

Rotation and Illumination Invariant Interleaved Intensity Order Based Local Descriptor

Shiv Ram Dubey, *Student Member, IEEE*, Satish Kumar Singh, *Member, IEEE*, and Rajat Kumar Singh, *Member, IEEE*

Abstract— The region descriptors using local intensity ordering patterns have become more popular recent years for image matching due to its enhanced discriminative ability. However, the dimension of these descriptors increases rapidly with the slight increase in the number of local neighbors under consideration and becomes unreasonable for image matching due to time constraint. In this paper, we reduce the dimension of the descriptor and matching time significantly while keeping up the comparable performance by considering the number of neighboring sample points in an interleaved manner. The proposed interleaved order based local descriptor (IOLD) considers the local neighbors of a pixel as a set of interleaved neighbors and constructs the descriptor over each set separately and finally combines them to produce a single pattern. We extract the local ordering pattern to cope up with the illumination effect in an inherent rotation invariant manner. The novelty lies with using multiple neighboring sets in an interleaved fashion. We also explored the local intensity order pattern in a multi-support-region scenario. Results are compared over three challenging and widely adopted image matching datasets with other prominent descriptors under various image transformations. Results based on experiments suggest that the proposed IOLD descriptor outperforms in terms of both improved matching performance and reduced matching time. We also found that the amount of improvement is significant under complex illumination difference while showing more robustness towards noise.

Index Terms— Complex illumination change, Image matching, Intensity order, Local feature description, Interleaved descriptor, Rotation invariance.

I. INTRODUCTION

Computer vision researchers widely studied local feature descriptors constructed over the detected interest regions. Recent years, local features have been frequently used in large number of vision application problems such as 3D reconstruction, panoramic stitching, object recognition, image classification, facial expression recognition, and structure from motion [1-6]. The main focus while describing the local image features is to enhance the distinctiveness and maintain the robustness to the various image transformations. The basic goal is to first find the affine invariant interest regions and then extract feature pattern description for each of them. Hessian–Affine and Harris–Affine [7-8] detectors have been widely used for the extraction of interest regions. After detecting region of interest, feature descriptors are constructed

over it in order to facilitate the region matching. In the literature, many feature descriptors have been proposed with the increasing interest in region detectors [37], and it is observed that the performance of image descriptors based on some distributions are significantly better than the descriptors based on the spin image, shape context, and steerable filters [9-11]. The distributions of the gradient are widely used by distribution-based methods. For example, a histogram of gradient orientation on 4×4 location cells are computed by the SIFT descriptor [4]. Many other local image feature descriptors such as GLOH, SURF, and DAISY similar to SIFT have been introduced in the literature encouraged by the success of the SIFT description [12-15]. Some recent works involved Gaussian shapes as a feature descriptor [16], descriptor with face recognition applications under image blur condition [17], using alternate Hough and inverted Hough transforms for robust feature matching [18]. Although theoretically rotation invariant feature descriptions (i.e. Rotation Invariant Feature Transform (RIFT) and spin image [10]) are also exists in the literature, but these descriptors discard spatial information and becomes less distinctive. Exact orders are also utilized for image feature description by Kim et al. [19]. They combined the exact global and local orders to generate the EOD descriptor. Orthogonal LBPs are combined with color information to describe the image regions in [20].

Distribution-based descriptors are partially or fully robust to many of the geometric image transformations such as rotation, scale, occlusions etc., but can't handle more complex illumination changes. To ease this problem, some researchers have proposed to consider the orders of local intensities rather than the raw intensity values, because invariance to monotonic difference is being obtained by using the order of the intensity values. Some common order based descriptors are OSID, LBP, uniform LBP, CS-LBP, HRI, CS-LTP and LIOP [21-27]. A local binary pattern (LBP) creates the pattern for every pixel based on the ordering information [24]. The major benefit of LBP is its simplicity in the computation and also it's invariance to the illumination changes, but LBP has some drawbacks also such as computation of feature having more dimension and sensitivity to Gaussian noise in the uniform areas. Observing that small subset of the LBP contains the most of the textual information, a uniform LBP is proposed in [27]. The CS-LBP reduces the dimension of LBP by comparing only center-symmetric pixel intensity differences [25]. The CS-LTP descriptor is introduced by considering only diagonal comparisons among the neighboring points [23]. HRI and CS-LTP contains complementary information and it is combined to construct a single HRI-CSLTP descriptor [23]. Recently, Wang et al. [26] proposed LIOP descriptor to encode the intensity order pattern among the neighbors located at a fixed radius from a given pixel. They assigned a unique order to each neighbor and partitioned the whole patch into

Copyright (c) 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

The authors are with the Indian Institute of Information Technology Allahabad, India (e-mail: shivram1987@gmail.com, sk.singh@iitaa.ac.in, rajatsingh@iitaa.ac.in).

The final paper is available from: <http://dx.doi.org/10.1109/TIP.2014.2358879>

different regions according to the global ordering of each pixel in the patch and calculated the LIOP in each region separately and concatenated them in order to obtain a single pattern.

An intensity order based interleaved local descriptor (IOLD) is proposed in this paper which assumes the properties of the intensity orders in an interleaved manner. The neighborhood of each pixel is divided into a set of neighborhoods and patterns are computed by merging the patterns extracted for each neighboring set. In this way, low dimensional descriptor is obtained while considering a large number of neighbors which reduces the matching time. We also use multiple support regions and order based region partitioning of an interesting region to make proposed descriptor more robust and discriminative [28-29]. We also tested LIOP using multiple support regions and compared with IOLD. Results obtained from experiments suggest that under various image transformations, IOLD has the better discriminative ability with low time complexity.

The remainder of the paper is set up in the following manner: section II presents the detailed construction process of proposed IOLD descriptor; section III illustrates the detailed results of image matching experiments; section IV discusses the effect complex illumination and noisy condition; and section V draws the deductions with conclusions.

II. PROPOSED DESCRIPTOR CONSTRUCTION

We present the construction process of proposed IOLD descriptor in this section. First we address the steps involved in preprocessing, region detection and normalization, and then we discuss the concept of generating rotation invariant local features, and then we proposed the partitioning of local neighboring pixels into interleaved sets. The final pattern is generated by concatenating the LIOP [26] calculated over each set. At the last of this section, we present the descriptor construction.

A. Pre-processing, Feature detection and Normalization

The steps involved for pre-processing, feature detection and normalization are similar to [8, 12, 26, 28, 29, 30]. To remove the noises, a Gaussian smoothing with σ_p is used initially. To find a position and neighboring structure of point of interest, Harris-Affine/Hessian-Affine region detectors are considered. A circular region of size 41×41 pixels with radius 20.5 similar to other approaches [12, 26, 28, 29, 31] is generated by normalizing the detected region (Fig. 1). Finally, a Gaussian filter with sigma σ_n is applied again to cope up with the noise introduced by interpolation.

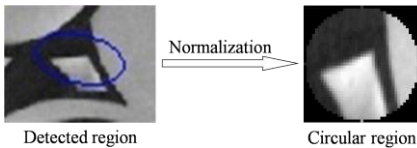


Fig. 1. Generating a circular patch by normalizing the detected patch of elliptical shape and arbitrary size.

