRF) based on local and overall order encoding of each pixel in every ordinal bin is obtained. Last, the MGRF descriptor is constructed by accumulating the MGRF of all pixels in each ordinal bin respectively, then by concatenating them together. The following subsections give a detailed description of the process of constructing our descriptor, mainly including a novel sampling pattern of tree structure, extraction of multiple gradient-related features, as well as both local and overall order encoding for each type of the proposed features.

2.1. Sampling Pattern of Tree Structure

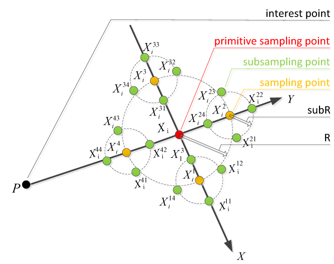
Before effective integration of the proposed gradient-related features, it is necessary to establish a new sampling pattern of tree structure shown in Fig.1(a), where nodes in each level are printed in the same color. Each node in tree structure represents a certain pixel in ROI. Specifically, the one-level node, known as the root node, denotes one of the original pixels in ordinal bins after region division [6]. Two-level nodes are those pixels sampled on circles with a certain radius centred at root node. Likewise, k-level nodes can be obtained respectively by sampling on circles with a certain radius centred at each of $(k-1)$-level nodes, where $k = 2, ..., N$, and N is the total levels of tree structure. As shown in Fig.1(a), each node in tree structure has four child nodes which can be used to compute the gradient orientation of the corresponding parent node.

In the following, we will simply set N as 3 to further describe the specific sampling method. The corresponding coordinate system established is shown in Fig.1(b), where the sampling points are denoted in the same color as the corresponding nodes in tree structure in Fig.1(a). In details, one-level node, two-level nodes, and three-level nodes respectively correspond to the primitive sampling point, sampling points, and sub-sampling points, which are labeled as characters with the corresponding color.

In Fig.1(b), P is an interest point and X_i is the i-th primitive sampling point. Then a local coordinate system can be established by setting vector $\overrightarrow{PX_i}$ as the positive y-axis and the positive x-axis is determined through the positive y-axis rotating 90 degrees in a clockwise order. The four sampling



(a) Sampling pattern of tree structure



(b) Local coordinate system ($N = 3$)

Fig. 1. Sampling pattern of tree structure and its corresponding local rotationally invariant coordinate system ($N = 3$).

points $X_i^j, (j = 1, 2, 3, 4.)$ around X_i are equally distributed on a circle with radius R centred at X_i. We set the intersection point between the positive x-axis and the circle boundary as the first sampling point X_i^1 and subsequent sampling points are X_i^2, X_i^3, X_i^4 in anticlockwise order as shown in Fig.1(b). Furthermore, three-level nodes, that is to say, subsampling points around sampling point $X_i^j, (j = 1, 2, 3, 4.)$ can be denoted as $X_i^{jp}, (p = 1, 2, 3, 4.)$ where j represents the j-th sampling point around pixel X_i, and p denotes the p-th sub-sampling point for the sampling point X_i^j. The four sub-sampling points X_i^{jp} are equally distributed on a subcircle with radius $subR$ centred at X_i^j, and their positions can be determined in the way similar to the determination of the sampling point X_i^j.

As shown in Fig.1(b), the construction of local coordinate system here is the special case in sampling pattern of tree structure where $N = 3$. When N increases, we can further extend the coordinate system with extra sampling points centered at each of nodes in the previous level of tree structure. Based on the extended coordinate system, a more informative feature can be extracted, which makes our descriptor have a more discriminative ability, unfortunately accompanied with the higher amount of calculation and weaker capability of resisting localization errors and noises. The performance of our method when $N = 3$ is satisfactory, and thus we use the coordinate system in Fig.1(b) to construct our descriptor.

