Aggregating Gradient Distributions into Intensity Orders: A Novel Local Image Descriptor

Bin Fan Fuchao Wu Zhanyi Hu
National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, 100190, Beijing, China
{bfan,fcwu,huzy}@nlpr.ia.ac.cn

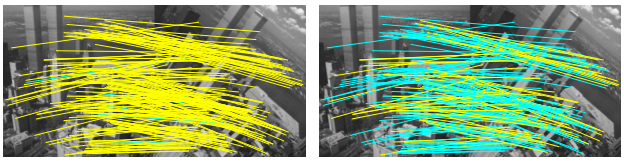
Abstract

A novel local image descriptor is proposed in this paper, which combines intensity orders and gradient distributions in multiple support regions. The novelty lies in three aspects: 1) The gradient is calculated in a rotation invariant way in a given support region; 2) The rotation invariant gradients are adaptively pooled spatially based on intensity orders in order to encode spatial information; 3) Multiple support regions are used for constructing descriptor which further improves its discriminative ability. Therefore, the proposed descriptor encodes not only gradient information but also information about relative relationship of intensities as well as spatial information. In addition, it is truly rotation invariant in theory without the need of computing a dominant orientation which is a major error source of most existing methods, such as SIFT. Results on the standard Oxford dataset and 3D objects have shown a significant improvement over the state-of-the-art methods under various image transformations.

1. Introduction

Local image descriptors computed from interest regions have been widely studied in the community of computer vision. Recent years, they have became more and more popular and have been shown to be useful for a variety of visual tasks, such as 3D reconstruction [1, 4], structure from motion [17], object recognition [10] and classification [24] as well as panoramic stitching [18], to name a few.

Many methods in the literature have been proposed for extracting interest regions. Widely used methods include Harris-affine/Hessian-affine [12], Maximally Stable Extremal Regions (MSER) [11], intensity and edge-based detectors [21]. Please see [14] for a comprehensive study of these affine regions. Once interest regions have been extracted, feature descriptors are computed from these interest regions (affine normalized) in order to distinguish them from each other. A main concern about the design of local descriptor is to make it distinctive while simultaneously robust to as many image transformations as possible. This paper is focused on the problem of designing image descriptors for local interest regions. More specifically, a novel local image descriptor is introduced. The novelties, and hence our main contributions include:



(a) With SIFT descriptor (b) With the proposed descriptor

Figure 1. Matching results of corresponding points with orientation assignment errors larger than 20 degrees. The corresponding points that are also matched by their descriptors are marked with cyan lines while yellow lines indicate those corresponding points that are un-matchable by their descriptors.

1) Given a support region for descriptor construction, gradients of sample points are computed in a rotation invariant way. The proposed descriptor is rotation invariant without resorting to computing a dominant orientation. According to our experimental study, errors in assigning dominant orientation are the main source of the false negatives (the true corresponding points that are not matched by their descriptors). Therefore, compared with those descriptors which achieve the rotation invariance by assigning a dominant orientation, our method is more stable. Fig. 1 shows the matching results of corresponding points whose orientation estimation errors are larger than 20 degrees. It is clear that many corresponding points can still be correctly matched by our proposed descriptor even if the orientation assignment errors are large.

2) In order to encode spatial information into the descriptor to enrich its discriminative power, an adaptive strategy is proposed for pooling gradients spatially. In our work, sample points are segmented based on their intensity orders and

their gradients are then pooled for each segment. Therefore, not only spatial information, but also the relationship of intensities among sample points are inherently encoded.

3) Multiple support regions are used for the descriptor construction to further enhance its discriminative ability. Although two non-corresponding points may have similar appearances in a certain size of support region, they usually can be easily distinguished in a different size of support region. Thus by constructing our descriptor from multiple support regions, discriminative ability is improved.

All these factors contribute to the good performance of our proposed descriptor. Its performance has been tested on two widely used datasets: one from the Oxford (2D objects) [13] and the other from the Caltech (3D objects) [15]. The experimental results are quite encouraging.

The rest of this paper is organized as follows: Section 2 gives a brief overview of the related works. Then, our new descriptor is presented in Section 3, followed by the experiments in Section 4. Finally, we conclude this paper in Section 5.

