Local Intensity Order Pattern for Feature Description

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

Abstract

This paper presents a novel method for feature description based on intensity order. Specifically, a Local Intensity Order Pattern(LIOP) is proposed to encode the local ordinal information of each pixel and the overall ordinal information is used to divide the local patch into subregions which are used for accumulating the LIOPs respectively. Therefore, both local and overall intensity ordinal information of the local patch are captured by the proposed LIOP descriptor so as to make it a highly discriminative descriptor. It is shown that the proposed descriptor is not only invariant to monotonic intensity changes and image rotation but also robust to many other geometric and photometric transformations such as viewpoint change, image blur and JEPG compression. The proposed descriptor has been evaluated on the standard Oxford dataset and four additional image pairs with complex illumination changes. The experimental results show that the proposed descriptor obtains a significant improvement over the existing state-of-the-art descriptors.

1. Introduction

Local image features have been widely used in many computer vision applications such as object recognition [13] , texture recognition [12], wide baseline matching [24], image retrieval [18] and panoramic image stitching [3]. The basic idea is to first detect interest points or interest regions and then compute invariant feature descriptors on each of them. Once the feature descriptors are computed, the feature correspondences between different images can be automatically established under some similarity measure, *e.g*. the Euclidean distance.

Many methods have been proposed to detect interest points or interest regions that are covariant with a class of transformations(*e.g*. affine transformation). For example, Harris corner [10] and DoG(Difference of Gaussian) [13] for interest point detection, and Harris-affine [15], Hessian-affine [17], MSER(Maximally Stable Extremal Region) [14], IBR(Intensity-Based Region) and EBR(Edge-Based Region) [24] for affine covariant region detection.

Concerning feature description, a number of methods have been proposed in the literature. The most popular methods are those based on histograms. For example, SIFT(Scale Invariant Feature Transform) [13], GLOH(Gradient Location-Orientation Histogram) [16] and DAISY [23] create a histogram of gradient orientations and locations, spin image [12] creates a histogram of pixel locations and intensities, and shape context [2] creates a histogram of edge point locations and orientations. This kind of descriptors usually obtains better performance than other kinds of descriptors such as filter-based descriptor [5, 20], derivative-based descriptor [21] and moment-based descriptor [6] *etc*. However, while the above descriptors have been shown to be fully or partially robust to many of the variations and distortions, they can not handle more complex illumination changes including gamma correction, small specular reflections, changes in exposure time *etc*. Recently, to alleviate the problem of complex illumination change, some researchers have proposed to use the intensity order rather than the raw intensities for feature description. Gupta and Mittal [8] presented a new feature descriptor by calculating a weighted sum of the order flips between point pairs chosen from the extremal regions. Tang *et al*. [22] created a 2D histogram encoding both the ordinal distribution and the spatial distribution. Heikkila *et al*. [11] proposed a CS-LBP descriptor which combines the strength of the SIFT descriptor and the LBP [19] texture operator. Goswami *et al*. [7] proposed another LBP-based method which computed pairwise ordinal information from adjacent circular neighborhoods. Gupta *et al*. [9] presented a more robust method which contains two part: a histogram of relative intensities and a histogram of CS-LTP codes. These methods generally obtained good performance vis-a-vis illumination changes.

In this paper, we propose a novel method for feature description based on intensity order. The basic principle of the proposed method is that the relative order of pixel intensities remains unchanged when the intensity changes are monotonic. In order to effectively exploit the ordinal infor-

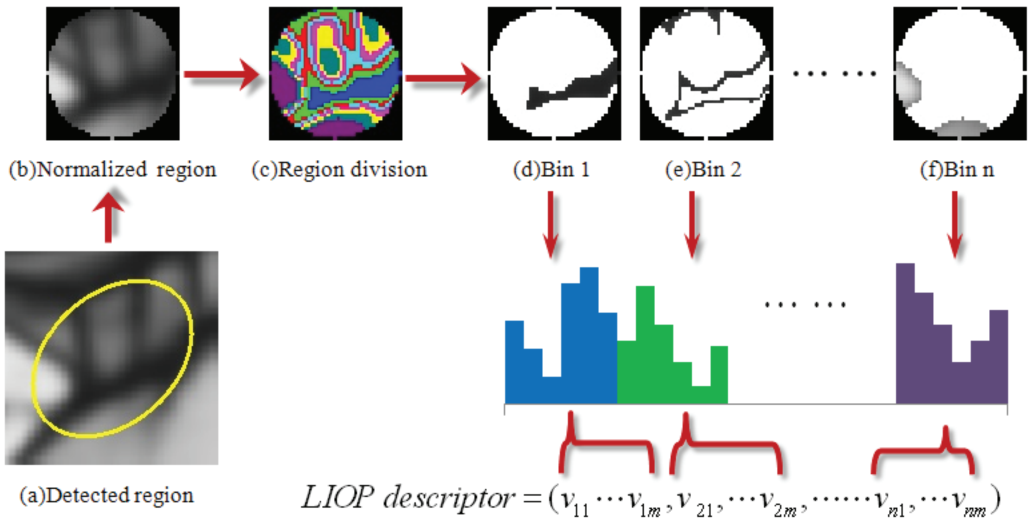


Figure 1. The workflow of our method.