2.2. Multiple Gradient-Related Features

As mentioned above, we can directly extract our features on the local coordinate system in Fig.1(b). Of course, we can extend our extration method to the general case. When N increases, tree structure can be divided into a certain number of subtrees with the depth of 3, based on which we compute our proposed features.Specifically, when $N = 4$, in tree structure there are five subtrees with the depth of 3, which are labeled as triangles with dashed line in the same color as the corresponding root node in each subtree as shown in Fig.1(a). Features can be extracted respectively based on each subtree, which indicates the completely similar process of feature extraction when N grows up.

The proposed features include multidirectional gradient, gradient orientation and first derivative of gradient orientation.We take the subtree containing the root node X_i for instance to describe the calculation procedure for each type of features.

For X_i in Fig.1(b), we can define multidirectional gradient using four sampling points $X_i^1, X_i^2, X_i^3, X_i^4$ around X_i as

$$G_i^k = X_i^k - X_i^{(k+1)\%4}, \qquad k = 1, 2, 3, 4. \tag{1}$$

where G_i^k denotes the k-th direction gradient of X_i in one ordinal bin, and % represents mod operation.

Gradient orientations of four sampling points are defined as

$$O_i^j = \arctan \frac{X_i^{j1} - X_i^{j3}}{X_i^{j2} - X_i^{j4}}, \qquad j = 1, 2, 3, 4. \tag{2}$$

where O_i^j represents the gradient orientation of j-th sampling point of X_i in one ordinal bin.

The first derivative of gradient orientation of four sampling points is equal to the difference value of the gradient orientations between two neighboring sampling points. Supposing that DO_i^j denotes the j-th difference value between gradient orientations of sampling points around X_i in one ordinal bin, then

$$DO_i^j = O_i^j - O_i^{(j+1)\%4}, \qquad j = 1, 2, 3, 4. \tag{3}$$

It needs to note that although the Equ.1,2, and 3 are formulated based on the subtree including the root node X_i, these equations can be applied to all other subtrees with the depth of 3 in tree structure when $N \geq 4$.

2.3. Encoding of Multiple Gradient-Related Features

Once three local features of all pixels in each ordinal bin are obtained, we can respectively encode them with both local and overall order information.Considering the similar encoding process in [6] and [9], we only take multidirectional gradient for instance.

• **Local encoding**: First, we sort multidirectional gradients of all sampling points around a pixel in a non-descending order and obtain the index list that represents the ranking of multidirectional gradients. Then, the index list is considered as a permutation Π and one-to-one mapping is established between the permutation Π and the corresponding index $Ind(\Pi)$ that is called local gradient order code (GOC). Finally, a feature vector, whose elements are all 0 except the $Ind(\Pi)$-th element which is 1, is obtained.

• **Overall encoding**: First, we sort multidirectional gradients of all sampling points around all pixels in all ordinal bins in a non-descending order to obtain their overall order. Then, we quantize their overall orders into C levels and get quantization values of multidirectional gradients of all pixels through quantization thresholds. Next, for each pixel in every ordinal bin, we can obtain a base-C number with a certain length by concatenating them together, which will be converted into a decimal number that is named as overall GOC. Last, a feature vector, whose elements are all 0 except the element corresponding to the decimal number which is 1, is obtained.

For two other features, following the same encoding process, we can obtain the respective feature vectors.The three types of features embody different aspects of sampling points and thus capture the complementary information.We concatenate three feature vectors obtained by local encoding and call the combined feature vector local multiple gradient-related features (LMGRF). Likewise, three other feature vectors obtained by overall encoding are concatenated as a complete feature vector, which is named as overall multiple gradient-related features (OMGRF). In addition, considering that LMGRF and OMGRF encode different ordinal information, which implies that they might have a certain degree of complementarity, and thus we further cascade LMGRF and OMGRF into a whole feature vector called multiple gradient-related features (MGRF) to improve performance.In the end, we accumulate MGRF of all pixels in each ordinal bin respectively and obtain the histograms for every ordinal bin. The MGRF descriptor can be formed by concatenating all histograms.