B. Rotation Invariant Local Features

In order to facilitate local feature extraction in rotation invariant manner, we considered a local coordinate approach, similar to [10, 26, 28, 29]. Fig. 2 illustrates such a coordinate system, where O represents the center of the patch, X_i is any

pixel within the patch. Then a local rotation invariant coordinate system is generated from O and X_i for the sample pixel X_i centered at X_i by considering positive y-axis as,

$$\text{positive y-axis} = \vec{OX_i} \quad (1)$$

$\{X_i^1, X_i^2, \dots, X_i^N\}$ are the N neighbors of the X_i equally spaced in a circle of radius R centered at X_i . Angle ϕ is defined as,

$$\phi = \tan^{-1}\left(\frac{P_y}{P_x}\right) \quad (2)$$

where P_x and P_y are the co-ordinates of the pixel X_i with respect to the center of the patch O . The coordinates of N neighbors of X_i with respect to X_i are given by,

$$P_x^\alpha(X_i) = R\cos((\alpha - 1)\theta); P_y^\alpha(X_i) = R\sin((\alpha - 1)\theta) \quad (3)$$

where $\alpha \in [1, N]$ and angle θ is defined as,

$$\theta = \frac{2\pi}{N} \quad (4)$$

We represent the coordinate of X_i^α w.r.t. X_i as $P_{X_i}^\alpha$ using $P_x^\alpha(X_i)$ and $P_y^\alpha(X_i)$ as,

$$P_{X_i}^\alpha = P_x^\alpha(X_i) + jP_y^\alpha(X_i) \quad (5)$$

From (3) and (5), $P_{X_i}^\alpha$ is written as,

$$P_{X_i}^\alpha = R(\cos((\alpha - 1)\theta) + j\sin((\alpha - 1)\theta)) \quad (6)$$

We represent (6) using Euler's formula and $P_{X_i}^\alpha$ in Euler form is given as,

$$P_{X_i}^\alpha = Re^{j((\alpha-1)\theta)} \quad (7)$$

where $e^{j\omega} = \cos(\omega) + j\sin(\omega)$.

The intensity value of any neighboring pixel X_i^α is determined by the gray value at the coordinate $P_{X_i}^\alpha$ w.r.t. X_i and it is denoted by $I_{X_i}^\alpha$ and we refer all the neighboring intensity values of pixel X_i as $\mathbf{E}(X_i)$. Using this coordinate system, the rotation invariance is obtained inherently. It is easily observable that the position of each X_i^k w.r.t. X_i remains unchanged if the whole patch rotates in any direction (i.e. clockwise or anti-clockwise).

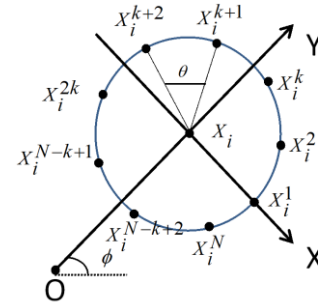


Fig. 2. Rotation invariant coordinate system to compute the location of the local features, O is the center of the patch and X_i is the sample point.

C. Local Neighbor Partitioning Into Interleaved Sets

The main problem with the earlier intensity order based descriptor [26] is with the rapid increase in the dimension of the descriptor with a slight increase in the number of neighboring points. In this section, we proposed to partition the N neighbors into k interleaved sets to overcome the problem of rapid increase in the descriptor's dimension with N . Fig. 3 illustrates the proposed approach to divide the original neighbors into multiple interleaved sets of local neighbors. Fig. 3(a) shows the original N neighbors $P_{X_i}^\alpha$ for $\alpha = [1, N]$ of a sample point X_i which are equally spaced at a distance R from center X_i .

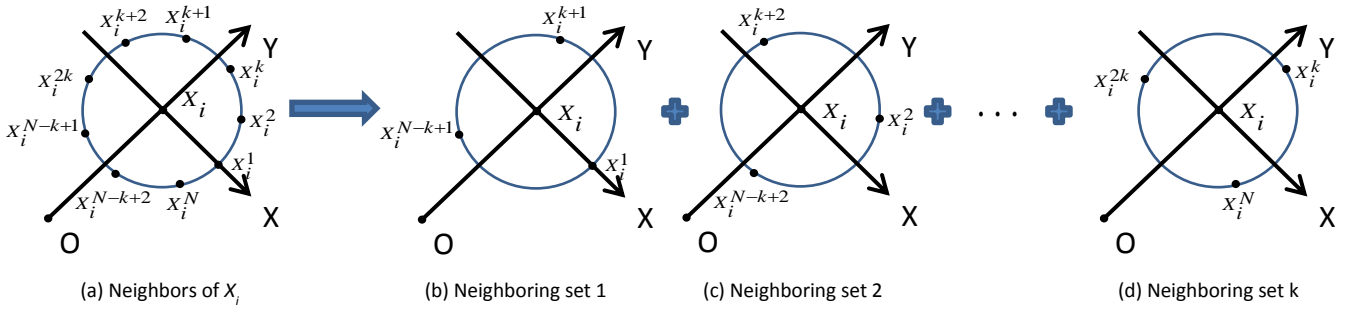


Fig. 3. Considering local neighborhood as a set of different interleaved local neighborhood. The original N neighbors are divided into k neighboring sets having $d=N/k$ neighbors each.

Fig. 3(b-d) represent k interleaved neighboring sets having d neighbors in each set generated from original N neighbors, where $d = N/k$. The coordinate of u^{th} neighbor of X_i in v^{th} neighboring set is given as,

$$P_{X_i}^{u,v} = Re^{j((u-1) \times \frac{2\pi}{d} + (v-1) \times \frac{2\pi}{d \times k})} \quad (8)$$

where $u \in [1, d]$ and $v \in [1, k]$.

By solving (8), $P_{X_i}^{u,v}$ is represented as,

$$P_{X_i}^{u,v} = Re^{j(\frac{[(u-1) \times k + (v-1)] \times 2\pi}{d \times k})} \quad (9)$$

If $(\beta - 1) = [(u - 1) \times k + (v - 1)]$, then $P_{X_i}^{u,v}$ is written as,

$$P_{X_i}^{u,v} = Re^{j((\beta-1) \times \frac{2\pi}{d \times k})} \quad (10)$$

The ranges of $(u - 1) \times k$ and $(v - 1)$ is computed from the ranges of u and v as,

$$(u - 1) \times k = [0, k, 2 \times k, \dots, (d - 1) \times k] \quad (11)$$

$$(v - 1) = [0, 1, 2, \dots, (k - 1)] \quad (12)$$

The range of $(\beta - 1)$ is computed from (11) and (12) as,

$$(\beta - 1) = [0, 1, 2, \dots, (k \times d - 1)] \quad (13)$$

If $k = 1$ and $d = N$, then the range of $(\beta - 1)$ becomes,

$$(\beta - 1) = [0, 1, 2, \dots, (N - 1)] \quad (14)$$

Now, the range of $(\beta - 1)$ is the same as the range of $(\alpha - 1)$ used in (7). By replacing $(\beta - 1)$ with $(\alpha - 1)$, k with 1 and d with N , $P_{X_i}^{u,v}$ in (10) becomes

$$P_{X_i}^{u,v} = Re^{j((\alpha-1) \times \frac{2\pi}{N})} \quad (15)$$

From (4) and (14), $P_{X_i}^{u,v}$ is written as,

$$P_{X_i}^{u,v} = Re^{j((\alpha-1) \times \theta)} \quad (16)$$

From (7) and (16), we conclude that,

$$P_{X_i}^{u,v} = P_{X_i}^{\alpha} \quad \text{if } k = 1 \text{ \& } d = N \quad (17)$$

It means, we consider original neighbors without division only if $k = 1$ and $d = N$ which is used by LIOP [26] (i.e., LIOP is a special case of our proposed approach). We also observed across the Fig. 3 that neighboring points in each neighboring set is also equally spaced in a circle of radius R having center at X_i . This is an advantage of our local neighborhood division and in this way it retains the symmetric information in the pattern.