mation, a Local Intensity Order Pattern(LIOP) is proposed to encode the local ordinal information and the overall ordinal information is used to divide the local patch into subregions which are used for accumulating the LIOPs respectively. Since both the region division and LIOP computation are based on the relative relationships of intensities, the proposed descriptor is inherently invariant to image rotation and monotonic intensity changes. According to the experimental results, it is also robust to many other geometric and photometric transformations such as view point change, image blur and JPEG compression.

The rest of this paper is organized as follows: Section 2 gives a detailed description of our method, the experimental evaluation is carried out in Section 3, and finally we conclude the paper in Section 4.

2. Our Method

We use the ordinal information in a novel way for descriptor construction. Firstly, the overall intensity order is used to divide the local patch into subregions called ordinal bins. Next, a Local Intensity Order Pattern (LIOP) of each point is defined based on the relationships among the intensities of its neighboring sample points. The LIOP descriptor is constructed by accumulating the LIOPs of points in each ordinal bin respectively, then by concatenating them together. The workflow of out method is shown in Figure 1.

2.1. Pre-processing, Feature Detection and Normalization

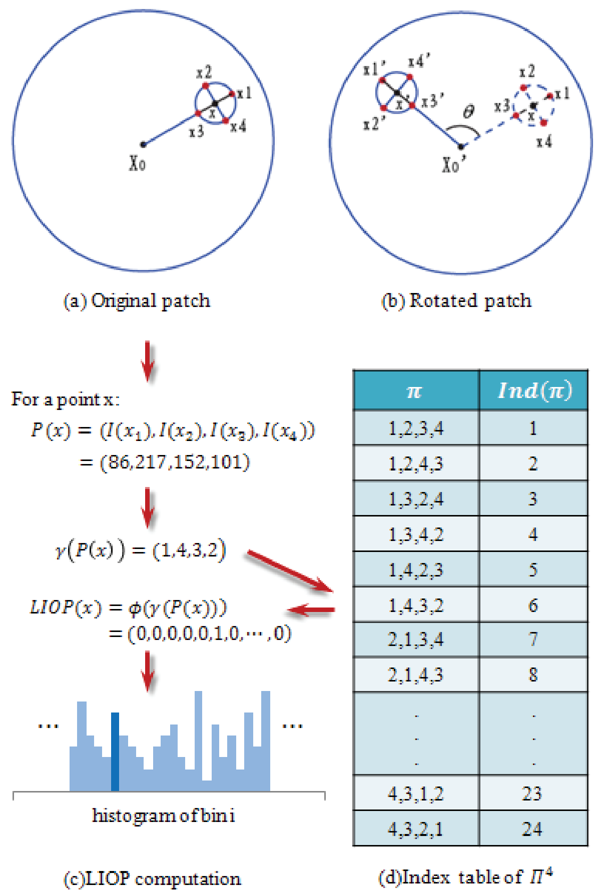
First, the image is smoothed by a Gaussian filter with sigma σ_p since the relative order is sensitive to noise. Then, an affine covariant region detector(such as Harris-Affine or Hessian-Affine) is used to localize the feature position and estimate the affine shape of its neighborhood. Since the detected regions usually have varying sizes and shapes, they are normalized to circular regions of a fixed diameter for feature description. Finally, a Gaussian smoothing with sigma σ_n is carried out again to remove noise which is introduced by interpolation in the normalization step. In this work, we refer to this resulting patch as a local patch.

It is worth noting that different from other methods such as [13, 11, 22, 9], we do not rotate the local patch according to the local consistent orientation(*e.g.* the dominant gradient orientation suggested by Lowe [13]) to achieve rotation invariance. The proposed descriptor is constructed in an orientation independent way which makes it inherently invariant to rotation. Detailed discussion will be presented in the following subsections.

2.2. Region Division

In order to improve the distinctiveness, the histogram-based methods usually divide the local patch into several subregions, and the descriptors are constructed by creating a histogram over each subregion respectively, then by concatenating them together.