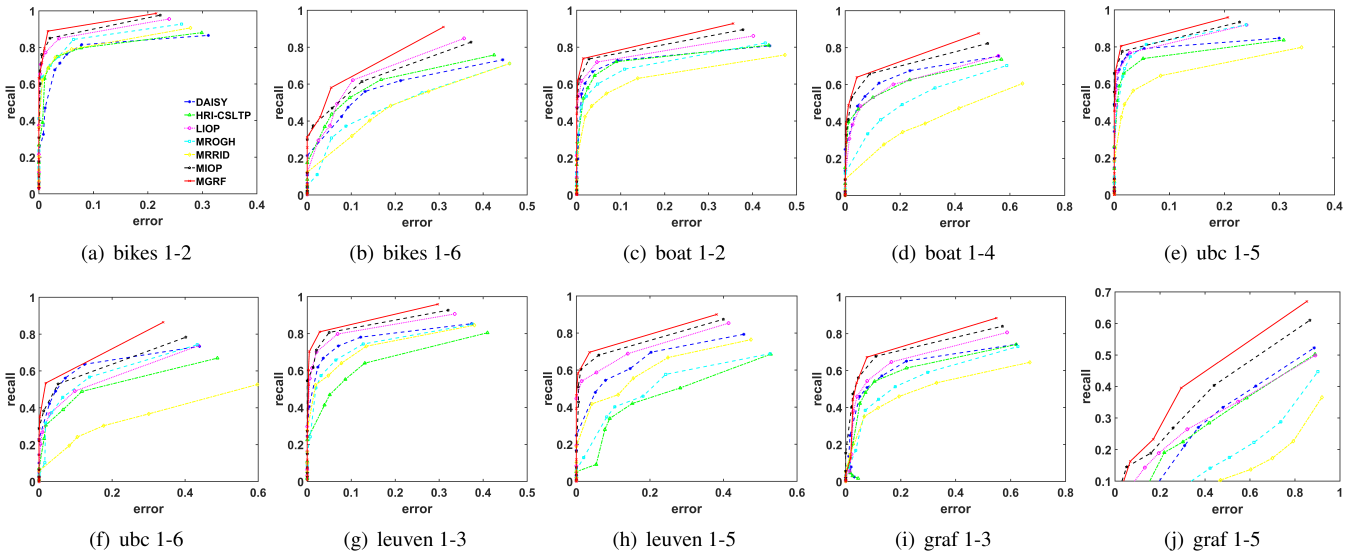
3. EXPERIMENTS

3.1. Dataset and Evaluation Criterion

We evaluate our descriptors on standard Oxford image matching dataset [11]. This dataset consists of six different geometric and photometric transformations, including image blur, scale change, image rotation, JPEG compression, illumination change and viewpoint change. And we use the evaluation criterion proposed by Mikolajczyk and Schmid [1]. The matching strategy is the nearest neighbor distance ratio (NNDR). Judging whether a match is correct is determined by the overlap error [12]. If the overlap error < 0.5, then the match is correct. The results are presented with recall versus *error* curves, where $error$=1-precision, and *error* is used as the title of the horizonal axis of recall versus 1-precision curves in replace of 1-precision shown in Fig.2.

3.2. Performance Evaluation

We compare our descriptor with DAISY [4], HRI-CSLTP [5], LIOP [6], MROGH [7], MRRID [8], and MIOP [9] descrip-



(a) bikes 1-2 (b) bikes 1-6 (c) boat 1-2 (d) boat 1-4 (e) ubc 1-5

(f) ubc 1-6 (g) leuven 1-3 (h) leuven 1-5 (i) graf 1-3 (j) graf 1-5

Fig. 2. Experiment results under various image transformations in Oxford dataset for Hessen-Affine region detector. "1-j" in caption of each sub-figure represents the image pair between the 1-st and j-th image in one image set, and $j \in \{2, 3, 4, 5, 6.\}$

tors. To be fair and representative, we utilize Hessian-Affine detector [10] to detect ROI and meanwhile adopt one support region for all the descriptors in the experiments. Furthermore, all the descriptors are extracted using parameters suggested by authors. As to our MGRF descriptor, the setting of common parameters follows the suggestions of the previous work in [9]. Meanwhile, a new parameter is introduced to represent sub-sampling radius of the sampling points around pixel X_i in one ordinal bin. As shown in Fig.1(b), sub-sampling radius is denoted as $subR$, which is set as 3 in this paper.