We also illustrated the proposed idea of local neighbor partitioning into multiple interleaved sets using an example in the Fig. 4(a-c). An example patch for any pixel X_i is shown in the Fig. 4(a). We have considered 8 local neighbors of X_i here in this example as depicted in the Fig. 4(b) and partitioned it into 2 interleaved sets consisting of the 4 local neighbors each. The intensity values of the local neighbors in each set are demonstrated in the Fig. 4(c).

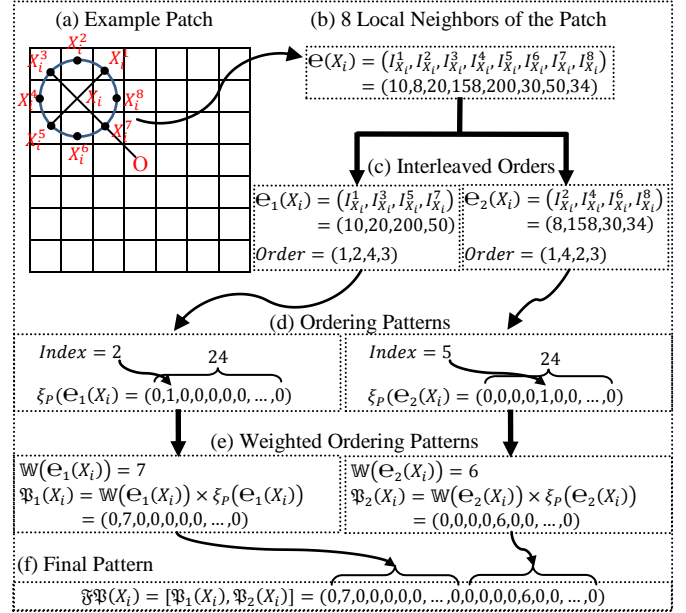


Fig. 4. Illustration of proposed concept of local neighborhood division into multiple interleaved sets and construction of IOLD pattern using an example. (a) example patch for pixel X_i , (b) intensity values of 8 local neighbors of considered patch, (c) partitioning of 8 local neighbors into 2 interleaved sets having 4 local neighbors each and its orders, (d) ordering patterns over each set, (e) weighted ordering patterns, and (f) final pattern for pixel X_i .

D. Computing Multiple Local Intensity Order Patterns

In this subsection, for each interleaved set, we construct the corresponding LIOP [26] pattern and then concatenate all the LIOPs to find the final pattern for a particular pixel. Let the intensity values of elements of the neighboring set v (i.e. points that fall in v^{th} interleaved set) are defined as,

$$\mathbf{E}_v(X_i) = (I_{X_i}^{1,v}, I_{X_i}^{2,v}, I_{X_i}^{3,v}, \dots, I_{X_i}^{d,v}) \quad (18)$$

where $v = [1, k]$ and $I_{X_i}^{u,v}$ is the intensity value of point $P_{X_i}^{u,v}$.

Note that the value of k is chosen in such a way that d should be a positive integer. We calculate a weighted ordering pattern $\mathfrak{P}_v(X_i)$ over each neighboring set $\mathbf{E}_v(X_i)$ using the method introduced in [26] as,

$$\mathfrak{P}_v(X_i) = \mathbb{W}(\mathbf{E}_v(X_i)) \times \xi_p(\mathbf{E}_v(X_i)) \quad (19)$$

where \mathbb{W} is a weight that encodes the dissimilarity information among the neighboring sample points and ξ_p is the ordering pattern of length $d!$.

The final interleaved order based local descriptor pattern is computed by concatenating the patterns for all neighboring set. Mathematically, we define the final pattern for pixel X_i as,

$$\mathfrak{F}\mathfrak{P}(X_i) = [\mathfrak{P}_1(X_i), \mathfrak{P}_2(X_i), \dots, \mathfrak{P}_k(X_i)] \quad (20)$$

According to the [26], the dimension \mathbb{D} of $\mathfrak{P}_j(X_i)$ is given by,

$$\mathbb{D}(\mathfrak{P}_j(X_i)) = d! \quad (21)$$

It means,

$$\mathbb{D}(\mathfrak{F}\mathfrak{P}(X_i)) = k \times (d)! \quad (22)$$

For two interleaved sets of intensity values of Fig. 4(c), its orders and ordering patterns computed using [26] are illustrated in the Fig. 4(c-e) respectively. Note that, in the example, we have partitioned 8 neighbors into 2 sets of 4 neighbors so the length of each ordering pattern is 24. Only the element corresponding to the index value of order is set to 1 in the ordering pattern and rest are zeros as illustrated in Fig. 4(d). We also calculated the weight for each set using its intensity values and multiplied it with the ordering patterns to get the weighted ordering patterns as depicted in the Fig. 4(e). Finally both weighted ordering patterns are concatenated to form the final pattern (see Fig. 4(f)) for the pixel X_i of Fig. 4(a) using its 8 local neighbors.

E. Descriptor Construction

The proposed IOLD descriptor construction workflow is demonstrated by Fig. 5. We consider B number of support regions centered at the feature point of minimal support region having the uniform increasing size similar to [28, 29]. Circular rings of size 41×41 are obtained by normalizing each support ring. Each support region is divided into C number of sub-regions based on the global intensity orders of each pixel in that support region similar to [26, 29]. The pattern over a sub-region is extracted by summing patterns of all pixels belonging to that sub-region. We refer j^{th} sub-region of i^{th} support region by \mathfrak{R}_i^j . Then descriptor \mathfrak{D}_i^j over sub-region \mathfrak{R}_i^j is calculated as follows,

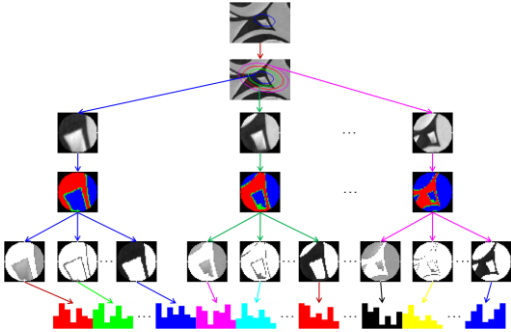


Fig. 5. IOLD descriptor construction process, B support regions is used with C sub-regions in each support region. The IOLD descriptor is constructed by accumulating local descriptor in each sub-region from all support regions.