Most of the previous methods for region division is based on the spatial location. For example, SIFT quantizes the spatial location into a 4×4 squared gird and GLOH uses a log-polar location grid with 3 bins in the radial direction and 8 in the angular direction. The disadvantage of such methods is that they have to estimate a local consistent orientation for each local patch and construct the descriptor relative to this orientation to achieve rotation invariance. Therefor, the performance of such methods heavily depends on the accuracy of the local consistent orientation estimation which



π	$Ind(\pi)$
1,2,3,4	1
1,2,4,3	2
1,3,2,4	3
1,3,4,2	4
1,4,2,3	5
1,4,3,2	6
2,1,3,4	7
2,1,4,3	8
⋮	⋮
4,3,1,2	23
4,3,2,1	24

Figure 2. The construction of the proposed LIOP descriptor.

is usually not robust to noise and distortion [4, 25]. To avoid the local consistent orientation estimation, spin image [12] divides the local patch into 5 rings. However, since it only quantizes the spatial location in radial direction, its discriminative power is lower than the grid-shaped region division.

In our work, we use the intensity order based region division proposed by Fan [4]. Specifically, all the pixels in the local patch are first sorted by their intensities in a non-descending order. Then, the local patch is equally quantized into B ordinal bins according to their orders. Figure 1(c) gives an illustration of such an intensity order based region division where each ordinal bin is marked with a different color. Note that it is not only invariant to monotonic intensity changes and image rotation, but also contains much more spatial information than the ring-shaped region division

2.3. Local Intensity Order Pattern Descriptor

The local information to be used for feature description varies from method to method. For example, spin image creates the histogram of intensity distribution, SIFT and GLOH create the histogram of gradient orientation. More recently, the local binary pattern [19] based methods are proposed and achieve high performance comparable to SIFT. CS-LBP [11] creates a histogram of central symmetric local binary pattern and CS-LTP [9] creates a histogram of central symmetric ternary pattern. Since the CS-LBP and CS-LTP only compare the intensities of central symmetric neighboring sample points, they do not effectively capture the relationships among the intensities of the neighboring sample points. In addition, they also need to sample the neighboring points relative to the local consistent orientation to achieve rotation invariance, which makes them vulnerable to the orientation estimation errors.

To overcome these problems, here we propose a novel Local Intensity Order Pattern(LIOP) for feature description. It effectively exploits the local information by using the intensity order of all the sampled neighboring points. What is more, it makes use of a rotation invariant sampling to avoid the estimation of the local consistent orientation. Thus, higher discriminative power is expected.

2.3.1 The Definition of LIOP

Before the formal definition of LIOP, some mappings are introduced. Let $\mathcal{P}^N = \{(p_1, p_2, \cdots, p_N) : p_i \in \mathbb{R}\}$ be the set of N-dimensional vectors and Π^N the set of all possible permutations of integers $\{1, 2, \cdots, N\}$, the mapping $\gamma : \mathcal{P}^N \longrightarrow \Pi^N$ is defined to map an N-dimensional vector $P \in \mathcal{P}^N$ to a permutation $\pi \in \Pi^N$ based on the order of the N elements of P. More specifically, the mapping γ sorts the N elements of P into a non-descending order, $p_{i_1} \leq p_{i_2} \leq \cdots \leq p_{i_N}$, and uses the subscript list $(i_1, i_2, \cdots, i_N)$ as the permutation. To avoid ambiguity, we define $p_s \leq p_t$ if and only if $(i) p_s < p_t$, or $(ii) p_s = p_t \; but \; s < t$. Mathematically, the mapping γ is defined as:

$$\gamma(P) = \pi, \quad P \in \mathcal{P}^N, \pi \in \Pi^N \tag{1}$$

where $\pi = (i_1, i_2, \cdots, i_N)$ and $p_{i_1} \leq p_{i_2} \leq \cdots \leq p_{i_N}$.

Since there are a total of $N!$ permutations in Π^N, the mapping γ divides the set $\mathcal{P}^N$ into $N!$ partitions and each partition corresponds to a unique permutation. For a permutation $\pi \in \Pi^N$, its corresponding partition of $\mathcal{P}^N$ is:

$$S(\pi) = \{P : \gamma(P) = \pi, \; P \in \mathcal{P}^N\} \tag{2}$$

According to this definition, the following equivalence relation holds:

$$\begin{array}{c} P, P' \in S(\pi) \\ \pi = (i_1, i_2, \cdots, i_N) \end{array} \iff \begin{array}{c} p_{i_1} \leq p_{i_2} \leq \cdots \leq p_{i_N} \\ p'_{i_1} \leq p'_{i_2} \leq \cdots \leq p'_{i_N} \end{array} \tag{3}$$

It means that the N-dimensional vectors which are in the same partition have the same order relationship among their N elements, and vice versa.