2. Related Work

The research and design of local image descriptors have received many attentions of researchers in the field of computer vision. Perhaps one of the most famous and popular descriptors is SIFT (Scale Invariant Feature Transform) [10]. According to the comparative study of Mikolajczyk and Schmid [13], SIFT and its variant GLOH (Gradient Location and Orientation Histogram) outperforms other local descriptors such as shape context, steerable filters, spin images, differential invariants and moment invariants. Inspired by the high discriminative ability and robustness of SIFT, many researchers have developed local descriptors following the way of SIFT. Ke and Sukthankar [8] applied PCA (Principal Component Analysis) to gradient patch of keypoint and introduced the PCA-SIFT descriptor which is said to be more compact and distinctive than SIFT. Bay et al. [2] proposed an effective implementation of SIFT via the integral image technique, achieved 3 to 7-fold speed-ups. Tola et al. [20] developed a fast descriptor named DAISY for dense matching. Winder et al. [22] proposed a framework to learn local descriptors with different combinations of local features and spatial pooling strategies. The SIFT and many other descriptors proposed before can be incorporated into their framework and it is said that a DAISY-like descriptor has the best performance among all configurations. Then, the best DAISY was picked in [23].

In order to deal with the problem of illumination changes, some researchers proposed to design local descriptors based on intensity orders since the intensity orders are invariant to monotonic illumination changes. Gupta and Mittal [5] proposed a monotonic change invariant feature descriptor based on intensity orders of point pairs in the interest region. The point pairs are carefully chosen from extremal regions in order to be robust to localization error as well as to intensity noise. Matching of this kind of descriptors is based on a distance function that penalizes flip of orders. In [19], Tang et al. used a 2D histogram of position and intensity order to construct a feature descriptor to deal with complex brightness changes. While their work directly used intensity order with respect to the entire patch, Marko Heikkila et al. [7] proposed to use LBP (Local Binary Pattern) which encodes ordering relationship locally for feature description. Instead of gradient features, LBP was incorporated in the framework of SIFT and the obtained descriptor CS-LBP was reported to have better performance than SIFT. In [6], Gupta et al. generalized the CS-LBP descriptor with a ternary coding style and proposed to incorporate a histogram of relative intensities in their work. Therefore, their proposed descriptor captures both local orders as well as overall distribution of pixel orders in the entire patch.

Besides the low-level features (e.g. the gradient orientation in SIFT, LBP in CS-LBP) which are used for descriptor construction, choosing an optimal support region size is also critical for feature description. Some researchers have found that a single support region is not enough to distinguish some incorrect matches from correct ones [3, 16]. In [16], Mortensen et al. proposed to combine the SIFT with global context to improve the performance of SIFT especially when there exist repeated textures in the matching images. In their method, the global context is computed from curvilinear shape information in a much larger neighborhood. Thus, by incorporating global context to local descriptor, the discriminative ability of their descriptor is improved and hence can disambiguate the confusion induced by repetitive patterns to some extent. In [3], the authors proposed to use multiple support regions of different size to construct a feature descriptor that is robust to general image deformations. In their work, a SIFT descriptor is computed for each support region, then they are concatenated together to form their descriptor. They further proposed a similarity measure model, Local-to-Global Similarity model, to match points described by their descriptors.

Our work is fundamentally different from the previous ones. In our work, the gradient is calculated in a rotation invariant way. Thus it is theoretically rotation invariant. While previous methods, such as SIFT, GLOH, DAISY, CS-LBP, are not totally rotation invariant since they need to assign a dominant orientation for each interest point, but unfortunately the computation of dominant orientation is not reliable according to our experiments. The need of a dominant orientation is in fact a drawback and bottleneck of the previous methods that utilize multiple support regions, hence largely differentiates our method from them. Although there are local descriptors that are also theoretically rotation invariant, e.g. spin image [9], RIFT (Rotation

Invariant Feature Transform) [9], they are less distinctive since spatial information is discarded. On the contrary, our descriptor not only contains rotation invariant gradient information, but also encodes spatial information which is an important cue for discrimination. The spatial pooling strategy used in our work is an adaptive one while the previous methods pool low-level features either in rectangular grids or in polar grids. Once again, these pooling strategies need to assign a dominant orientation in advance in order to be rotation invariant. In our method, the pooling regions are decided by the intensity orders of sample points and so no dominant orientation is required for reference. What is more, the ordering information is encoded in our descriptor by such a pooling strategy.

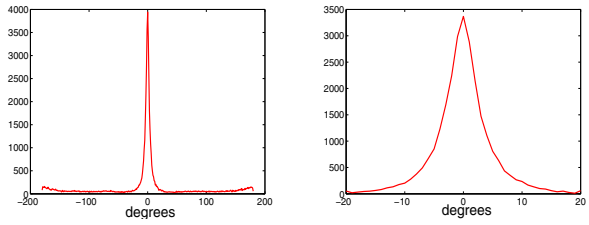
3. Descriptor Construction

3.1. Aggregating Gradient Distributions into Intensity Orders

In this work, we aggregate gradient distributions into intensity orders by a 2D histogram of gradient orientation and intensity order, in order to construct a robust and distinctive descriptor for a given support region. Such a 2D histogram not only contains gradient information but also encodes relative relationship of intensities. Moreover, spatial information is also encoded indirectly by pooling sample points according to their intensity orders. Theoretically, it is rotation invariant without the help of a dominant orientation, because both the computations of gradient and intensity ordering are rotation invariant in our work. Whereas, for the most of existing methods, their rotation invariance is achieved by assigning a dominant orientation to each interest point based on local image statistics [10]. Then the descriptor is constructed relative to the assigned orientation. Take the SIFT descriptor as an example, the partition of subregions and calculation of gradient are relative to the dominant orientation.