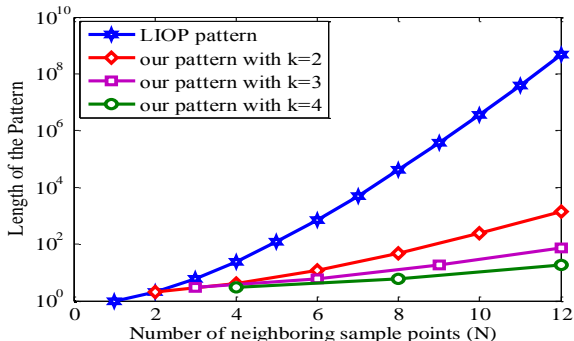


Fig. 6. Comparison between the pattern dimension using LIOP and proposed approach.

$$\mathfrak{D}_i^j = \sum_{t \in \mathfrak{R}_i^j} (\mathfrak{F}\mathfrak{P}(t)) \quad (23)$$

where $i \in [1, B]$, $j \in [1, C]$ and $\mathfrak{F}\mathfrak{P}$ is given by (20). The descriptor over a support region is computed by concatenating the descriptor computed over each sub-region of that support region. So descriptor over i^{th} support region \mathfrak{D}_i becomes,

$$\mathfrak{D}_i = [\mathfrak{D}_i^1, \mathfrak{D}_i^2, \dots, \mathfrak{D}_i^C] \quad (24)$$

The descriptor extracted over each support region is concatenated to compute the final IOLD descriptor.

Mathematically IOLD descriptor is given by,

$$IOLD = [\mathfrak{D}_1, \mathfrak{D}_2, \dots, \mathfrak{D}_B] \quad (25)$$

From (24) and (25), it is derived that IOLD descriptor can also be represented as,

$$IOLD = [\mathfrak{D}_1^1, \mathfrak{D}_1^2, \dots, \mathfrak{D}_1^C, \mathfrak{D}_2^1, \mathfrak{D}_2^2, \dots, \mathfrak{D}_2^C, \dots, \mathfrak{D}_B^1, \mathfrak{D}_B^2, \dots, \mathfrak{D}_B^C] \quad (26)$$

The dimension of IOLD for $B = 1$ and $C = 1$ using (22) is given as,

$$\mathbb{D}(IOLD) = k \times (d)! \quad (27)$$

The dimension of LIOP [26] for $C = 1$ is given as,

$$\mathbb{D}(LIOP) = N! \quad (28)$$

It is shown in Fig. 6 that $k \times (d)!$ is much lesser than the $N!$. It means,

$$\mathbb{D}(IOLD) \ll \mathbb{D}(LIOP) \quad (29)$$

By adapting local neighborhood division into several local neighborhoods (i.e. k neighboring sets), we reduce the pattern size significantly with comparable performance. The proposed ordering pattern is distinctive because it holds the invariance property for the rotation and illumination difference; moreover the symmetric information around the center pixel makes it more discriminative. It has also been shown that neighborhood division approach greatly reduces the descriptor size while maintaining the comparable results under noisy condition (Fig. 6 and 16).

III. EXPERIMENTS AND RESULTS

We compare SIFT [4], HRI-CSLTP [23] and LIOP [26] descriptors with IOLD descriptor to measure the effectiveness and discriminative ability of proposed descriptor. For evaluation purpose three widely used standard datasets namely Oxford image matching dataset [34], Complex illumination change dataset [35] and a large image matching dataset [36] have been used. The Oxford dataset comprises different geometric and photometric transformed image sets with textured and structured scenes. We used the Harris-Affine and Hessian-Affine detectors to detect the interest regions [34]. All the matching experiments is conducted using a personal computer having Intel(R) Core(TM) i5 CPU 650@3.20 GHz processor, 4 GB RAM, and 32-bit Windows 7 Ultimate operating system.

A. Evaluation Criteria

The criterion introduced in [12] is used for the evaluation of the descriptors in this paper. Each region of one image is matched with the every region of the second image and according to the number of false and correct matches the precision and recall values are generated. We calculate *all matches* using nearest neighbor distance ratio (NNDR) matching strategy. According to this scheme, a distance ratio is computed between 1st and 2nd nearest region.

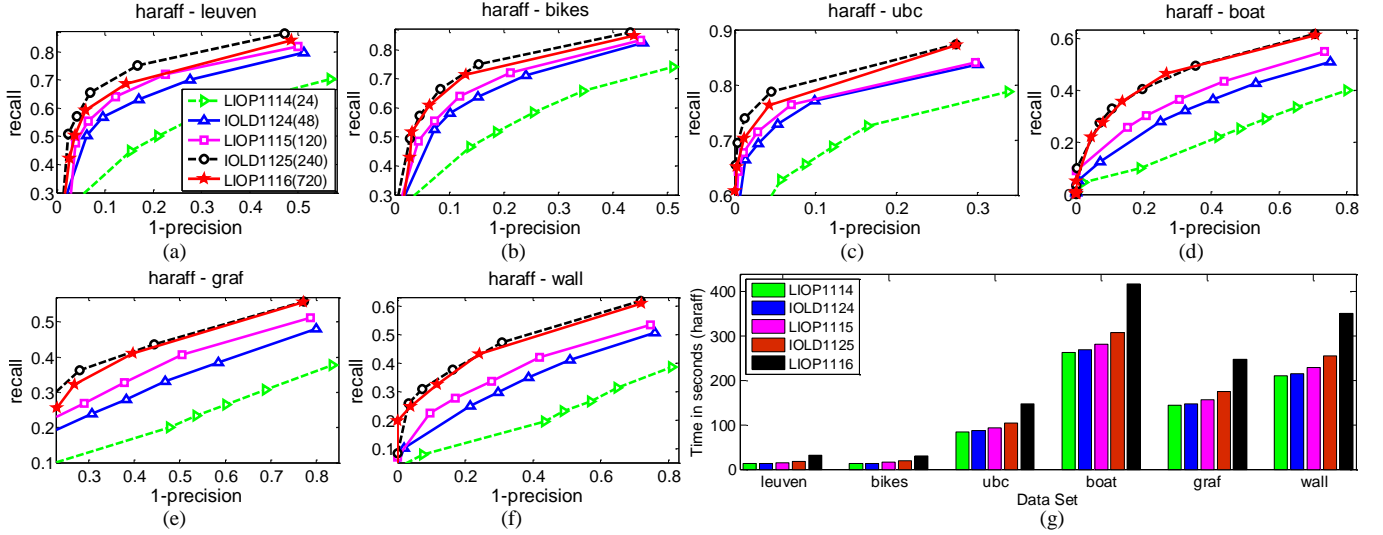


Fig. 7. Descriptors performance for $kd=14, 24, 15, 25$ and 16 when $B=1$ and $C=1$ using Harris-Affine region detector over Oxford dataset.

If the value of distance ratio is above a threshold then only a match is declared with 1st nearest region. By changing this distance threshold, different precision and recall values are obtained. We use overlap error [8] to determine the ground truth correspondences and the number of correct matches. The target region is transformed over the source region using a homography. The ratio of area of intersection and union between both regions (i.e. original source region and transformed target region) is used to find the overlap error. A match between two regions is acceptable if the overlap error < 0.5 . If $A1$ and $A2$ are the two regions, then overlap error between $A1$ and $A2$ is defined as,

$$\text{overlap error}(A1, A2) = 1 - \frac{A1 \cap A2}{A1 \cup A2} \quad (30)$$

We used *recall vs 1-precision* plots to present the matching results. If the number of correctly, falsely, all and ground truth matches are represented by $\#correct\ matches$, $\#false\ matches$, $\#all\ matches$, and $\#correspondences$ respectively, then,