The partitions of $\mathcal{P}^N$ can be encoded by setting up an index table of all the possible permutations in Π^N due to the one-to-one correspondence between the partition and the permutation. Figure 2(d) shows such an index table in the case of $N = 4$. With the index table, a feature mapping function ϕ is defined to map a permutation π to an $N!$-dimensional feature vector $V_{N!}^{i}$ whose elements are all 0 except for the i-th element which is 1. The mathematical definition of ϕ is:

$$\phi(\pi) = V_{N!}^{Ind(\pi)},\ \pi \in \Pi^N \tag{4}$$

where $Ind(\pi)$ is the index of π in the index table and $V_{N!}^{Ind(\pi)} = (0, \cdots, 0, \underset{(Ind(\pi))}{1}, 0, \cdots, 0)$.

With the above definitions, let $P(x)$ be an N-dimensional vector which consists of the intensities of N neighboring sample points of a point x in the local patch, the LIOP of the point x can be defined as:

$$\begin{aligned} LIOP(x) &= \phi(\gamma(P(x))) \\ &= V_{N!}^{Ind(\gamma(P(x))} \\ &= (0, \cdots, 0, \underset{(Ind(\gamma(P(x))))}{1}, 0, \cdots, 0) \end{aligned} \tag{5}$$

where $P(x) = (I(x_1), I(x_2), \cdots, I(x_N)) \in \mathcal{P}^N$ and $I(x_i)$ denotes the intensity of the i-th neighboring sample point x_i. Since there are a total of $N!$ different LIOPs, the local patch is divided into $N!$ partitions, each of which is represented by an LIOP.

The N neighboring sample points of the point x are equally distributed on a circle of radius R centered at x. To obtain a rotation invariant sampling, the first point is sampled along the radial direction which is from the center of the local patch to the point x. Since there are two points along the radial direction on the circle, the one which is farther from the center of the local patch is selected as the first sample point. Then, the rest $N - 1$ points are sampled on the circle in an anticlockwise direction. An example of such a rotation invariant sampling in the case of $N = 4$ is shown in Figure 2(a) and (b). It can be seen that the 4 neighboring sample points x_1, x_2, x_3 and x_4 of the point x remain the same in the rotated patch, *i.e.* x'_1, x'_2, x'_3 and x'_4 respectively.

2.3.2 Descriptor Construction

The descriptor is constructed by accumulating the LIOPs of points in each ordinal bin respectively, then by concatenating them together. The construction of the LIOP descriptor is shown in Figure 2. Mathematically, the LIOP descriptor of the local patch is computed as:

$$\begin{gathered} LIOP\ descriptor = (des_1, des_2, \cdots, des_B) \\ des_i = \textstyle\sum_{x \in bin_i} LIOP(x) \end{gathered} \tag{6}$$

where B is the number of the ordinal bins. The dimension of the descriptor is $N! \times B$.

It is worth noting that the LIOP descriptor is both invariant to monotonic intensity changes and image rotation. Suppose x denotes a point in the local patch, x' denotes the point x after a monotonic intensity change and image rotation, $P(x) = (I(x_1), I(x_2), \cdots, I(x_N))$ and $P(x') = (I(x'_1), I(x'_2), \cdots, I(x'_N))$ are their N-dimensional vectors respectively. Since the intensity order of $I(x_1)$, $I(x_2)$, $\cdots$, $I(x_N)$ and $I(x'_1)$, $I(x'_2)$, $\cdots$, $I(x'_N)$ are the same under monotonic intensity changes, $P(x)$ and $P(x')$ will belong to the same partition of $\mathcal{P}^N$ according to the equivalence relation shown in Eq.(3). In other words, $\gamma(P(x)) = \gamma(P(x'))$. Thus, we obtain $LIOP(x) = LIOP(x')$ according to Eq.(5). As discussed before, the intensity order based region division is invariant to monotonic intensity changes and image rotation. Thus, the point x and x' will belong to the same ordinal bin, which makes the the LIOP descriptor unchanged.

Since the order of similar intensities is less reliable than that of dissimilar ones due to Gaussian noise, the LIOP of point which has more dissimilar neighboring sample points is more stable and should be given a larger weight. A weighting function is proposed to improve the robustness of the LIOP descriptor, which is defined as:

$$w(x) = \sum_{i,j} \mathrm{sgn}(|\mathrm{I(x_i)} - \mathrm{I(x_j)}| - \mathrm{T_{lp}}) + 1 \tag{7}$$

where $\mathrm{sgn}()$ is the sign function and T_{lp} is a preset threshold. This weighting function measures the intensity dissimilarities among the neighboring sample points of point x by counting the number of dissimilar sample pairs. Thus, the descriptor becomes:

$$\begin{gathered} LIOP\ descriptor = (des_1, des_2, \cdots, des_B) \\ des_i = \textstyle\sum_{x \in bin_i} w(x) LIOP(x) \end{gathered} \tag{8}$$

Experiments show that this weighting scheme performs better than the naive uniform weighting and Gaussian weighting(see Figure 5).