However, here we would claim that the dominant orientation assignment based on local image statistics is an error-prone process and we will experimentally show that the inevitable error in the orientation assignment will make many true corresponding points un-matchable by their descriptors. To assess this, we have collected 40 pairs of images with rotation transformation from the Internet[1], each of which is related by a homography and the homography is supplied along with the image pair. For each image pair, we extracted SIFT keypoints and matched them by the nearest neighbor of the distances of their SIFT descriptors. We focus on orientation assignment errors between corresponding points which satisfy the homography. Fig. 2 presents some statistical results on these 40 image pairs. Fig. 2(a) shows the orientation assignment errors between corresponding points. A similar histogram of orientation estimation errors between corresponding points was obtained in [22] through applying random synthetic affine warps. Here we use real image data with mainly rotation transformation. In Fig. 2(b), it shows the errors between those corresponding points that are also matched by their SIFT descriptors. Fig. 2(b) implies that for the SIFT descriptor, it requires the orientation assignment errors within 20 degrees in order to match corresponding points correctly. However, it can be clearly seen from Fig. 2(a) that there are only 63.77% corresponding points whose orientation assignment errors are within $[-20, 20]$. Thus many corresponding points whose orientation assignment errors are larger than 20 degrees would not be correctly matched by comparing their descriptors. In other words, 36.23% corresponding points will not be correctly matched mainly due to the incorrect orientation assignment. Therefore, orientation assignment has a significant impact on distinctive descriptor construction.

Instead of assigning a dominant orientation to each interest point, we propose to calculate the gradient of sample point in a rotation invariant way. Meanwhile, in order to encode spatial information which is an important factor to the discriminative power of the descriptor, we use intensity orders of sample points to adaptively pool gradient information into different groups. Since intensity orders of sample points are rotation invariant, such a spatial pooling strategy is also rotation invariant inherently and so no dominant orientation is required. Hence, our descriptor is completely rotation invariant. It can be seen from Fig. 1 that many corresponding points with large orientation assignment errors can be correctly matched by our proposed descriptor which are un-matchable by the SIFT descriptor due to incorrect orientation estimation.



(a) (b)

Figure 2. Orientation assignment errors. (a) Between corresponding points, only 63.77% of errors are in the range of [-20,20]. (b) Between corresponding points that are also matched by SIFT.

3.1.1 The Computation of Rotation Invariant Gradient

In Fig. 3, suppose P is an interest point and P_i is one of the sample points in its support region. Then a local $x - y$

[1] http://lear.inrialpes.fr/people/mikolajczyk/

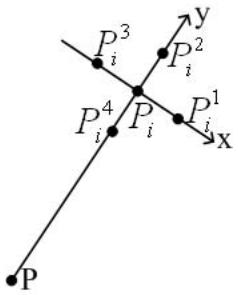


Figure 3. The computation of rotation invariant gradient.

coordinate system can be established by P and P_i for each sample point as shown in Fig. 3, where $\overrightarrow{PP_i}$ is defined as the positive y-axis. Obviously, such a coordinate system is rotation invariant. Therefore, in this coordinate system, the calculated gradient is rotation invariant too. Such a rotation invariant gradient can be computed as follows:

$$Dx(P_i) = I(P_i^1) - I(P_i^3) \quad (1)$$

$$Dy(P_i) = I(P_i^2) - I(P_i^4) \quad (2)$$

where $P_i^j, j = 1, 2, 3, 4$ are P_i's neighboring points along the x-axis and y-axis in the local $x - y$ coordinate system and $I(P_i^j)$ stands for the intensity at P_i^j.

Note that in RIFT [9], it uses a similar way to calculate the gradients. However, since it accumulates histogram of orientation in rings around the interest point to achieve rotation invariance, the spatial information is lost which results in less distinctiveness. In our work, spatial information is encoded into the descriptor by an adaptive pooling strategy, which will be described in the next subsection.