$$\text{recall} = \frac{\#correct\ matches}{\#correspondences} \quad (31)$$

$$1 - \text{precision} = \frac{\#false\ matches}{\#all\ matches}$$

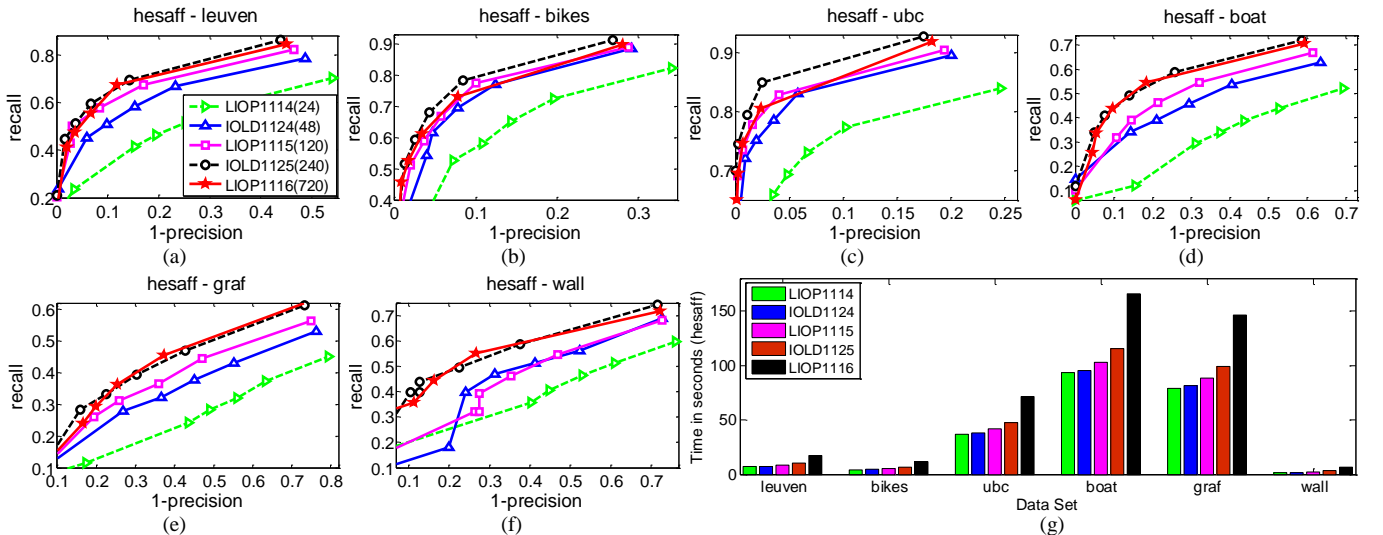


Fig. 8. Descriptors performance for $kd=14, 24, 15, 25$ and 16 when $B=1$ and $C=1$ using Hessian-Affine region detector over Oxford dataset.

We have used 1.0, 1.2 and 6 as the values of σ_p , σ_n and R in this paper similar to [26] for all experiments such that a fair comparison can be made between LIOP and IOLD.

B. Performance Evaluation on the Oxford Dataset

We used standard Oxford image matching dataset [34] for the evaluation of IOLD descriptor. IOLD is evaluated and compared for both Harris and Hessian Affine (i.e. haraff and hesaff) detectors. We considered 6 sequences of oxford dataset namely leuven (illumination change), bikes (image blur), ubc (jpeg compression), boat (rotation and scale), graf (viewpoint change) and wall (viewpoint change). Each sequence consists of the 6 images with increasing degree of the corresponding transformation. For a particular sequence, first image is matched with remaining five images (i.e., 5 pairs). Results are depicted in the terms of average performance over each pair for each sequence in Fig. 7-8 using recall and 1-precision. We compared the average performance and matching time by changing the number of neighboring sets k and the number of elements in each neighboring set d . To illustrate the effect of k and d the value of B (number of support regions) and C (number of partitions in a support region) is considered as 1.

Table 1. Image matching time reduced by IOLD1125 over LIOP1116 in % over each category of the Oxford dataset

Detector Used	Image Category of Oxford Dataset					
	leuven	bikes	ubc	boat	Graf	wall
Harris-Affine	70.53	59.66	41.32	35.27	41.23	37.83
Hessian-Affine	69.60	75.34	50.86	43.66	47.71	102.81

The value of k is considered as 1 and 2 and the value of d is considered as 4, 5 and 6. When $k=1$, we denoted it by LIOP because in this case IOLD is equivalent to the LIOP. In the source paper of LIOP, only 3 and 4 neighboring sample points are considered whereas in this paper we also experimented with the LIOP using more number of neighboring sample points (N) to test the effect of N over performance and matching time. Five combinations of $BCkd(\mathbb{D})$ (i.e. 1114(24), 1124(48), 1115(120), 1125(240) and 1116(720) respectively) are compared (see Fig. 7-8) where \mathbb{D} is the dimension of the descriptor. Fig. 7(a-f) and Fig. 8(a-f) shows the results when haraff and hesaff detector is used respectively while Fig. 7(g) and Fig. 8(g) shows the matching time by each combination of $BCkd$ for each sequence of Oxford dataset using haraff and hesaff detectors respectively. The significant improvement in the performance is reported when the value of k is increased to 2 for a particular d (i.e. between IOLD1124 and LIOP1114 and between IOLD1125 and LIOP1115). Consider the case of LIOP1116 and IOLD1125 combinations, the image matching time consumed by earlier one is much higher than later one for each sequence because $\mathbb{D}_{LIOP1116} > \mathbb{D}_{IOLD1125}$, while the performance of the IOLD1125 is either better or nearly equal to LIOP1116.

Table 1 depicts the % of matching time reduced by IOLD1125 over LIOP1116 for each set of images of Oxford dataset using both detectors. The highest improvement in the matching complexity is 102.81% reported for the wall sequences while hesaff detector is used. The plots of the Fig. 7-8 convey that by increasing only N the dimension \mathbb{D} increases more rapidly as compared to the performance but this problem can be overcome by dividing N into k interleaved sets. We also tested the proposed approach in conjunction with the multiple support region and region division concept. Fig. 9 reports the results and time consumption using haraff detector for $B=2$, $C=2$, $k=1, 2$ and $d=3$ combinations (i.e. comparison among LIOP2213 and IOLD2223).

We implemented LIOP2213 as the LIOP over multi-support-regions to show the effect of multiple support regions over LIOP and to compare it with IOLD implemented over multiple support regions. Here we reported the average performance and matching time over full oxford dataset for all combinations using each detector. It is observed that if k is increased the performance of the descriptor is still improved significantly using haraff detector but the degree of improvement is less whereas the matching time is nearly same (see Fig. 9(b)). The results for hesaff detector also follow the same trend as haraff detector.

Fig. 10 demonstrates the average image matching performance and matching time over Oxford dataset for SIFT, HRI-CSLTP, LIOP1116 and IOLD1125 descriptors using

haraff detector. We considered $BC=11$ for both LIOP and IOLD such that a fair comparison can be made in view of introduced concept. It is evident from this figure that the performance of IOLD1125 is better than the remaining descriptors for haraff detector and it is 49.40% and 15.53% faster than LIOP1116 and HRI-CSLTP respectively (i.e. the matching time is also better than other descriptors except SIFT).