3. Experiments

3.1. Dataset and Evaluation Criterion

We have evaluated our descriptor on the standard Oxford dataset [1]. It contains images with different geometric and photometric transformations of structured and textured scenes. The six different transformations are: viewpoint change, scale change, image rotation, image blur, illumination change, and JPEG compression. In order to study in more detail about the performance of our descriptor to complex illumination changes, we captured two additional image pairs 'desktop' and 'corridor' which have even more drastic illumination changes(see Figure 3). We also synthesized two images by performing a square root and square

(a) desktop

(b) corridor

Figure 3. Two captured image pairs with drastic illumination changes.

(a) square (b) square root

Figure 4. Two synthesized images from the second image of Figure 3(a).

operation on the second image of 'desktop'. Such nonlinear transformations produce images with a monotonic intensity change. Figure 4 shows these two synthesized images.

We used the evaluation criterion proposed by Mikolajczyk and Schmid [16] which is based on the number of correct and false matches between two images. The matching strategy is the nearest neighbor distance ratio(NNDR) which declares a match if the distance ratio between the first and second nearest neighbors is below a threshold. The number of correct matches and ground truth correspondences is determined by the overlap error [17]. A match is correct if the overlap error < 0.5. The results are presented with *recall* versus *1-precision* curves:

$$recall = \frac{\#correct\ matches}{\#correspondences} \quad 1-precision = \frac{\#false\ matches}{\#all\ matches}$$

where $\#correspondences$ is the ground truth number of matches.

3.2. Parameters Selection

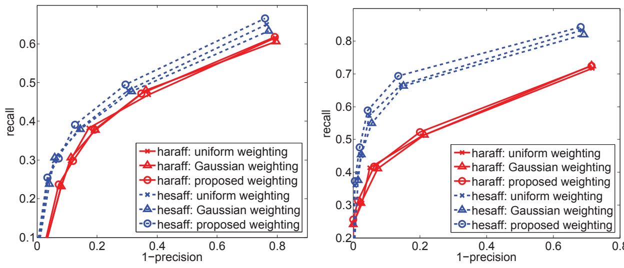
There are six parameters in our method: 1) the smoothing sigma σ_p before region detection, 2) the smoothing sigma σ_n after region normalization, 3) the number B of ordinal bins, 4) the number N of the neighboring sample points, 5) the sampling radius R, 6) the threshold T_{lp} of the weighting function.

Two image sequences of the standard Oxford dataset('graf' and 'wall') were used to investigage the effect of the parameters. We simply tried all combinations of these parameters and compared the matching performance of them. Due to the space limit, we only show the results between the 1^{st} and the 4^{th} images in Figure 6 by varying N(3 and 4) and B(4, 6 and 8). As can be seen, $N = 4$ got a better performance than $N = 3$ while both $B = 8$ and $B = 6$ got a better performance than $B = 4$. To obtain a smaller dimension, $B = 6$ is selected. Thus, the dimension of the proposed LIOP descriptor is $4! \times 6 = 144$. All the selected parameters are shown in Table 1 and kept unchanged in our subsequent experiments.

Parameters	σ_p	σ_n	N	B	R	T_{lp}
Values	1.0	1.2	4	6	6	5

Table 1. The selected parameters

We also compared the performance of the three weighting schema: uniform weighting, Gaussian weighting and the weighting function proposed in Eq.(7). It can be seen from Figure 5 that, the proposed weighting function obtains the best performance.

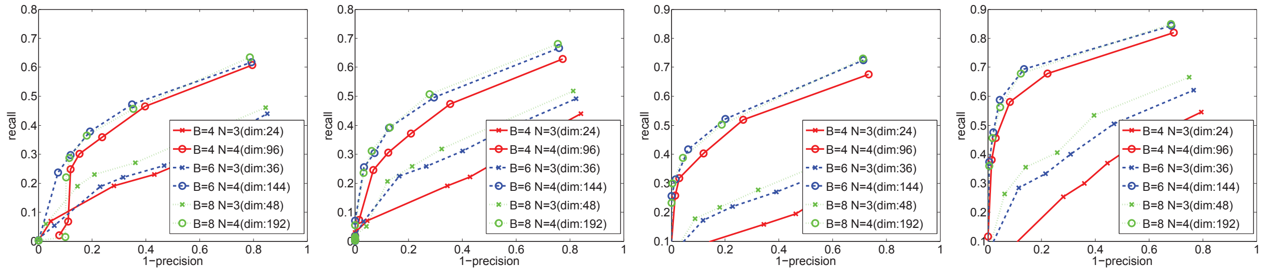


(a) graf 1-4 (b) wall 1-4

Figure 5. Performance comparison of the three weighting schema: uniform weighting, Gaussian weighting and the proposed weighting(Eq.(7)).Note that the scales are different for different figures to improve the clarity of the plot.