3.1.2 Adaptive Spatial Pooling based on Intensity Orders

Given a support region, one can divide it into several rings and pool together gradient information of sample points in each ring in a similar way as spin image or RIFT [9] does, so as to achieve a rotation invariant description of the region. However, such a method does not take into account the spatial information that is important for distinguishing different interest regions. In other words, pooling gradient information circularly achieves a rotation invariant representation at the cost of degrading the descriptor's discriminative power. Therefore, many popular and state-of-the-art methods divide the support region into subregions in order to take into considerations of the spatial information, such as SIFT [10], DAISY [20], CS-LBP [7], OSID [19] and so on [6, 16, 2]. Unfortunately, these pre-defined subregions need to assign a relative orientation in order to be rotation invariant. As we have said before, the orientation assignment is not stable enough, here we propose an adaptive strategy for pooling gradient information spatially. The proposed strategy is based on the intensity orders of sample points. Specifically, we first sort sample points in the support region according to their intensities. Then we divide them into k segments equally according to their orders. Finally, gradient information of the sample points in each segment are pooled, and the gradient orientation histograms in these k segments are concatenated to form the representation of this support region. Fig. 4 shows an example of spatial segmentations based on intensity orders.

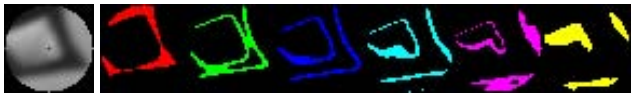


Figure 4. An example of spatial pooling: 6 pooling segments (indicated by different colors) based on intensity orders for a region.

To sum up, given a support region, firstly we calculate rotation invariant gradient for each sample point in the region according to Eq. (1) and Eq. (2). Secondly, sample points are sorted according to their intensities and they are divided into k segments according to their orders. Finally, gradient information are pooled together in each of the k segments. That is, in each of the k segments, a histogram of gradient orientation is accumulated. In the histogram, the gradient orientation of a sample point is linearly allocated to the two adjacent orientation bins according to its distances to them and also weighted by the gradient magnitude of this sample point. Suppose the number of orientation bins is d, then we can obtain a $d \times k$ vector as the representation for the given support region. The obtained vector is then normalized to counter illumination changes as SIFT does [10].

3.2. Multiple Support Regions

We believe that only one single support region is not enough to distinguish incorrect matches from correct ones in general cases. Generally speaking, two non-corresponding points may accidently have similar appearances in a certain local region. However, it is less likely that two non-corresponding interest points have similar appearances in several local regions of different sizes. In contrast, two corresponding interest points should have similar appearances in a local region of any size, although some differences may exist due to localization error of interest points. That is to say, with multiple support regions, it is much easier to distinguish whether two points are matched or not than using a single support region. Therefore, in our proposed method we utilize multiple support regions to construct the descriptor to further improve its discriminative ability.

In Fig. 5, it gives an overview of our proposed method. For an interest point (x, y), let $R_i(x, y)$ denotes its ith support region. Then a vector D_i can be obtained from $R_i(x, y)$ by aggregating the rotation invariant gradient distributions

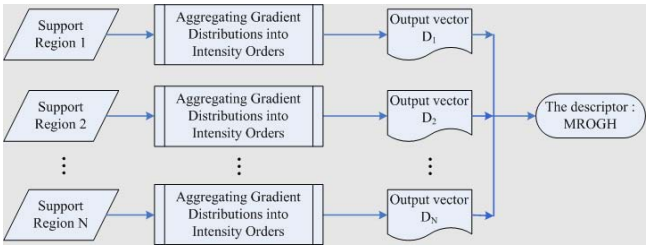


Figure 5. The proposed method for descriptor construction.

into intensity order as described in the proceeding subsection. Finally, those vectors of multiple support regions are concatenated to form our proposed descriptor, which we called MROGH, i.e. $\{D_1 D_2 \cdots D_N\}$ in which N is the number of support regions. In this paper, we choose support regions as the N nested regions centered at the interest point with an equal increment of size. All the support regions are affine normalized to an unified circular region of radius 20.5 for descriptor construction.

4. Experiments

4.1. Parameter Selection

The proposed descriptor has three parameters: the number of orientation bins *d*, the number of order segmentations *k* and the number of support regions *N*. In order to evaluate their influences on the performance of the proposed descriptor, we have conducted experiments on 142 pairs of images[2] with different parameter settings. These 142 image pairs are mainly selected from zoom and rotation transformations. Note that here we do not use image pairs in the standard Oxford dataset[3] because those image pairs will be used for the descriptor's performance evaluation later on.