We test our descriptor using all the image sequences in the dataset. Due to the limited length of the paper, we only show the evaluation results of partial image pairs in Fig.2, which involve different image transformations. From the results, we can see that the recall versus *error* curves of our descriptor are always above those of other tested descriptors in all cases.

Meanwhile, to make quantitative comparisons more convincing, we introduce area under curves (AUC) to represent the area formed by the recall versus *error* curve and the horizontal axis. The larger AUC means the better performance of the corresponding descriptor. Detailed results are shown in Tab.1, where each element denotes the average AUC of all image pairs in a certain type of image sequences for the specific descriptor. From Tab.1, we can see that the AUC of our descriptor is larger than those of other descriptors in terms of each image sequence and the average of all image sequences.

From comprehensive results of Fig.2 and Tab.1, we can conclude that our descriptor outperforms all other descriptors in all cases. Superior performance in illumination changes (leuven 1-3,1-5) and rotation and scale changes (boat 1-2,1-4) demonstrates the effectiveness of aggregating order into gradient-related features. Meanwhile, good performance in terms of image blur (bikes 1-2,1-6) and JEPG compression (ubc 1-5,1-6) indicates the fact that gradient-related features can preserve more detailed local contrast and edge information and improve the discriminative ability.

Table 1. AUC comparisons for different descriptors

Descriptors	Image Sequences in Oxford dataset						
	bikes	boat	ubc	leuven	graf	wall	average
HRI-CSLTP[5]	0.7595	0.6310	0.8748	0.7840	0.4915	0.5297	0.6889
DAISY[4]	0.7768	0.6103	0.8507	0.6463	0.4836	0.6004	0.6613
LIOP[6]	0.8772	0.6211	0.9061	0.8324	0.5236	0.6252	0.7309
MROGH[7]	0.7786	0.5432	0.9064	0.6856	0.4422	0.5459	0.6503
MRRID[8]	0.7511	0.4613	0.8185	0.7332	0.3775	0.4685	0.6017
MIOP[9]	0.8751	0.6854	0.9214	0.8574	0.5799	0.7372	0.7761
MGRF	**0.9200**	**0.7315**	**0.9455**	**0.8815**	**0.6192**	**0.7464**	**0.8073**

4. CONCLUSION

In this paper, three types of gradient-related features are introduced for local feature description. Each of these features represents different aspect of ROI. Meanwhile, we propose a novel sampling pattern of tree structure to extract them. Furthermore, by aggregating both local and overall order information into each type of these features, we can see that our descriptor has both highly discriminative ability and strong robustness. Therefore, we achieve superior performance in illumination changes and rotation and scale changes, as well as image blur and JEPG compression. Experiment results have shown that our proposed descriptor outperforms all the other state-of-the-art methods under various geometric and photometric image transformations.

5. REFERENCES

[1] 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, Oct 2005.

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

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, *SURF: Speeded Up Robust Features*, pp. 404–417, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[4] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, May 2010.

[5] R. Gupta, H. Patil, and A. Mittal, "Robust order-based methods for feature description," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 334–341.

[6] Zhenhua Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 603–610.

[7] B. Fan, F. Wu, and Z. Hu, "Aggregating gradient distributions into intensity orders: A novel local image descriptor," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, June 2011, pp. 2377–2384.

[8] B. Fan, F. Wu, and Z. Hu, "Rotationally invariant descriptors using intensity order pooling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2031–2045, Oct 2012.

[9] Z. Wang, B. Fan, G. Wang, and F. C. Wu, "Exploring local and overall ordinal information for robust feature description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2016.

[10] Krystian Mikolajczyk and Cordelia Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[11] "Oxford dataset," `http://www.robots.ox.ac.uk/~vgg/research/affine/`, accessed Jun. 2016.

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.