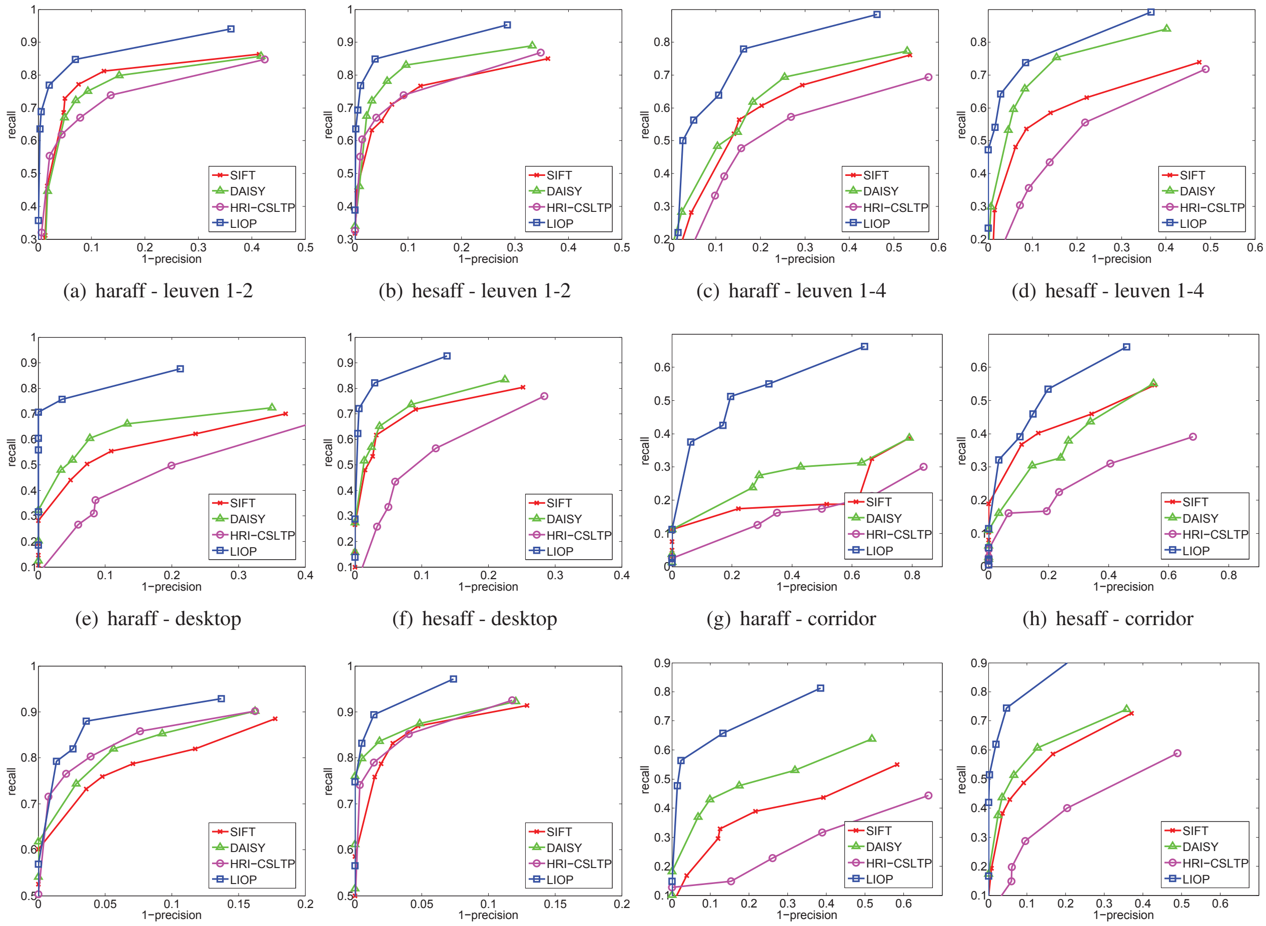
3.3. Performance Evaluation

Experiments have been carried out on five popular affine covariant regions: Harris-Affine(haraff), Hessian-Affine(hesaff), MSER, EBR and IBR. According to the results, the ranking of descriptors almost remains the same on different affine regions as indicated by [16]. Due to space limit, we only show the results on haraff and hesaff regions since they provide more interest regions than others. The hesaff detects blob-like structures while haraff detects corner-like structures, and both of them output elliptic regions of varying size. These elliptic regions are normalized to circularz regions of a fixed diameter as described in subsection 2.1. To make the comparison fair and representa-



(a) haraff - graf 1-4 (b) hesaff - graf 1-4 (c) haraff - wall 1-4 (d) hesaff - wall 1-4

Figure 6. Performance comparison of the LIOP descriptor on the Harris-Affine and Hessian-Affine regions under different parameter configurations, by varying the number of ordinal bins B and the number of neighboring sample points N. Note that the scales are different for different figures to improve the clarity of the plot.