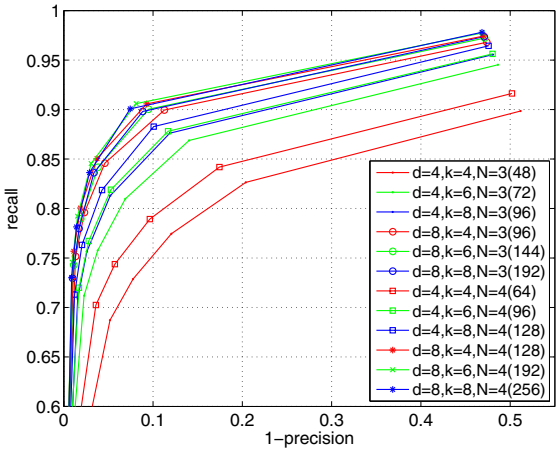


Figure 6. Parameters evaluation of the proposed descriptor.

Fig. 6 shows the curves of *average recall* vs. *average 1-precision* for different parameter settings. The definition of a correct match and a correspondence is the same as [13] which is determined with the overlap error [12]. The matching strategy we used here is the nearest neighbor distance ratio [13]. In the remaining experiments, we used the same definition of a match, a correct match and a correspondence unless otherwise specified.

It can be seen from Fig. 6 that the performance of the proposed descriptor is improved with the increase of the number of orientation bins, the number of order segments as well as the number of support regions. This could be that with more bins, more information can be captured by the descriptor. Thus higher performance is expected. Among all these plots, it is clear that both the settings of $\{d = 8, k = 8, N = 4\}$ and $\{d = 8, k = 6, N = 4\}$ give the best performance, but the dimension of the latter one is much less, i.e. 192 vs. 256. Therefore, we set orientation bins to 8, the number of order segments to 6 and the number of support regions to 4 for our subsequent experiments.

4.2. Multi-Support Regions vs. Single Support Region

This experiment aims to show the superiority of using multi-support regions to a single support region. We used the same dataset as in the experiment of parameter selection. As said in the previous subsection, we use 4 support regions to construct our descriptor, obviously we can also calculate a descriptor for each of the 4 used support regions respectively. Thus we have 5 respective descriptors in total for one interest point: **SR-i** denotes the descriptor calculated with the information in the ith support region and **MR** is the descriptor concatenating **SR-1**, **SR-2**, **SR-3**, **SR-4**. For each image pair, we respectively extracted these 5 kinds of descriptors to perform point matching and obtained the curves of *average recall* vs. *average 1-precision* as before. The comparative results are shown in Fig. 7. It can be seen from Fig. 7 that the performance of **SR-i** improves with the increasing size of support region, mainly because that more information can be captured by a larger support region. As expected, by combining multiple support regions for descriptor construction, the performance of our proposed descriptor has a significant improvement over the best performance when using a single support region, c.f. the curves of **MR** and **SR-4**. For comparison, the performance of SIFT is also included.

4.3. Performance Evaluation on the Oxford Dataset

To evaluate the performance of the proposed descriptor, we have tested it on the Oxford dataset which is widely used for local descriptors evaluation. We followed the evaluation procedure proposed by Mikolajczyk and Schmid [13]. The codes for evaluation are downloaded from their website[4]

[2] They can be downloaded from http://lear.inrialpes.fr/people/mikolajczyk/

[3] http://www.robots.ox.ac.uk/~vgg/research/affine/

[4] http://www.robots.ox.ac.uk/~vgg/research/affine/desc_evaluation.html

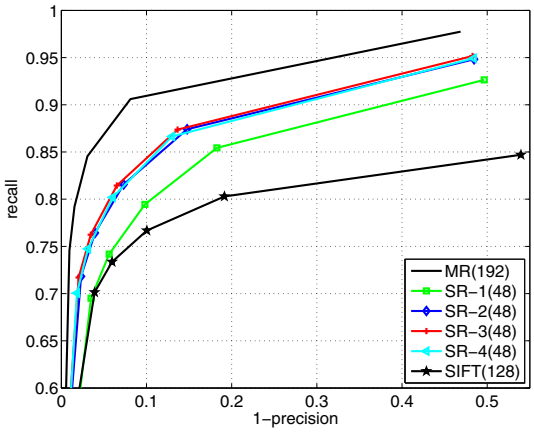


Figure 7. Performance comparison between multi-support regions and each single support region.

too. Two other state-of-the-art local descriptors are also evaluated in our experiments for comparison: SIFT [10] and DAISY [20, 23]. In our experiments, the implementation of SIFT is downloaded from the Oxford University website which is the same as the one used by Mikolajczyk and Schmid [13] while the DAISY is our own implementation according to the optimal parameters provided by the authors. The **T1-8-2r8s** configuration [23] of DAISY is used, whose dimension is 136. Since the proposed descriptor does not need to estimate orientation, its computational cost is lower than SIFT and DAISY in a single support region. Generally speaking, its computational cost is proportional to the number of the used regions (4 in this paper). In our experiments (Intel Core2 CPU 1.86GHz), the time of constructing descriptor for a feature point is: 12.3ms for DAISY, 4.6ms for SIFT, 8.3ms for MROGH, 1.9 ms for the MROGH with a single region.