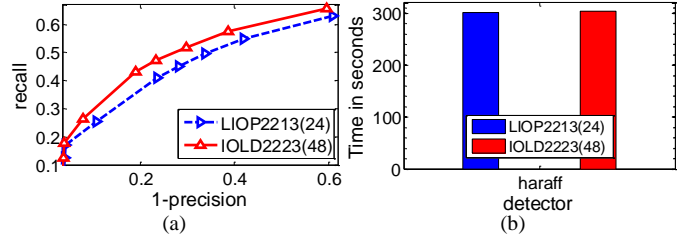


Fig. 9. (a) Matching results and (b) matching time over Oxford dataset in conjunction with B and C while $kd=13$ and 23 and $BC=22$ for haraff detector.

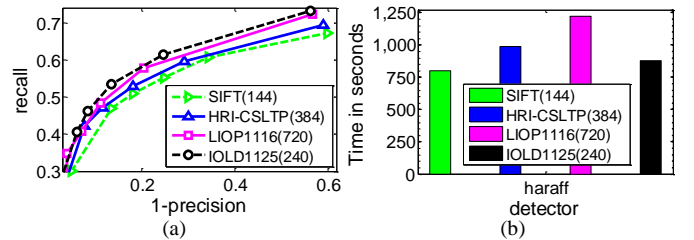


Fig. 10. Comparison of IOLD with LIOP, SIFT and HRI-CSLTP over Oxford dataset in terms of (a) ROC and (b) matching time using haraff detector.

C. Performance Evaluation on the Complex Illumination Change Dataset

We used a Complex illumination change dataset [35] in order to evaluate proposed descriptor for large illumination changes. Two image sets corridor and desktop of 6 images each having drastic illumination differences are used in this paper as shown in Fig. 11. We synthesized the 6th image of the corridor (i.e. corridor 6) from the 1st image of corridor having largest illumination difference from the 1st image of the corridor. The 5th and 6th image of the desktop is square and square root of the 4th image of the desktop respectively. Fig. 12 demonstrates the descriptors performance and matching time for $kd=14, 24, 15, 25$ and 16 when $BC=11$ using both region detector over Complex illumination change dataset. It is observed here also that the performance of IOLD descriptor with $k=2$ is improved significantly as compared to the LIOP descriptor with $k=1$ for a particular d .

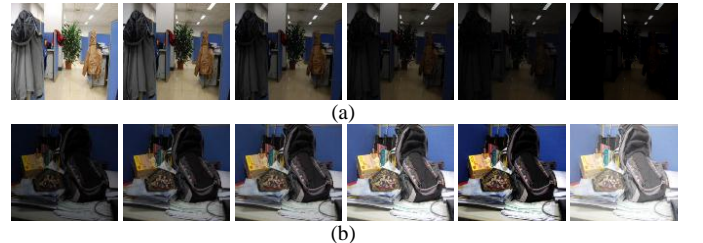


Fig. 11. Images of (a) Corridor and (b) Desktop category (6 images in each) of the complex illumination change dataset.

The results of IOLD1125 (dim: 240) is better than the results of LIOP1116 (dim: 720) whereas the matching time with LIOP1116 is much higher than the matching time with LIOP1125. In Fig. 13, we compared the IOLD descriptor with SIFT, HRI-CSLTP and LIOP descriptors using haraff detector over full Complex illumination change dataset in terms of average precision, average recall and matching time. Both LIOP and IOLD descriptors outperform SIFT and HRI-CSLTP descriptors because LIOP and IOLD descriptors are inherently invariant to the monotonic intensity change. The performance of IOLD is still comparable with the LIOP while maintaining the low dimensional feature description and the matching time with IOLD is significantly lower than the matching time with LIOP. It is observed across the plots (Fig. 12-13) that IOLD is able to maintain better results with low dimensional feature description under drastic illumination difference scenario.

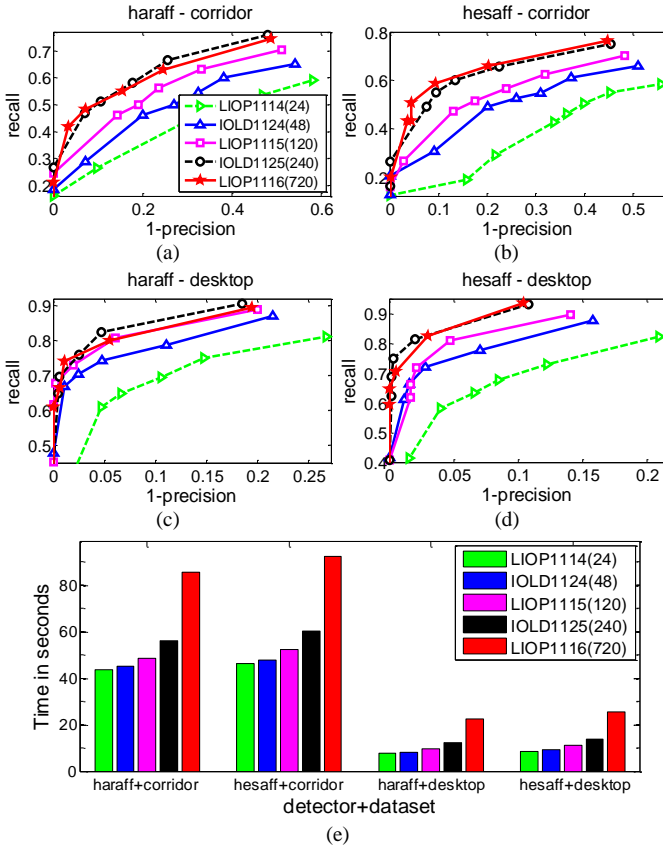


Fig. 12. (a-d) Descriptors performance and (e) matching time for $kd=14, 24, 15, 25$ and 16 when $BC=11$ using both region detector over Complex illumination change dataset.

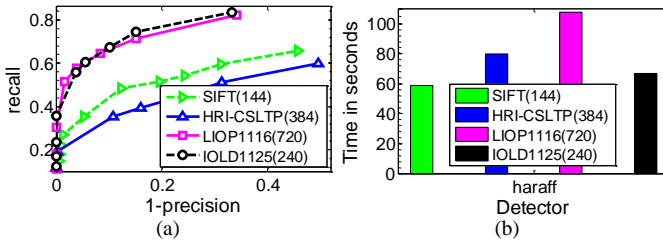


Fig. 13. Comparison of IOLD with LIOP, SIFT and HRI-CSLTP over Complex illumination change dataset in terms of (a) recall-precision and (b) matching time using haraff detector.

D. Performance Evaluation on Large Image Matching Dataset

To demonstrate the performance of proposed descriptor over large image matching dataset, we considered 190 pair of images which consists of 84, 63 and 43 pairs from rotation, illumination and zoom category respectively [36]. The image pairs already used in Oxford dataset are excluded in this experiment. The average results and matching time over large image matching dataset using SIFT, HRI-CSLTP, LIOP1116 and IOLD1125 are shown in Fig. 14. IOLD outperforms other descriptors using Harris-Affine region detector (see Fig. 14(a)). We observed that, in the case of Hessian-Affine region detector also, the performance of IOLD descriptor is comparable. The matching using IOLD is faster than LIOP and HRI-CSLTP by 1.43 and 1.13 times respectively and slower than SIFT by 0.89 times using hesaff detector and similar speedup also gained using haraff detector as shown in Fig. 14(b). The results and matching time suggest that the IOLD descriptor match the images more precisely and accurately with reasonable speed.