(a) haraff - leuven 1-2 (b) hesaff - leuven 1-2 (c) haraff - leuven 1-4 (d) hesaff - leuven 1-4

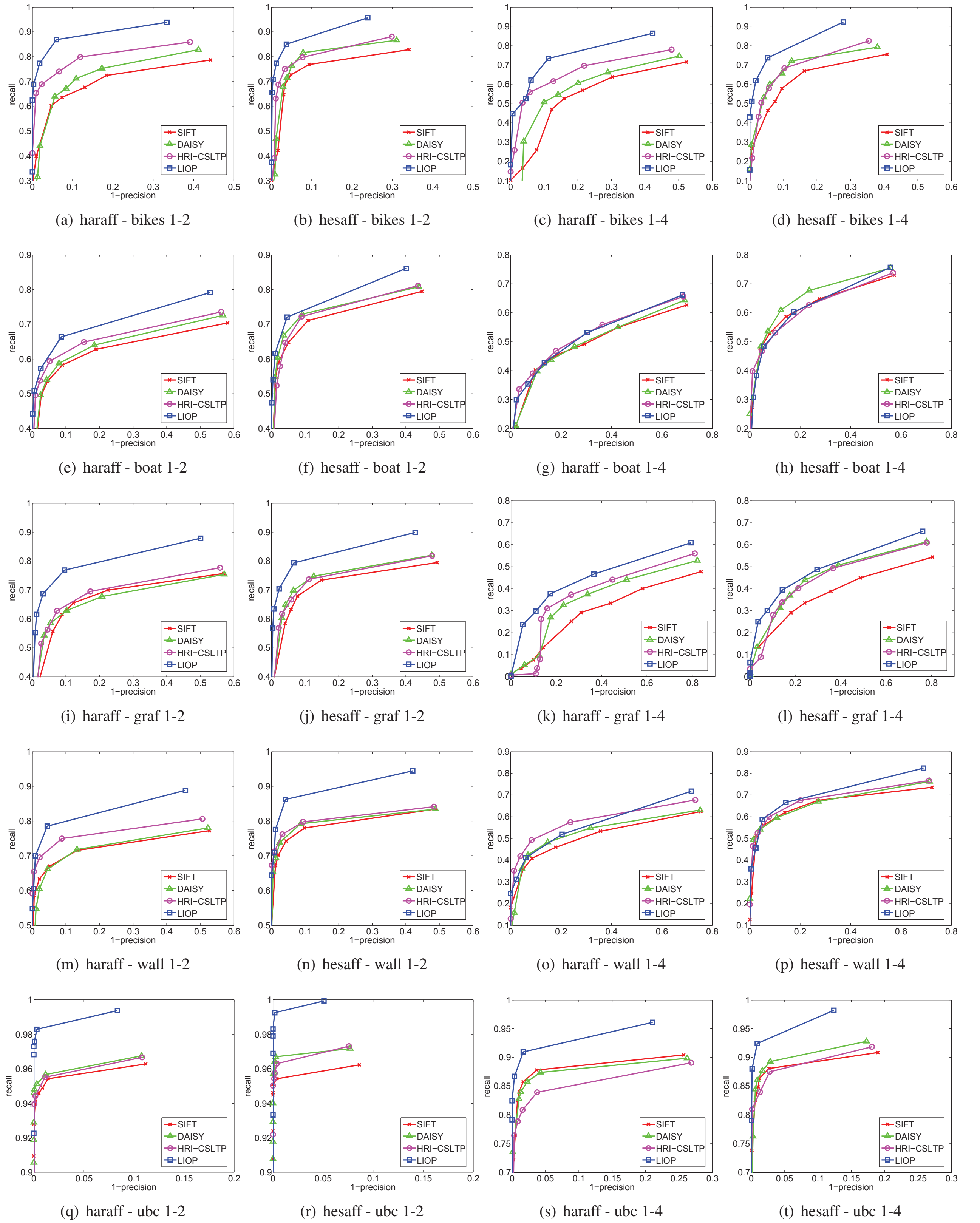
(e) haraff - desktop (f) hesaff - desktop (g) haraff - corridor (h) hesaff - corridor

(i) haraff - square (j) hesaff - square (k) haraff - square root (l) hesaff - square root

Figure 7. Experimental results for illumination changes. (a)-(d) are the results of 'leuven' in the Oxford database. (e)-(h) are the results of images shown in Figure 3. (i)-(l) are the results of synthesized images shown in Figure 4. Note that the scales are different for different figures to improve the clarity of the plot.

tive, we use the same diameter(41 pixels) as the evaluation paper [16].