As in [13], we have evaluated the performance of the descriptors using both Hessian-Affine and Harris-Affine detectors [12]. Fig. 8 shows the experimental results on images under various transformations, including viewpoint changes, rotation and scale changes, illumination changes, image blur and JPEG compression. In Fig. 8, **MROGH** is our proposed local descriptor indicated by red plots, while the results of **DAISY** are indicated by green plots and the results of **SIFT** are marked by blue plots. Results using Hessian-Affine detector are plotted by solid curves while the dashed curves are the results using Harris-Affine detector. It can be found that although the performance of each descriptor varied with different feature detectors, the relative performance among different descriptors is more or less consistent. Obviously, our proposed descriptor outperforms SIFT and DAISY in all cases, either for Hessian-Affine regions or for Harris-Affine regions. Such a good performance of our proposed descriptor may attribute to its well-designed properties. It not only uses multiple support regions to improve its discriminative ability, but also encodes local gradient information as well as ordering and spatial information. Moreover, the gradients are calculated in a rotation invariant way which further improves its robustness.

4.4. Performance Evaluation based on 3D Objects

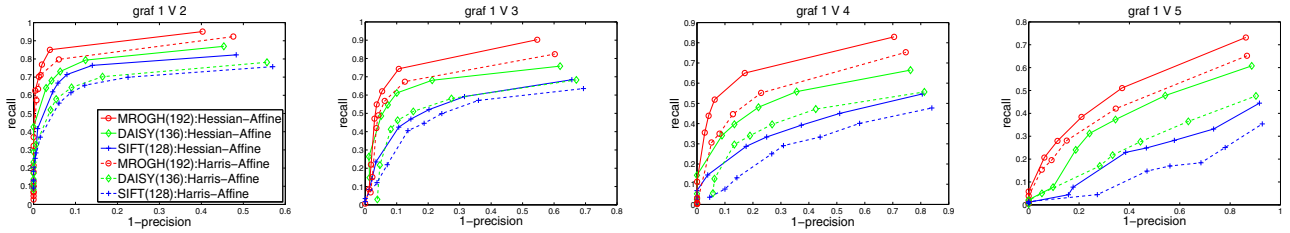
In [15], Moreels and Perona have evaluated different combinations of feature detectors and descriptors based on 3D objects. We have downloaded the database from their website[5] and evaluated our proposed descriptor following their work. In order to mimic the process of image retrieval/object recognition, interest point matching is conducted on a database containing both the target features and a large amount of features from unrelated images. In our experiments, 10^5 features are randomly chosen from 500 unrelated images obtained from Google by typing 'things' as in [15]. Please refer to [15] for more details about the dataset and experimental setup.

In this experiment, the Hessian-Affine detector is used for feature detection since it has been reported with the best results combined with SIFT descriptor in [15]. Fig. 9 shows the comparative results of the evaluated descriptors. In Fig. 9(a), the ROC curves of 'detection rate vs. false alarm rate' are obtained by varying the threshold of nearest neighbor distance ratio which defines a match. The detection rate is the number of detections against the number of tested matches, while the false alarm rate is the number of false alarms divided by the number of tested matches. A tested match is classified as a non-match, a false alarm or a correct match (a detection) according to the distances between descriptors and whether the geometric constraints are satisfied or not [15]. In Fig. 9(b), it shows the detection rate as a function of the viewpoint changes at a fixed false alarm rate. The false alarm rate 10^{-6} is chosen which implies that one false alarm over every 10 attempts since the false alarm rate is normalized by the number of database features (10^5). From Fig. 9 we can see that our proposed descriptor outperforms the other tested descriptors.

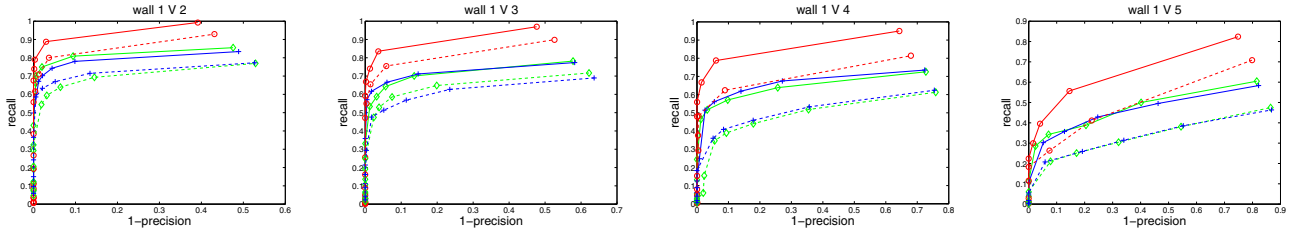
5. Conclusion

This paper presents a novel local image descriptor, which has the following nice features:
(1) Unlike the undertaking in many popular descriptors where a dominant orientation is assigned to the descriptor for it to be rotation invariant, our descriptor is inherently rotation invariant thanks to a rotation invariant way of gradient computation.
(2) To encode spatial information and take into consideration of intensity distributions, sample points are segmented based on their intensity orders, rather than their geometric locations. Thus no orientation is required for reference.