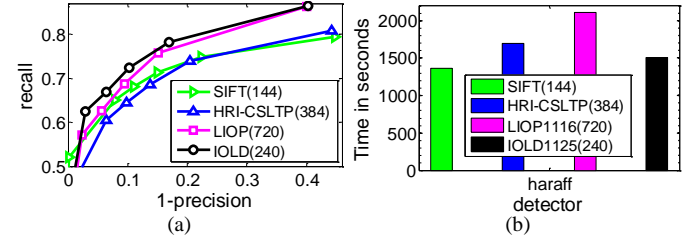


Fig. 14. Comparison of IOLD with LIOP, SIFT and HRI-CSLTP over large image matching dataset in terms of (a) recall-precision and (b) matching time using Harris-Affine detector.

IV. OBSERVATIONS AND DISCUSSIONS

In this section, we present some observations and discussions about the matching performance of IOLD descriptor under drastic illumination change and noisy conditions. In the last of this section, we analyze the matching time in terms of the number of matched key-points.

A. Effect of drastic illumination change over descriptor

To visualize the effect of the proposed approach on drastic illumination change, we considered a patch from the 1st image of the corridor and also same patch from the 4th image of the corridor. SIFT, LIOP1164 and IOLD1125 are computed from both the patches. LIOP and IOLD are quantized to the size of the SIFT such that the dimension of patterns becomes same for each descriptor. The difference between the patterns of both patches is computed for each descriptor. Fig. 15 presents the patches and similarity plot, two corresponding patches are shown in (a) and (b), and the similarity plot for the pattern difference is shown in (c). We observe that the global peak values (both +ive and -ive) is lowest for IOLD and highest for the SIFT descriptor. Another important factor is the overall deviation from zero in both directions (i.e. +ive and -ive) for each bin which is lowest for IOLD. We compared the normal distribution in (d) at zero mean ($\mu=0$). It is observed that the plot for IOLD is more tend to mean value and also have the highest peak value. From Fig. 15(d), it is concluded that the pattern of both patches are more similar using IOLD.

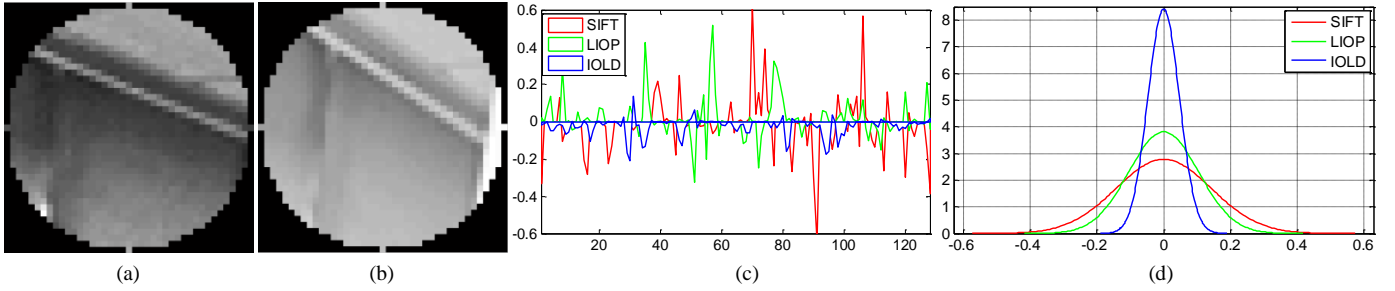


Fig. 15. Visualization of the performance of SIFT, LIOP and IOLD under illumination change, (a) a patch from the 1st image of the corridor, (b) same patch as of (a) but from the 4th image of the corridor, (c) the difference between the pattern of both patches for each descriptor, and (d) normal distribution of dissimilarities (c) at zero mean ($\mu=0$).

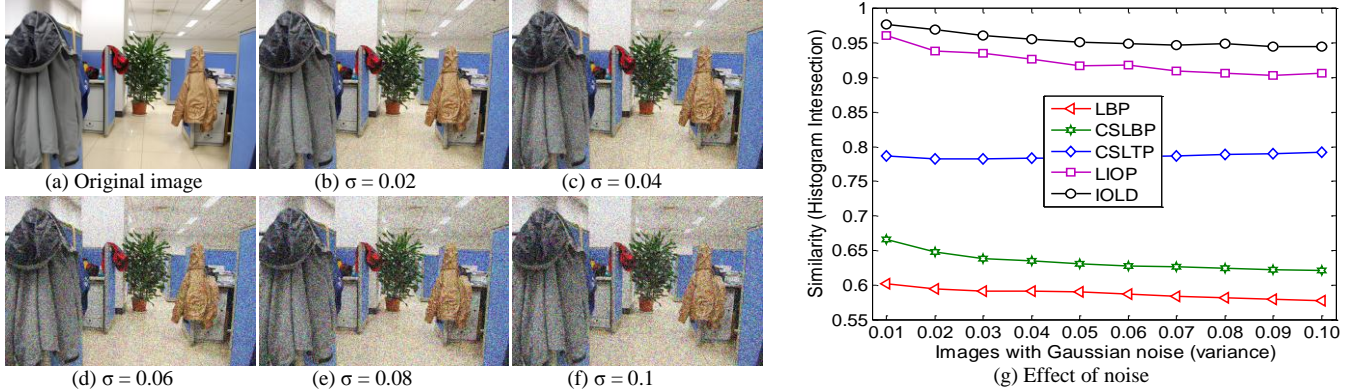


Fig. 16. Similarity between histograms of original and noised frames (effect of Gaussian noise), first desktop frame is the original frame and remaining frames are obtained after adding Gaussian noise in original frame with zero mean and σ variance.

Thus, we believe that incorporating order based approach proposed descriptor becomes more robust towards monotonic intensity change and provides more similar pattern for the similar patches under large illumination differences.

B. Effect of noise over descriptor

While good performance is achieved by LBP, CS-LBP, CS-LTP and LIOP operators (i.e. local order based methods), these methods are sensitive to the noises. We synthesized ten noisy frames from a desktop frame by adding Gaussian noise with zero mean and σ variance ($\sigma = [0.01, 0.1]$) at an interval of 0.01 to illustrate the effect of noise over descriptors. We compared LBP, CS-LBP, CS-LTP, LIOP1164 [23-26] and IOLD3224 (i.e. all methods based on the local ordering) using these noisy frames.

Fig. 16(a) depicts the original desktop frame used in this experiment and Fig. 16(b-f) shows some noisy frames obtained after adding the Gaussian noise in the original frame. The original frame's descriptor is compared with the same of each noisy. We used histogram intersection method [33] to compare two histograms. If the value of the similarity is tending towards one, it means that histograms are similar (i.e. that method is robust to noise). The performance of each method is shown in Fig. 16(g). LIOP and IOLD are having less sensitivity to the noise than LBP, CS-LBP and CS-LTP. LIOP are more robust than CS-LTP because it is the generalization of CS-LTP but IOLD are more robust to noise than LIOP because LIOP is a special case of proposed descriptor. IOLD descriptor is also consistent with the amount of the noise added.

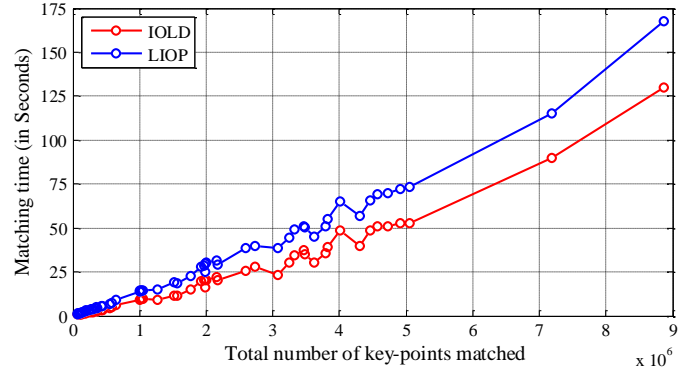


Fig. 17. Matching time vs number of matched key-points using IOLD descriptor and LIOP descriptor.