We have compared our descriptor with SIFT [13], DAISY [23] and HRI-CSLTP [9] since SIFT and DAISY are the well known state-of-the-art feature descriptors while HRI-CSLTP is the recently proposed order based descriptor similar to ours. In our experiments (Intel Core2 Quad CPU 2.83GHz), the average time for constructing a feature descriptor is: 2.1ms for SIFT, 3.8ms for DAISY, 5.3ms for HRI-CSLTP and 5.5ms for LIOP. The evaluation results are



(a) haraff - bikes 1-2 (b) hesaff - bikes 1-2 (c) haraff - bikes 1-4 (d) hesaff - bikes 1-4

(e) haraff - boat 1-2 (f) hesaff - boat 1-2 (g) haraff - boat 1-4 (h) hesaff - boat 1-4

(i) haraff - graf 1-2 (j) hesaff - graf 1-2 (k) haraff - graf 1-4 (l) hesaff - graf 1-4

(m) haraff - wall 1-2 (n) hesaff - wall 1-2 (o) haraff - wall 1-4 (p) hesaff - wall 1-4

(q) haraff - ubc 1-2 (r) hesaff - ubc 1-2 (s) haraff - ubc 1-4 (t) hesaff - ubc 1-4

Figure 8. Experimental results for (i) image blur (a)-(d), (ii) image rotation and scale change (e)-(h), (iii) viewpoint change (i)-(p), and (iv) JPEG compression (q)-(t). Note that scales are different for different figures to improve the clarity of the plot.

shown in Figure 7 and Figure 8. For the Oxford database, we only show the results of two image pairs(the 1^{st} vs. the 2^{nd} and the 1^{st} vs. the 4^{th}) for each case here since they represent small and large image transformation respectively.

For illumination changes(see Figure 7), the LIOP descriptor performs consistently better than all the other tested descriptors in all cases. It significantly improves the performance on illumination changes due to the novel usage of intensity order. For other image transformations(see Figure 8), the LIOP descriptor also outperforms SIFT in all cases, and it outperforms DAISY and HRI-CSLTP in almost all cases except for *hesaff - boat 1-4*(Figure 8(h)) and *haraff - wall 1-4*(Figure 8(o)). As can be observed, the proposed LIOP descriptor obtains a high discriminative power while stays robust to many image transformations.

4. Conclusion

This paper proposed a novel Local Intensity Order Pattern(LIOP) for feature description. Compared with the previous proposed intensity order based methods, LIOP is quite different in the sampling strategy, comparison rule and encoding scheme. More specifically, it employs a rotation invariant sampling and fully explores the local intensity relationships by considering the intensity order among all the sample points. Meanwhile, a permutation-based encoding scheme is proposed to compress the dimension which makes LIOP more suitable for constructing local descriptor. By accumulating the LIOPs of points in each ordinal bin respectively, the descriptor is constructed totally based on the relative relationships of intensities, which makes it invariant to image rotation and monotonic intensity changes. Experimental results on various image transformations have shown that the proposed LIOP descriptor outperforms the state-of-the-art methods.

5. Acknowledgements

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

References

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

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24:509–522, 2002.

[3] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 74:59–73, 2007.

[4] B. Fan, F. C. Wu, and Z. Y. Hu. Aggregating gradient distributions into intensity orders: A novel local image descriptor. In *Proc. CVPR*, pages 2377–2384, 2011.

[5] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *PAMI*, 13:891–906, 1991.

[6] L. J. V. Gool, T. Moons, and D. Ungureanu. Affine/ photometric invariants for planar intensity patterns. In *Proc. ECCV*, pages 642–651, 1996.

[7] B. Goswami, C. H. Chan, J. Kittler, and B. Christmas. Local ordinal contrast pattern histograms for spatiotemporal, lip-based speaker authentication. In *Biometrics: Theory Applications and Systems (BTAS)*, pages 1–6, 2010.

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

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

[10] C. Harris and M. Stephens. A combined corner and edge detection. In *Proc. Alvey Vision Conference*, pages 147–151, 1988.

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

[12] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *PAMI*, 27(8):1265 – 1278, 2005.

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

[14] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC*, volume 1, pages 384–393, 2002.

[15] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60:63–86, 2004.

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

[17] 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*, 65:43–72, 2005.

[18] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proc. CVPR*, volume 2, pages 2161 – 2168, 2006.

[19] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 24:971–987, 2002.

[20] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or "how do i organize my holiday snaps?". In *Proc. ECCV*, pages 414–431, 2002.

[21] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19:530–535, 1997.

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

[23] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *PAMI*, 32:815–830, 2010.

[24] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *IJCV*, 59:61–85, 2004.

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