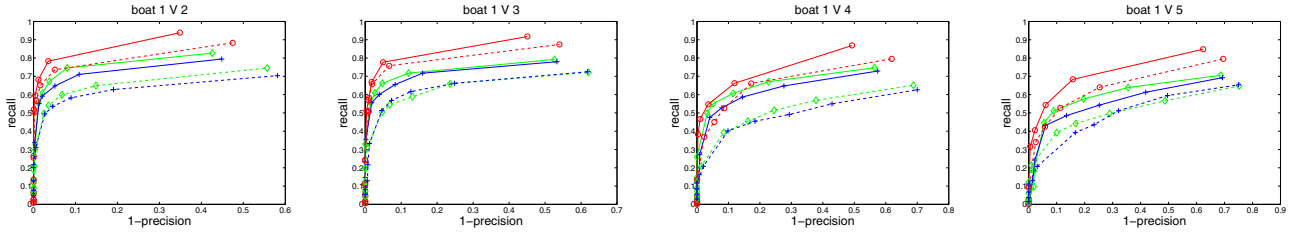
[5] http://www.vision.caltech.edu/pmoreels/Datasets/TurntableObjects



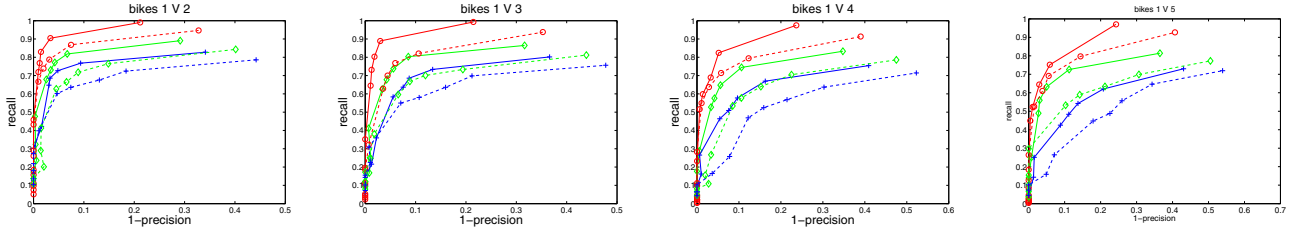
(a) viewpoint changes (graf)



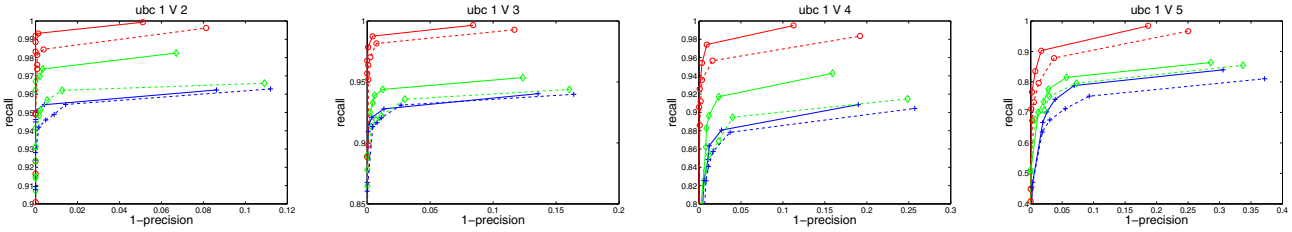
(b) viewpoint changes (wall)



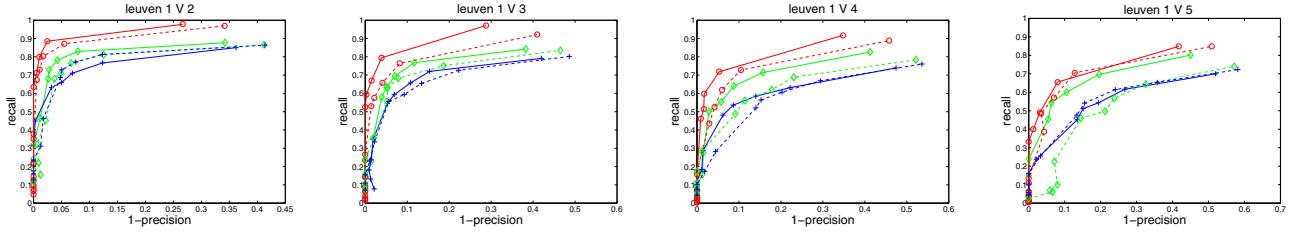
(c) rotation and scale changes (boat)