C. Matching time analysis in terms of the number of matched key-points

We have shown in the previous section that the dimension of the proposed IOLD descriptor is significantly lower than LIOP descriptor [26] whereas the performance of IOLD descriptor is either improved or nearly same than the LIOP descriptor. Here, we analyze the matching time in terms of the number of matched key-points using IOLD descriptor and the LIOP descriptor. Consider LIOP is constructed from the 6 local neighbors (i.e. LIOP1116 with dimension 720) and IOLD is constructed from the 10 local neighbors and two neighboring sets (i.e. IOLD1125 with dimension 240). We calculated the matching time for each pair of the images of Oxford image matching dataset [34]. The total number of image pairs is 30 in the Oxford dataset but we matched each

pair using both Harris-Affine and Hessian-Affine detectors so the total number of image pair comparison is 60. If the two images of any pair are having \mathcal{K}_1 and \mathcal{K}_2 number of key-points returned by any particular detector then total number of matched key-points for that pair will be $\mathcal{K}_1 \times \mathcal{K}_2$. Fig. 17 presents the matching time vs total number of matched key-points for both LIOP and IOLD descriptor. It is observed that the matching time for IOLD descriptor is always less than the LIOP descriptor. In other words the proposed IOLD descriptor is more efficient as compared to the LIOP descriptor in terms of matching time. Moreover, the degree of improvement increases with number of matched key-points. So, it is clearly deduced that the proposed approach is more efficient when the images are containing more details and of-course more number of extracted key-points.

The experiment shows that introduced interleaved order based local descriptor (IOLD) is better than other order based descriptors such as LIOP and HRI-CSLTP in terms of both performance and time complexity. The performance of proposed descriptor is better under each geometric and photometric transformation considered in this paper (i.e. scale change, JPEG compression, viewpoint change, image rotation, image blur and illumination difference). IOLD descriptor also performs very well under drastic illumination differences. We also compared proposed descriptor under noisy condition and found that IOLD is less prone to noise as compared to LBP, CSLBP, CSLTP and LIOP. The multiple intensity orders computed from different neighboring sets provide the discriminative ability to proposed descriptor and make it robust for different image transformations. The results obtained using IOLD descriptor points that IOLD outperforms other prominent descriptors proposed recently.

V. CONCLUSION

To overcome the problem of rapid growth in the descriptor's dimension with slight increase in the number of neighboring sample points, an interleaved neighbor division approach is presented in this paper. An interleaved order based local descriptor (IOLD) is introduced by computing the ordering patterns over multiple neighboring sets. IOLD incorporates the advantage of local features extracted in a rotation invariant manner. It computes the local intensity orders to achieve the invariance property towards monotonic intensity change. The robustness and discriminating ability of proposed descriptor is increased by using more than one interleaved intensity orders derived from multiple neighboring sets of the neighboring sample points. Multiple support regions and region partitioning into sub-regions further improve our descriptor. By incorporating all these, proposed descriptor becomes more robust and invariant towards various geometric and photometric image transformations. IOLD greatly reduces the matching time on an average with a factor of 49.40% while having the comparable performance. Results obtained on the image matching experiments suggest that the proposed IOLD descriptor is more time efficient and able to discriminate the images more robustly. In the presence of noise also IOLD is performing more robustly. IOLD outperforms other state-of-the-art descriptors under different imaging conditions.

REFERENCES

- [1] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 72–79, 2009.
- [2] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, Aug. 2007.
- [3] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 835–846, July 2006.
- [4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [5] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, June 2007.
- [6] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, May 2009.
- [7] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *Proc. British Machine Vision Conference*, pp. 384–393, 2002.
- [8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A Comparison of Affine Region Detectors," *International Journal of Computer Vision*, vol. 65, no. 1–2, pp. 43–72, Nov. 2005.
- [9] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, Sept. 1991.
- [10] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.
- [11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–521, April 2002.
- [12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [13] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Proc. European Conf. Computer Vision*, vol. 1, pp. 404–417, 2006.
- [14] E. N. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 184–190, 2005.
- [15] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, 2010.
- [16] L. Gong, T. Wang, and F. Liu, "Shape of Gaussians as Feature Descriptors," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 2366–2371, 2009.
- [17] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A Blur-Robust Descriptor with Applications to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1220–1226, June 2012.
- [18] H. Y. Chen, Y. Y. Lin, and B. Y. Chen, "Robust Feature Matching with Alternate Hough and Inverted Hough Transforms," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 2762–2769, 2013.
- [19] B. Kim, H. Yoo, and K. Sohn, "Exact order based feature descriptor for illumination robust image matching" *Pattern Recognition*, vol. 46, no. 12, pp. 3268–3278, 2013.
- [20] C. Zhu, C.-E. Bichot, and L. Chen, "Image region description using orthogonal combination of local binary patterns enhanced with color information" *Pattern Recognition*, vol. 46, no. 7, pp. 1949–1963, 2013.
- [21] F. Tang, S. H. Lim, N. L. Chang, and H. Tao, "A novel feature descriptor invariant to complex brightness changes," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 2631–2638, 2009.
- [22] F. Tang, S. H. Lim, and N. L. Chang, "An improved local feature descriptor via soft binning," *Proc. IEEE Int'l Conf. Image Processing*, pp. 861–864, 2010.
- [23] R. Gupta, H. Patil, and A. Mittal, "Robust order-based methods for feature description," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 334–341, 2010.

- [24] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [25] M. Heikkila, M. Pietikainen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, March 2009.
- [26] Z. Wang, B. Fan, and F. Wu, "Local Intensity Order Pattern for feature description," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 603–610, 2011.
- [27] T. Ojala, M. Pietikainen, and M. Maenpaa, "Multi resolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [28] B. Fan, F. Wu, and Z. Hu, "Rotationally Invariant Descriptors Using Intensity Order Pooling," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2031–2045, Oct. 2012.
- [29] B. Fan, F. Wu, Z. Hu, "Aggregating Gradient Distributions into Intensity Orders A Novel Local Image Descriptor," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 2377–2384, 2011.
- [30] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," *Proc. 7th European Conf. Computer Vision - Part I*, pp. 128–142, 2002.
- [31] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 511–517, 2004.
- [32] S. Winder and M. Brown, "Learning local image descriptors," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2007.
- [33] G. Finlayson, S. Hordley, G. Schaefer, and G. Y. Tian, "Illuminant and device invariant colour using histogram equalization," *Pattern Recognition*, vol. 38, no. 2, pp. 179–190, 2005.
- [34] <http://www.robots.ox.ac.uk/~vgg/research/affine/>.
- [35] http://vision.ia.ac.cn/Students/wzh/datasets/illumination/Illumination_Datasets.zip.
- [36] <http://lear.inrialpes.fr/people/mikolajczyk/>.
- [37] Z. Wang, B. Fan and F. Wu, "FRIF: Fast Robust Invariant Feature", *Proc. British Machine Vision Conference*, 2013.