(d) image blur (bikes)

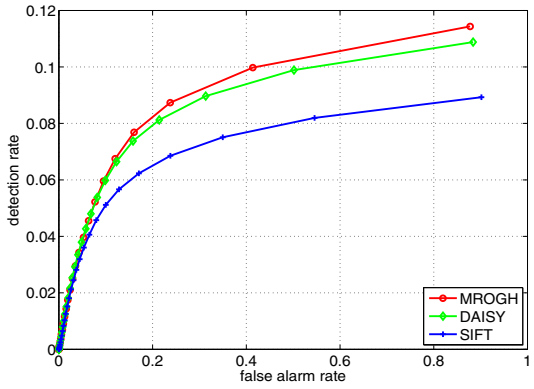


(e) JPEG compression (ubc)

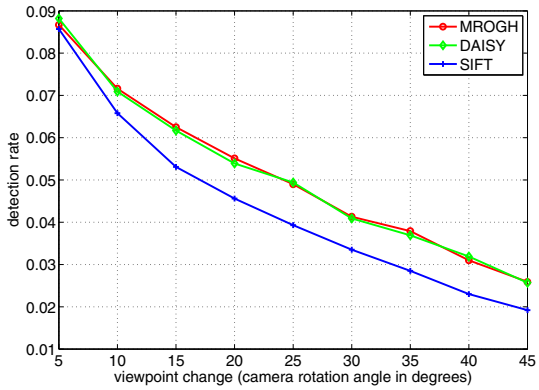


(f) illumination changes (leuven)

Figure 8. Experimental results under various image transformations.



(a) Detection Rate vs. False Alarm Rate



(b) Detection Rate vs. Viewpoint Changes at false alarm rate of 10^{-6}

Figure 9. Experimental results on dataset of 3D objects.

(3) Gradient distributions are pooled within such order segments, rather than in fixed subregions.
(4) It uses multiple support regions to further improve its discriminative ability.
Experimental results on two popular datasets have shown that our proposed descriptor outperforms many state-of-the-art methods under various image transformations.

6. Acknowledgements

This work is supported by the National Science Foundation of China (60835003, 61075038).

References

[1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building rome in a day. In *Proc. ICCV*, pages 72–79, 2009.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[3] H. Cheng, Z. Liu, N. Zheng, and J. Yang. A deformable local image descriptor. In *Proc. CVPR*, 2008.

[4] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *PAMI*, 32(8):1362–1376, 2010.

[5] R. Gupta and A. Mittal. SMD: A locally stable monotonic change invariant feature descriptor. In *Proc. ECCV*, pages 265–277, 2008.

[6] R. Gupta, H. Patil, and A. Mittal. Robust order-based methods for feature description. In *Proc. CVPR*, pages 334–341, 2010.

[7] M. Heikkila, M. Pietikainen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42:425–436, 2009.

[8] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Proc. CVPR*, volume I, pages 511–517, 2004.

[9] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *PAMI*, 27:1265–1278, 2005.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[11] J. Matas, O. Chum, M. Urba, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC*, pages 384–396, 2002.

[12] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proc. ECCV*, pages 128–142, 2002.

[13] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005.

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, 2005.

[15] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3D objects. *IJCV*, 73(3):263–284, 2007.

[16] E. Mortensen, H. Deng, and L. Shapiro. A SIFT descriptor with global context. In *Proc. CVPR*, volume 1, pages 184–190, 2005.

[17] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (TOG)*, 25:835–846, 2006.

[18] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2:1–104, 2006.

[19] F. Tang, S. H. Lim, N. L. Change, and H. Tao. A novel feature descriptor invariant to complex brightness changes. In *Proc. CVPR*, pages 2631–2638, 2009.

[20] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *Proc. CVPR*, 2008.

[21] T. Tuytelaars and L. V. Gool. Matching widely separated views based on affine invariant regions. *IJCV*, 59(1):61–85, 2004.

[22] S. Winder and M. Brown. Learning local image descriptors. In *Proc. CVPR*, pages 1–8, 2007.

[23] S. Winder, G. Hua, and M. Brown. Picking the best DAISY. In *Proc. CVPR*, pages 178–185, 2009.

[24] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: a comprehensive study. *IJCV*, 73(2):213–238, 2007.